© 2020. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

RESEARCH ARTICLE

A highly stable, non-digestible lectin from *Pomacea diffusa* unveils clade-related protection systems in apple snail eggs

Tabata R. Brola[1], Marcos S. Dreon[1,2,*], Jian-Wen Qiu[3] and Horacio Heras[1,4]

ABSTRACT

The acquisition of egg protection is vital for species survival. Poisonous eggs from *Pomacea* apple snails have defensive macromolecules for protection. Here we isolated and characterized a novel lectin called PdPV1 that is massively accumulated in the eggs of *Pomacea diffusa* and seems part of its protective cocktail. The native protein, an oligomer of *ca* 256 kDa, has high structural stability, withstanding 15 min boiling and denaturing by SDS. It resists *in vitro* proteinase digestion and displays structural stability between pH 2.0 and pH 12.0, and up to 85°C. These properties, as well as its subunit sequences, glycosylation pattern, presence of carotenoids, size and global shape resemble those of its orthologs from other *Pomacea.* Furthermore, like members of the *canaliculata* clade, PdPV1 is recovered unchanged in feces of mice ingesting it, supporting an anti-nutritive defensive function. PdPV1 also displays a strong hemagglutinating activity, specifically recognizing selected ganglioside motifs with high affinity. This activity is only shared with PsSC, a perivitelline from the same clade (*bridgesii* clade). As a whole, these results indicate that species in the genus *Pomacea* have diversified their egg defenses: those from the *bridgesii* clade are protected mostly by non-digestible lectins that lower the nutritional value of eggs, in contrast with protection by neurotoxins of other *Pomacea* clades, indicating that apple snail egg defensive strategies are clade specific. The harsh gastrointestinal environment of predators would have favored their appearance, extending by convergent evolution the presence of plant-like highly stable lectins, a strategy not reported in other animals.

KEY WORDS: Ampullariidae, Egg defense, Gastropods, Glycan array, Lectin activity, Protein structure

INTRODUCTION

Pomacea freshwater snails (Caenogastropoda, Ampullariidae), particularly those of invasive species belonging to the *canaliculata* clade, lay colorful and poisonous egg masses on emergent substrates above the water level, thus exposing eggs to environmental stressful agents and terrestrial predators (Przeslawski, 2004; Hayes et al., 2015). These defenses pay off, and as a result their eggs have almost no reported predators except for the fire ant *Solenopsis geminata* (Yusa, 2001). We have found that these defenses are provided by proteins (perivitellins) already active in the albumen gland of females where the egg components are synthesized (Cadierno et al., 2018a,b), a likely explanation why the brown rat (*Rattus norvegicus*) and the snail kite (*Rostrhamus sociabilis*), which predate on adult snails, systematically discard the albumen gland (Yusa et al., 2000).

Extensive studies of the most abundant perivitellins in *Pomacea* species, collectively called PV1s, have shown that they are all oligomeric carotenoproteins of high molecular weight, representing ∼70% of the perivitelline fluid (PVF) soluble protein (Dreon et al., 2003; Ituarte et al., 2008; Pasquevich et al., 2014). These carotenoproteins are highly stable in a wide range of pH and temperatures (Dreon et al., 2007, 2008; Sun et al., 2012; Pasquevich et al., 2017) and withstand gastrointestinal digestion (Dreon et al., 2010, 2013; Ituarte et al., 2012; Pasquevich et al., 2017). Nevertheless, some differences emerged: although the PV1 carotenoprotein of the non-invasive species *Pomacea scalaris* (PsSC) is a lectin (carbohydrate-binding protein) with a remarkably strong activity (Ituarte et al., 2012, 2018), its orthologs from the invasive species *Pomacea canaliculata* and *Pomacea maculata* (PcOvo and PmPV1, respectively) lack lectin activity.

Many lectins have been experimentally identified in molluscs, mostly in eggs or in organs involved in the synthesis of egg components. Probably the best-known gastropod egg lectin is *Helix pomatia* agglutinin, which nowadays has important biomedical applications as a histopathological biomarker of tumors displaying differential glycosylation patterns (Sanchez et al., 2006). In ampullariids, lectin activity has been reported in the eggs of *Pila ovata* (Uhlenbruck et al., 1973) and *Pomacea urceus* (Baldo and Uhlenbruck, 1974) and assumed to be part of the innate immune system against microbial invasions (Prokop et al., 1968). However, the only ampullariid lectin isolated and functionally characterized so far is the PV1 carotenoprotein PsSC from *P. scalaris* eggs. This lectin role immediately suggests an unexpectedly functional divergence within *Pomacea* egg carotenoproteins (Ituarte et al., 2018), but this hypothesis needs further comparative analysis. In addition to PV1s, the conspicuous pink-reddish eggs of the invasive species have a neurotoxic perivitelline lethal to mice (Heras et al., 2008; Giglio et al., 2016) that is absent in the non-invasive species *P. scalaris* which has pale pinkish eggs. This prompted us to study whether the different defensive roles of perivitellins against predation was associated with their phylogeny (Hayes et al., 2009), which clusters *P. maculata* and *P. canaliculata* in the more defined *canaliculata* clade and *P. scalaris* and *P. diffusa* in the more basal *bridgesii* clade.

To better understand carotenoprotein roles and evolutionary trends in *Pomacea* egg defenses, we studied *P. diffusa*, a member of the basal clade, analyzing its egg perivitellin fluid and characterizing PdPV1, a novel PV1 carotenoprotein with lectin activity, reporting its structure, structural stability and functional features in a phylogenetic framework. We suggest a putative role of

[1]Instituto de Investigaciones Bioquímicas de La Plata Prof. R. R. Brenner (INIBIOLP), Universidad Nacional de La Plata (UNLP) – CONICET, C. P. 1900 La Plata, Argentina. [2]Cátedra de Bioquímica y Biología Molecular, Facultad de Ciencias Médicas, UNLP, C. P. 1900 La Plata, Argentina. [3]Department of Biology, Hong Kong Baptist University, Hong Kong, PRC. [4]Cátedra de Química Biológica, Facultad de Ciencias Naturales y Museo, UNLP, C. P. 1900 La Plata, Argentina.

*Author for correspondence (msdreon@gmail.com)

T.R.B., 0000-0002-7803-795X; M.S.D., 0000-0002-3427-4523; J.-W.Q., 0000-0002-1541-9627; H.H., 0000-0003-3379-0216

Received 23 June 2020; Accepted 21 July 2020

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

PdPV1 in embryo protection as both an anti-nutritive protein (highly stable, non-digestible) and as a toxic lectin. Furthermore, we describe an evolutionary association among *Pomacea* egg defense systems, their phylogenetic position and invasiveness.

MATERIALS AND METHODS

Ethics statement

Studies performed with animals (protocol number P01-01-2016) were approved by the 'Comité Institucional para el Cuidado y Uso de Animales de Laboratorio' (CICUAL) of the School of Medicine, Universidad Nacional de La Plata (UNLP) and in accordance with the Guide for the Care and Use of Laboratory Animals.

Animals

Female BALB/c mice were obtained from the Experimental Animals Laboratory, School of Veterinary Science, National University of La Plata (UNLP), and weighed around 16 g. *Pomacea diffusa* Blume 1957 adults were purchased from a local aquarium and genetically identified by sequencing the cytochrome oxidase C subunit I gene, using LCO 1490 and HCO 2198 primers for DNA amplification (Matsukura et al., 2013). Sequence was then compared with those deposited in GenBank (NCBI). Snails were raised in the laboratory in 50-liter aquaria, at 25°C and fed with lettuce and fish food *ad libitum*. Egg masses were collected as soon as they were laid and were never beyond the morula stage.

Reagents

The reagents used were obtained from Carlo Erba Reagents (Barcelona, Spain) or Merck/Sigma. Molecular standards and nitrocellulose strips were purchased from GE Healthcare Bio-Sciences (Uppsala, Sweden). Lectins, streptavidin conjugate and ABC system for electrochemiluminescence were from Vector Labs (Burlingame, CA, USA). Goat anti-rabbit horseradish peroxidase conjugated secondary antibody was bought from BioRad Laboratories, Inc. Alexa Fluor 488 Protein Labeling kit was from Invitrogen. U-shaped microtiter plates for hemagglutinating assays were from Greiner Bio-One (Frickenhausen, Germany).

Protein isolation and purification

Egg homogenate was prepared in ice-cold 20 mmol l^{-1} Tris–HCl, pH 7.4, 3:1 v/w, using a Potter homogenizer. The homogenate was then sequentially centrifuged at 10,000 ***g*** for 20 min in an Avanti JE centrifuge (Beckman, Palo Alto, CA, USA) and at 100,000 ***g*** for 45 min on a Beckman L8M centrifuge at 10°C. Both pellets were discarded and the supernatant (henceforth referred to as perivitelline fluid, PVF) was layered on NaBr δ=1.28 g ml^{-1} and centrifuged at 207,000 ***g*** for 22 h at 10°C on a Beckman L8M centrifuge. A tube layered with 20 mmol l^{-1} Tris–HCl in lieu of the sample was used as a blank for density calculations. Subsequently, aliquots of 200 µl were collected sequentially starting at the top of the tube. Absorbance of each aliquot was determined at 280 nm to obtain the protein profile. The refractive index of the blank tube aliquots was determined at 26°C with a refractometer (Bausch and Lomb, New York, NY, USA) and converted to density using tabulated values (Orr et al., 1991). Fractions showing absorbance peaks at 280 nm were pooled and purified by size exclusion chromatography (SEC) in a HPLC system (1260 Infinity; Agilent Technologies, Waldron, Germany) using a Superdex 200 HR 10/300 GL column (GE Healthcare Bio-Sciences) as described by Ituarte et al. (2018). Purification process of the protein (hereafter named PdPV1) was checked by polyacrylamide gel electrophoresis (PAGE) as described below. Protein content was determined using bovine serum albumin (BSA) as standard (Lowry et al., 1951) in an Agilent 8453 UV-visible diode array spectrophotometer. PsSC was prepared as described elsewhere (Ituarte et al., 2008).

Polyacrylamide gel electrophoresis

Native (non-denaturing) PAGE was performed in a 4–20% gradient gel in a Mini-Protean III System (Bio-Rad). High molecular mass standards were run in the same gels. To assay denaturant conditions, 4–20% SDS-PAGE containing 0.1% SDS was carried out; samples were denatured at 100°C for 10 min with and without reducing agent (2-ME). Gels were stained with Coomassie Brilliant Blue R-250. *Pomacea diffusa* PVF was also analyzed by native PAGE and protein bands were quantified by calibrated scanning densitometry using ImageJ software (Bourne and Bourne, 2010).

Size and global shape

The molecular weight of PdPV1 was estimated by 4–20% Native PAGE using high molecular weight standard and by SEC as previously described in Ituarte et al. (2008), calibrated with thyroglobulin, ferritin, catalase, aldolase and ovoalbumin as molecular weight standards.

Protein global shape was determined by small angle X-ray scattering (SAXS). Assays were performed at the D02A-SAXS2 line, in the Laboratorio Nacional de Luz Sincrotron, Campinas (SP, Brazil). The scattering pattern was detected using a MARCCD bidimensional charge-coupled device assisted by Fit2d software (Hammersley, 2016). The experiments were performed using a wavelength of 1.448 Å. The distance between the sample and the detector was 1044 mm, allowing a Q-range between 0.012 and 0.25 $Å^{-1}$ (D_{max}=260 Å). Temperature was controlled using a circulating water bath, and kept at 25°C. Each individual run was corrected for sample absorption, photon flux, buffer scattering and detector homogeneity. At least three independent curves were averaged for each single experiment, and buffer blank scattering was subtracted. To rule out a concentration effect in the data, SAXS experiments with a protein concentration range of 3.0–0.2 mg ml^{-1} were performed. Data were analyzed using the ATSAS package 2.6.0 (Petoukhov et al., 2012). The low-resolution model of PdPV1 was obtained from the algorithm built with DAMMIN and DAMMIF programs (Svergun, 1999). The average of the best 10 models fitting the experimental data was obtained with DAMMIF. Raw data were deposited in the Small Angle Scattering Biological Data Bank (SASBDB).

Absorption spectra

UV-visible absorption spectra of PdPV1 were measured between 250 and 650 nm using an Agilent 8453 UV-visible spectrophotometer at room temperature. Three spectra were recorded for each sample and its buffer contribution (50 mmol l^{-1} $NaHPO_4$, 150 mmol l^{-1} NaCl; pH 7.0) was subtracted.

Carbohydrate content

Carbohydrate content of PdPV1 was determined following the phenol/sulfuric method described by Dubois et al. (1956) using D-glucose as standard. Samples were diluted to a final concentration of 50 µg ml^{-1} (final volume 150 µl). All samples were treated with a 5% (v/v) phenol solution and concentrated sulfuric acid. After 30 min at 37°C, absorbance was read at 485 nm to determine hexose and pentose content.

Determination of glycan motifs using lectin dot blot assay

A set of 13 biotinylated lectins were used, namely ConA, WGA, PSA, PNA, JAC, RCA 1, SBA, DBA, UEA, PVL, LCA, ECL and

BSL 1. Lectins were reconstituted in 10 mmol l^{-1} Hepes buffer and 0.1 mmol l^{-1} $CaCl_2$ and then diluted in 3% (w/v) BSA in phosphate-buffered serum (PBS). Dots containing 3 µg of PdPV1 were spotted onto nitrocellulose strips and incubated for 30 min at room temperature. Membranes were then blocked with 3% (w/v) BSA in PBS overnight at 4°C and incubated for 1.30 h with 0.02 mg of the lectin. They were then incubated with horseradish peroxidase streptavidin conjugate diluted 1:3 in water for 20 min at 25°C. Between every step, strips were washed five times with 0.1% Tween in PBS for 3 min. Positive reactions were visualized by electrochemiluminescence using the ABC system following the manufacturer's instructions.

Immunoblotting

Cross-reactivity of PdPV1 with anti-PcOvo and anti-PsSC polyclonal antibodies (Dreon et al., 2003; Ituarte et al., 2008) was analyzed by western blot assays (WB). Briefly, purified PdPV1 as well as PcOvo and PsSC perivitellins were run in 4–20% SDS-PAGE and then transferred onto nitrocellulose membranes using 25 mmol l^{-1} Tris–HCl, 192 mmol l^{-1} glycine, 20% (v/v) methanol, pH 8.3 buffer in a Mini Transblot Cell (Bio-Rad) at 100 V for 1 h. Membranes were blocked with 5% (w/v) casein in PBS overnight at 4°C and then incubated with each polyclonal antibody at different dilutions in 3% (w/v) casein–PBS for 2 h. Finally, a goat anti-rabbit horseradish peroxidase conjugated secondary antibody was prepared in 3% (w/v) casein–PBS, incubated for 1 h with the membranes and the reaction was revealed by luminol-coumaric reaction.

Subunit sequences and bioinformatic analysis

Internal amino acid sequences of purified PdPV1were obtained by mass spectrometry at the CEQUIBIEM (UBA-CONICET, Argentina). N-terminal sequences of subunits were determined by Edman degradation at the Laboratorio Nacional de Investigación y Servicios en Péptidos y Proteínas (LANAIS-PRO, UBA-CONICET, Argentina). These partial sequences were used to search the *P. diffusa* albumen gland transcriptome in AmpuBase (http://www.comp.hkbu.edu.hk/~db/AmpuBase/blast.php) for the corresponding full protein sequences. Sequence alignment was performed using MUSCLE. PdPV1 sequences were analyzed using PsiBlast, FFAS server and Hhpred. A search in PROSITE was performed to look for conserved domains. For low complexity regions, we used SEG server. The signal peptide cleavage sites in the amino acid sequences were predicted using the SignalP 4.1 server. The theoretical molecular weight and isoelectric point of each mature subunit were estimated using the ProtParam tool-Expasy server. Potential phosphorylation and glycosylation sites were predicted with DISPHOS 1.3 (Iakoucheva et al., 2004) and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc) and NetOGlyc 4.0 (Steentoft et al., 2013) servers, respectively. PdPV1, PsSC, PcOvo and PmPV1 subunit sequences were subjected to phylogenetic analysis using the maximum likelihood method in MEGA X (Kumar et al., 2018).

Chemical denaturation

Protein chemical denaturation by guanidine hydrochloride (GdnHCl) of PdPV1 was studied following intrinsic tryptophan fluorescence emission. Increasing concentrations (0–6.5 mol l^{-1}) of GdnHCl buffered with 0.2 mol l^{-1} Tris–HCl were incubated with the samples (final concentration 0.15 g l^{-1}) at pH 7.4. Fluorescence spectra were recorded on a Fluorolog 3 Horiba spectrofluorometer (Jobin Yvon Technology, Moulton Park, Northampton, UK) at 25°C. Excitation was set at 295 nm (slit of 4 nm) and recorded from 300 to 410 nm (slit of 5 nm). Three spectra were recorded and averaged, and the corresponding buffer contribution subtracted for each sample. The protein unfolded fraction (f_u) was calculated using the fluorescence intensity at 327.9 nm and assuming a two-state unfolding process. The equilibrium reached in each GndHCl concentration allows the calculation of an equilibrium constant $K=f_u/(1-f_u)$ and the Gibb's free energy for the unfolding reaction in terms of this mole fraction ($\Delta G^o=-RT\ln K$) was calculated. The dependence of ΔG^o on GndHCl concentration can be approximated by the linear equation $\Delta G^o=\Delta G^o_{H_2O}-m$[GndHCl], where the free energy of unfolding in the absence of denaturant ($\Delta G^o_{H_2O}$) represents the conformational stability of the protein. The GdnHCl concentration in which half of the protein is unfolded (C_m) was estimated as a function of denaturant concentration from the linear extrapolation method.

Resistance to SDS

Denaturation induced by SDS is used to identify proteins whose native conformations are kinetically trapped in a specific conformation because of an unusually high unfolding barrier that results in very slow unfolding rates (Jayaprakash and Surolia, 2017). To test PdPV1 resistance to denaturation by SDS, an experiment following the Manning and Colón procedure was performed (Manning and Colón, 2004). In short, the protein in Laemmli sample buffer (pH 6.8) containing 1% SDS was either boiled for 10 min or unheated. Samples were then analyzed by 4–20% SDS gel electrophoresis and visualized with Coomassie Brilliant Blue.

Structural stability against temperature and pH

Three-hundred microliters of PdPV1 (0.28 g l^{-1}) in 50 mmol l^{-1} Na_2HPO_4 and 150 mmol l^{-1} NaCl buffer at pH 7.0 were used for the assays. For thermal stability assays, the sample was subjected to rising temperatures from 25 to 85°C and, to assess its stability at different pH values, the protein was incubated in buffers ranging from pH 2.0 to pH 12.0 overnight at 4°C. Buffer formulations were taken from Merril (1990). PdPV1 tertiary structure was monitored by absorbance and fluorescence spectroscopy and global shape and size by SAXS. Absorbance was measured between 250 and 800 nm with an Agilent 8453 spectrophotometer and the 4th derivative spectra calculated. Fluorescence emission was recorded from 300 to 420 nm in a Varian Cary Eclipse spectrofluorometer (Varian Inc., Australia), excitation set to 295 nm with a slit of 8 nm, in 5 mm optical path quartz cells. For both techniques three spectra were recorded and averaged for each sample and its corresponding buffer blank was subtracted. SAXS experiments were performed as described above, except for thermal stability of PdPV1 for which the sample holder temperature was controlled using a water circulating bath between 25 and 85°C. The effect of extreme thermal conditions was analyzed by boiling PdPV1 for 15 min and evaluating the oligomer integrity using native (non-denaturing) gel electrophoresis and the carotenoprotein fine spectra using absorbance spectrophotometry.

Capacity to withstand *in vivo* and *in vitro* gastrointestinal digestion

PdPV1 resistance to protease degradation was studied *in vitro* and *in vivo* as previously described (Dreon et al., 2010; Pasquevich et al., 2017). For the simulated gastrointestinal digestion, the purified protein (105 µg) in phosphate buffer was first subjected to pepsin at an enzyme:substrate ratio of 1:20 (w/w), at 37°C in a simulated gastric fluid (SGF) containing 150 mmol l^{-1} NaCl

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

(pH 2.5). BSA was used as control. Aliquots of the sample and control were taken after 60 and 120 min of incubation. To stop the reaction, pH was raised to 8.5 with 150 mmol l^{-1} Tris-Cl buffer and samples were kept at −20°C until processing. The remnant of the solutions was then incubated with trypsin to simulate the gut phase in a 1:2.7 enzyme:substrate ratio (w/w) at 37°C in SGF. The pH of the solution was adjusted with 0.1 mol l^{-1} NaOH, and 0.25 mol l^{-1} taurocolic acid and 0.15 mol l^{-1} Tris–HCl buffer (pH 8.5) were added. Aliquots of the sample and control were taken at the same times as in gastric phase and stored at −20°C. Finally, all samples were boiled at 100°C for 10 min and run in 4–20% SDS-PAGE gradient to check the integrity of the protein.

To evaluate PdPV1 resistance to *in vivo* gastrointestinal digestion, BALB/c mice were orally administrated with 0.6 mg of PdPV1 in 50 mmol l^{-1} Na_2HPO_4 and 150 mmol l^{-1} NaCl buffer and their feces were collected every hour for 4 h after oral administration. Before and during the experiment, animals were fed *ad libitum*. Control animals were only given buffer. Fecal extracts were resuspended in 30 mmol l^{-1} EDTA (pH 8.4) in PBS with 1:100 protease inhibitor (Sigma) and homogenized in a Teflon homogenizer before centrifugation at 13,000 ***g*** at 4°C for 10 min. Samples were run in 4–20% Native PAGE gradient and stained with Coomassie Brilliant Blue to check for the presence of the protein in the feces. WB was performed as stated in the 'Immunoblotting' section to visualize the protein subunits using purified PdPV1 as control.

Neurotoxicity

To evaluate whether *P. diffusa* eggs contain neurotoxins as reported in other *Pomacea* species (Heras et al., 2008; Giglio et al., 2016), groups of five mice were intraperitoneally injected with either 200 µl of PBS or 4.6 mg kg^{-1} (200 µl) of PVF. Mice behavior was observed over 96 h and behavioral changes or neurological signs recorded. Special attention was given to those changes already reported for *Pomacea* neurotoxins including weakness and lethargy, tachypnea, hirsute hair, abduction of the rear limbs preventing them from supporting their own weight and spastic movements of tail muscles.

Glycan binding specificity using glycan array

Glycan binding specificity of PdPV1 was determined at the Core H of the Consortium for Functional Glycomics (http://www.functionalglycomics.org; Emory University, Atlanta, GA, USA). The protein was fluorescently labeled using the Alexa Fluor 488 Protein Labeling kit according to the manufacturer's instructions. Labeled PdPV1 was assayed on a glycan array consisting of 585 glycan targets (version 5.4), and the data were analyzed as described by Ituarte et al. (2018).

Hemagglutination assays

PdPV1 lectin activity was tested by hemagglutination of red blood cells (RBC) from rabbit, human, goat, horse, rat and chicken. Cells were obtained from the animal facilities of National University of La Plata (UNLP). Blood was taken by venous puncture and collected in 300 mmol l^{-1} sodic EDTA in sterile conditions. Human RBCs of type A, B and O from healthy donors were donated by the Institute of Hematology (Ministerio de Salud, Provincia de Buenos Aires, Argentina). Erythrocytes were prepared as stated in Ituarte et al. (2012). Twofold serial dilutions of PdPV1 in PBS (50 µl) were incubated with an equal volume of 2% (v/v) erythrocytes in PBS in U-shaped microtiter plates (Greiner Bio-One) at 37°C for 2 h. The initial protein concentration was 0.4 mg ml^{-1}. Results are expressed as the lowest protein concentration showing visible hemagglutinating activity by the naked eye.

PdPV1 carbohydrate affinity was inferred by comparing the inhibitory activity of the addition of fucose, *N*-acetyl glucosamine, *N*-acetyl galactosamine, glucosamine, galactosamine, mannose, galactose and glucose on hemagglutination. Twofold serial dilutions of purified protein starting from 0.4 mg ml^{-1} were made in 200 mmol l^{-1} of each carbohydrate and incubated for 1 h at 37°C. Later, 50 µl of 2% rabbit erythrocyte suspension in PBS were added to these solutions in a U-shaped microtiter plate, and after 2 h hemagglutination activity was verified. The inhibitory capacity (%) was expressed as the percent of inhibition, compared with the agglutination titer of PdPV1 without inhibitors: inhibitory capacity=$[100-(T_i \times 100)/T]$, where T is the hemagglutinating titer of the lectin without inhibitors, and T_i is the titer with inhibitors. In some experiments, RBCs were pre-treated with 0.1 mg ml^{-1} trypsin or 0.1 U ml^{-1} neuraminidase (New England Biolabs, Ipswich, MA, USA) before agglutinating analysis.

Statistical analysis

Data were analyzed by one-way ANOVA. When P-values were <0.05, the significance between groups was estimated by Tukey's test.

RESULTS

Structure

PdPV1 was purified and isolated by ultracentrifugation and HPLC. After NaBr gradient ultracentrifugation of the egg soluble fraction, two protein fractions were obtained: a colorless fraction of ∼1.22 g ml^{-1} and a colored fraction of ∼1.24 g ml^{-1} (Fig. 1A). When the colored fraction was subjected to SEC, two peaks were observed (Fig. 1B). The larger one was a carotenoprotein we named PdPV1, while the other was a PdPV1 aggregate, both peaks having the same apoprotein composition in SDS PAGE (Fig. 1B, inset). Native (non-denaturing) PAGE of PVF proteins of *P. diffusa* eggs (Fig. 1C) showed that PdPV1 is the most abundant, representing *ca* 60% of the total protein. The estimated molecular mass by SEC was 256 kDa, although the apparent mass by non-denaturing PAGE was ∼470 kDa (Fig. 1C). PdPV1 is composed of several subunits ranging from 23 to 33 kDa as determined by SDS-PAGE under reducing conditions (Fig. 1B). These subunits are not held together by disulphide bonds, as SDS-PAGE performed under non-reducing conditions indicates (not shown). The absorption spectrum of PdPV1 has the carotenoid-characteristic bands, with a single absorption maximum at 382 nm (Fig. 1D). However, unlike the other apple snail carotenoproteins, its intensity (and hence the egg coloration) is very weak. Total carbohydrate content of PdPV1 represents 8.9% (by weight), which probably resulted in an inaccurate molecular weight estimation by Native PAGE. The protein glycosylation pattern was determined by lectin dot-blot (Table 1). The absence of reactivity with PSA together with positive reactivity of ConA suggest the presence of tri-mannoside core glycans, while DBA, SBA and PNA positive reactivity indicates the presence of galactosides. PdPV1 has non-substituted T-antigen as suggested by PNA positive reactivity. The global shape of the particle in solution was determined by SAXS. A gyration radius (R_g) of 4.55 nm was estimated from the Guinier plot. PdPV1 is a globular protein as evidenced from the bell-shaped form of the curve in the Kratky plot (see SASBDB). A low-resolution 3D model for PdPV1 is depicted in Fig. 1E. SAXS raw data were deposited in the SASBDB repository (https://www.sasbdb.org/data; accession code SASDG75).

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

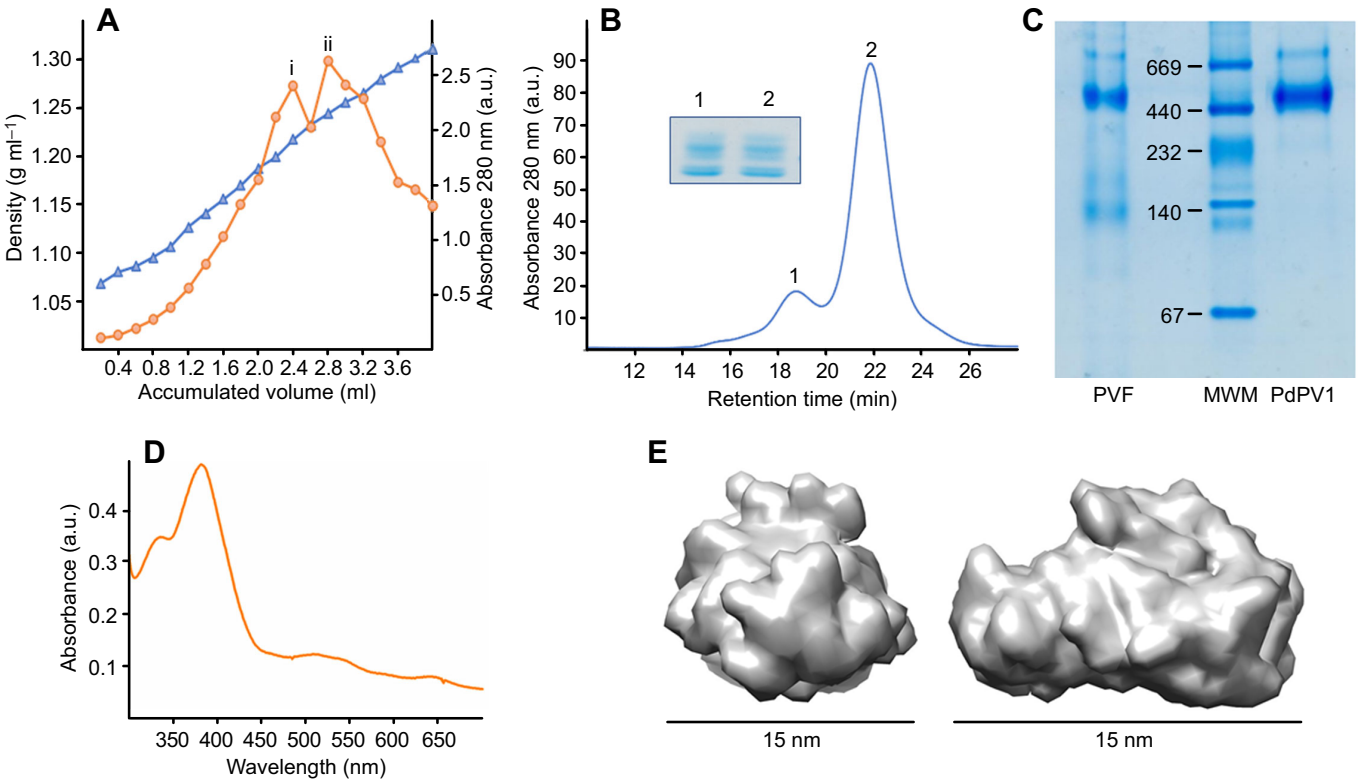


Fig. 1. Purification and structure of PdPV1. (A) Protein (orange circles) and hydration density (blue triangles) profile of *P. diffusa* egg soluble fraction; (i) uncolored and (ii) colored fraction. (B) Size exclusion chromatogram of colored (ii) fraction: (1) aggregate and (2) PdPV1. Inset: SDS-PAGE of 1 and 2 peaks showing the same apoprotein composition. (C) Native PAGE of the egg soluble fraction (PVF) and purified PdPV1. MWM, molecular weight markers. (D) Absorption spectrum of purified PdPV1. (E) *Ab initio* model of SAXS data of PdPV1 by DAMMIF (representative of best model cluster).

The internal sequences, together with three N-terminal amino acid sequences (Table S1), were used to identify six full-length subunits (Fig. S1A) in the *P. diffusa* albumen gland transcriptome (Ip et al., 2018), hereafter referred to as PdPV1-1 to -6. All six amino acid sequences have a signal peptide and a theoretical molecular mass around 21 kDa, in agreement with the molecular weight estimation by SDS-PAGE. All subunits have at least one potential phosphorylation site as follows: PdPV1-1 has four Tyr predicted to be phosphorylated; PdPV1-2 has six predicted sites (one Ser, two Thr, three Tyr); PdPV1-3 has two Thr and one Tyr; PdPV1-4 has one Ser and one Tyr; PdPV1-5 has two Ser; and PdPV1-6 has one Thr and two Tyr. Regarding potential glycosylation sites on PdPV1, although the six subunits present at least one N-glycosylation site, predicted O-glycosylation sites are only present in PdPV1-4, PdPV1-5 and PdPV1-6 (Fig. S1A). The SEG server indicates the absence of low-complexity regions in these sequences. Sequence alignment of the six PdPV1 subunits is shown in Fig. S1B.

Table 1. Acronyms and specificity of lectins used in PdPV1 oligosaccharide recognition

Acronym	Binding specificity	PdPV1
ConA	Man α3[Man α6] man; terminal αMan; oligomannose and hybrid type *N*-glycans	+
WGA	Neu5Acα3/6/8 R	+
PSA	*N*-acetylchitobiose-linked fucose	–
PNA	Terminal Gal β13-GalNAc	+
JAC	GalNAc	+
RCA-1	Terminal βGal	+
SBA	Tn antigen; terminal α GalNAc	+
DBA	GalNAc α1,3 GalNAc/Gal	+
UEA	α(1,2) L-Fucose	+
PVL	GlcNAc Neu5Ac α26-Gal	–
LCA	Bi- and triantennary complex-type *N*-glycan	+
ECL	Terminal LacNAc	–
BSL-1	αGalNAc; αGal	+

Major specificities (Goldstein and Hayes, 1978). Gal, galactose; GalNAc, *N*-acetyl galactosamine; Glc, glucose; GlcNAc, *N*-acetyl glucosamine; Man, Mannose; Neu5Ac, acetyl neuraminic acid (sialic acid).

An NCBI nr database search revealed no homology with any known sequences other than those of *Pomacea* perivitellins. Using PsiBLAST search revealed homologs only within the genus; no conserved domains were identified by PROSITE. In addition, no sequence similarity with any known lectin could be found for any of the six PdPV1 subunits, except for PsSC lectin from the sister species *P. scalaris*.

Evolutionary relationships among *Pomacea* perivitellins

A phylogenetic analysis performed including all *Pomacea* perivitellin subunit sequences (Fig. 2A) indicates that PdPV1 subunits did not group together, but each one grouped with those of the other species into six separate clusters. Each orthologous sequence cluster among PdPV1, PsSC, PcOvo and PmPV1 was separated into two subgroups that corresponded to *bridgesii* and *canaliculata* clades (Fig. 2B). Comparison of amino acid sequences within each ortholog cluster indicated high similarity. However, among the six PdPV1 subunits, the similarity between any two subunits was significantly lower. The immune cross-reactivity among PdPV1, PcOvo and PsSC was evaluated by WB assays. Anti-PsSC polyclonal antibodies strongly recognized each of the six

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

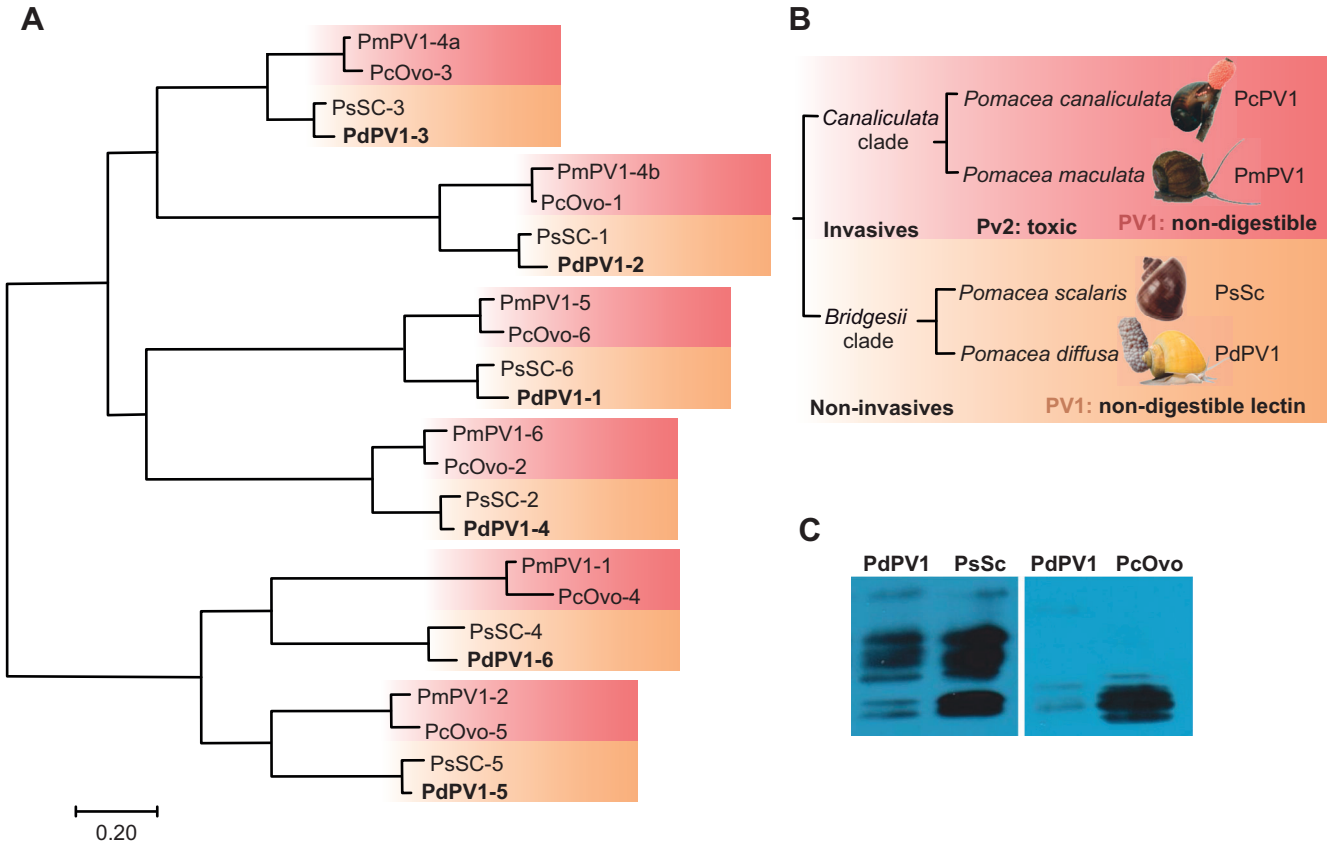


Fig. 2. Phylogenies of PV1 perivitellin subunits versus morphological phylogenies of *Pomacea canaliculata* and *bridgesii* clades. (A) Maximum-likelihood phylogram obtained with RAxML under the best partition model for *Pomacea* 'PV1' perivitellins. Proteins from this study are given in bold. (B) Phylogeny of *P. canaliculata* and *bridgesii* clades (after Hayes et al., 2009) indicating species, invasiveness and some of the perivitelline defenses present in each. (C) Cross-reactivity of PdPV1 with polyclonal antibodies anti-PsSC (left panel) and anti-PcOvo (right panel).

PdPV1 subunits, whereas anti-PcOvo polyclonal antibodies weakly cross-reacted with only a few PdPV1subunits (Fig. 2C).

Lectin activity

PdPV1 showed hemagglutination activity against rabbit, rat and human A and B RBCs up to 0.025, 0.05, 0.05 and 0.05 mg ml^{-1}, respectively, and non-detectable activity towards RBCs from human O and from other species. This activity was largely inhibited by *N*-acetyl galactosamine (approximately 75% inhibition), galactosamine and galactose, while pre-incubation of rabbit RBCs with neuraminidase did not block hemagglutination. Glycan specificity and relative affinity were further evaluated by a glycan array screening of PdPV1. Out of 523 unique glycan structures, we observed significant binding of PdPV1 to a range of structures that included 2,8-linked *N*-acetylneuraminic acid, a common sialic acid in vertebrate ganglioside motifs. In fact, PdPV1 recognizes GD1b, GT2, GD2, GT1b and GD1a gangliosides with high affinity. Interestingly, PdPV1 higher affinity scores were obtained for sialylated structures that featured a GalNAcb1-4Gal motif. PdPV1 affinity for GalNAc-Gal would explain the presence of aggregated PdPV1 lectin observed in the chromatographic analysis (Fig. 1B) because these glycans, which predominate in the glycan decoration of PdPV1 (Table 1), would promote self-aggregation. Very few examples of PdPV1 binding to complex N-linked glycans were observed. Glycans with higher scores are listed in Table 2.

Table 2. Main glycan structures recognized by PdPV1

	Oligosaccharide	Mean RFU	%CV
1	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	47,094±3961	8
2	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	31,842±3388	11
3	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	26,833±2216	8
4	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	24,275±2348	10
5	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	21,408±1403	7
6	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	12,349±185	1
7	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	12,313±1317	11
8	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	9642±871	9
9	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	9360±1641	18
10	GalNAcb1-3Galb-Sp8	6334±2003	32

RFU, relative fluorescence units. Results are expressed as the mean RFU±1 s.d., *N*=4. %CV=100×s.d. The complete glycan array data are available at: https://www.functionalglycomics.org/static/index.shtml (glycan array 3546).

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

Neurotoxicity

Intraperitoneal injection of PVF had no effect on mouse survival. Moreover, after 96 h mice did not show either neurological signs or behavioral changes.

Structural stability

Structural stability of PdPV1 at different pH values was monitored by fluorescence and absorption spectroscopy and by SAXS. Absorption and fluorescence spectroscopy showed that only at pH 2.0 did the protein partially alter its conformation, as seen by an intensity decrease in the carotenoid absorbing (Fig. 3A) and a blue shift in the fourth derivate of the spectra (Fig. 3B). No changes in tryptophan fluorescence emission spectra were recorded (Fig. 3C), indicating that the protein remains properly folded up to pH 2.0, where a slight red shift of the emission maximum was observed. However, changes at pH 2.0 might be considered as small changes in protein structure because SAXS experiments showed that the particle shape remains globular in the pH 2.0–12.0 range (Fig. 3D) with no changes of its R_g except for a small reduction at pH 12.0 (Fig. 3E). Regarding thermal stability, no structural perturbations in the conditions assayed could be detected in PdPV1 either by absorption (Fig. 4A,B) or fluorescence spectroscopy (Fig. 4C),

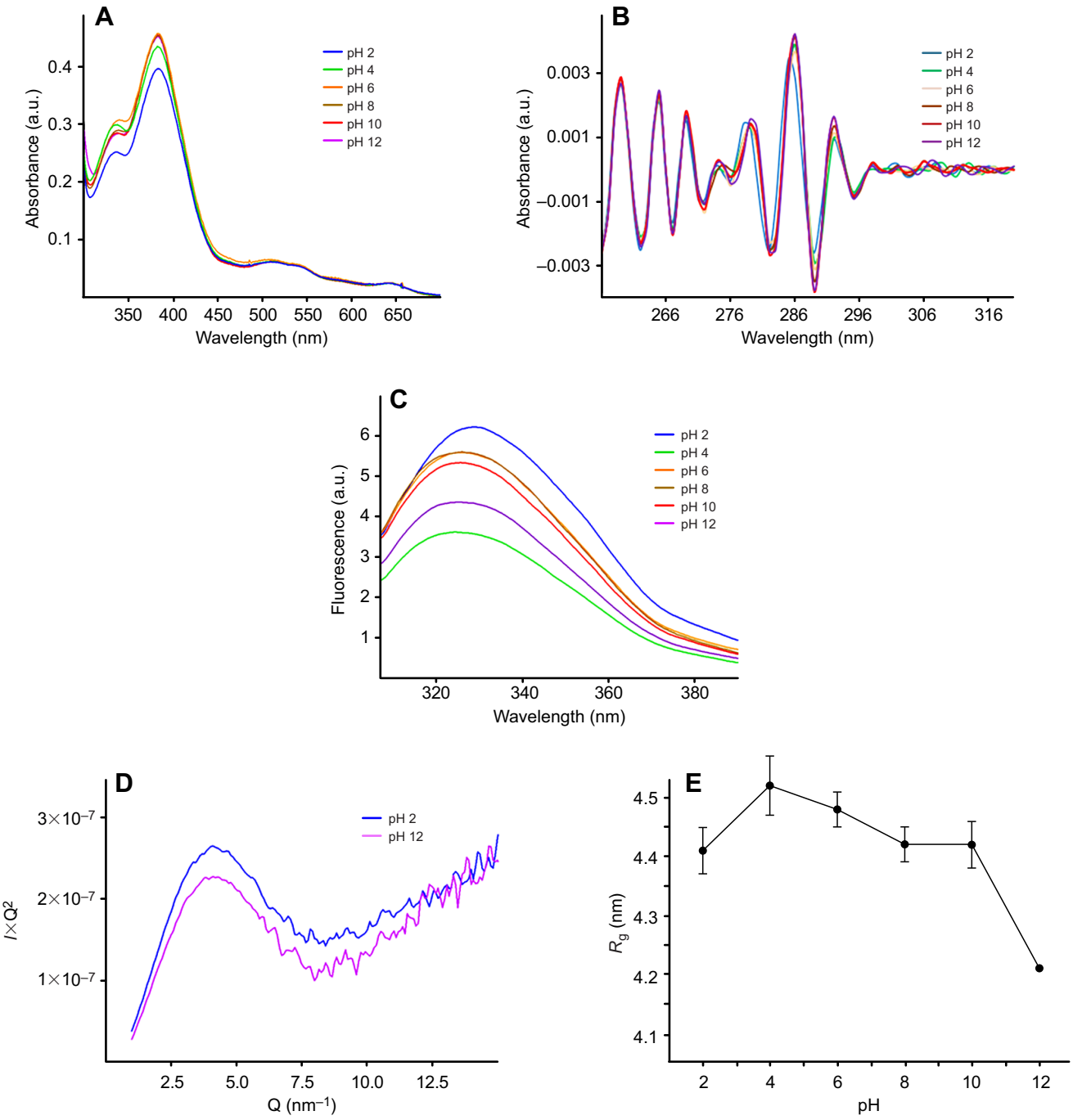


Fig. 3. Effect of pH on PdPV1 structural stability. (A) Absorption spectra. (B) Fourth derivative absorption spectra. (C) Intrinsic fluorescence emission spectra and (D) Kratky plots of PdPV1 at pH 2.0 (blue) and pH 12.0 (violet). (E) Gyration radii (R_g) at different pH values.

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

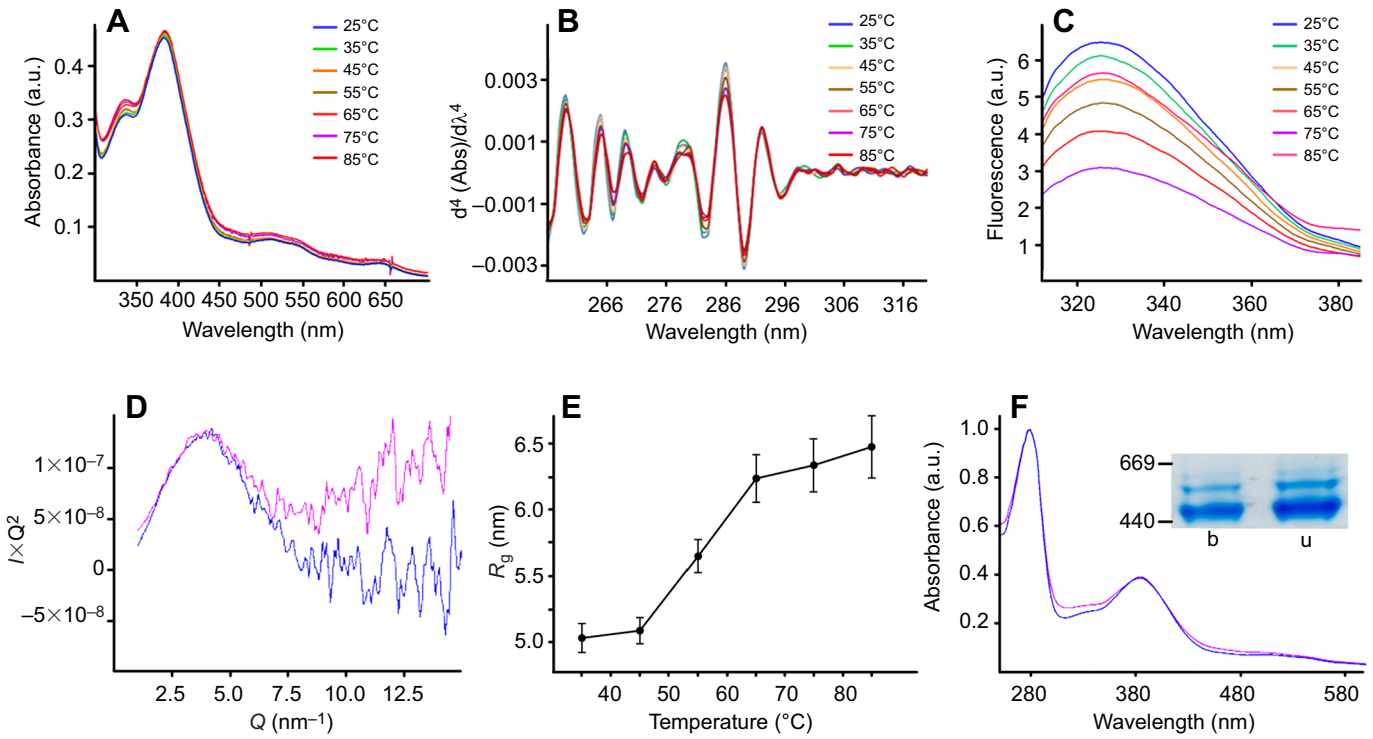


Fig. 4. Effect of temperature on PdPV1 structural stability. (A) Absorption spectra. (B) Fourth derivative absorption spectra. (C) Intrinsic fluorescence emission spectra. (D) Kratky plots obtained from SAXS data at 25°C (blue) and 85°C (violet). (E) Gyration radii (R_g) at different temperatures. (F) Absorption spectra of unboiled (blue) and boiled (violet) PdPV1; inset: Native 4–20% PAGE; lane b: PdPV1 boiled for 15 min; lane u: PdPV1 unboiled.

indicating protein stability up to 85°C. These results are supported by SAXS, although at 85°C a partial loss of protein globularity occurs (Fig. 4D), associated with a slight R_g increase (Fig. 4E). Moreover, extreme thermal treatment of PdPV1 (boiled for 15 min) did not affect its oligomer integrity as seen by Native PAGE (Fig. 4F, inset).

Chemical stability of PdPV1 was studied by equilibrium unfolding experiments. The protein unfolded fraction versus [GndHCl] is depicted in Fig. 5A, showing cooperative unfolding transitions. The unfolding process fits to a two-state model with the protein completely unfolded from 4.5 mol l^{-1} GnHCl onwards. The calculated standard free energy for the unfolding process was 4.76 kJ mol^{-1}, and the m value, which characterized the influence of the chaotrope on the process, was 1416.76 kJ mol^{-1} (mol l^{-1})$^{-1}$. Finally, the midpoint unfolding GndHCl concentration for PdPV1 was 3.4 mol l^{-1}. Kinetically stable proteins show no alteration in their electrophoretic mobility when incubated with SDS; they need to be heated to be disassembled into their subunits. In this regard, PdPV1 did not change its migration pattern after incubation with SDS. Only when PdPV1 was boiled in the presence of SDS did the oligomer disassemble into its six subunits, as shown by SDS PAGE (Fig. 5B).

The particle was also very resistant to enzymatic digestion, withstanding a simulated gastrointestinal digestion for 2 h. The integrity of the protein after gastric and duodenal phases was analyzed by SDS-PAGE, showing no significant alterations in its electrophoretic migration (Fig. 6A). Furthermore, when the protein was administered to mice, the protein was recovered almost intact in the feces as shown by Native PAGE and WB analysis, indicating that PdPV1 can pass through the digestive system without altering its structure (Fig. 6B,C).

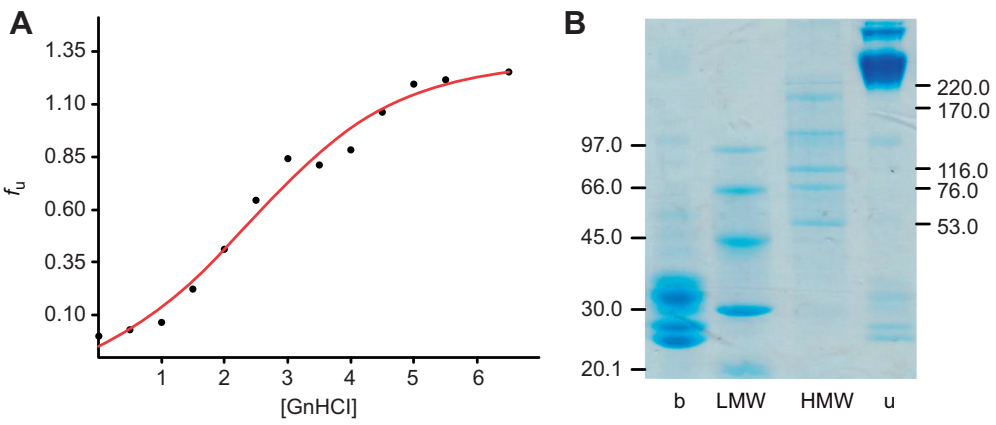


Fig. 5. Structural stability of PdPV1. (A) Chemical stability evaluated by the unfolding induced by GnHCl. f_u, protein unfolded fraction. (B) Kinetic stability evaluated with an SDS-resistance assay. The same PdPV1 sample was unheated (u) or boiled (b) in the presence of SDS for 10 min immediately prior to being loaded onto the gel. LMW and HMW, low and high molecular weight markers, respectively.

Journal of Experimental Biology (2020) 223, jeb231878. doi:10.1242/jeb.231878

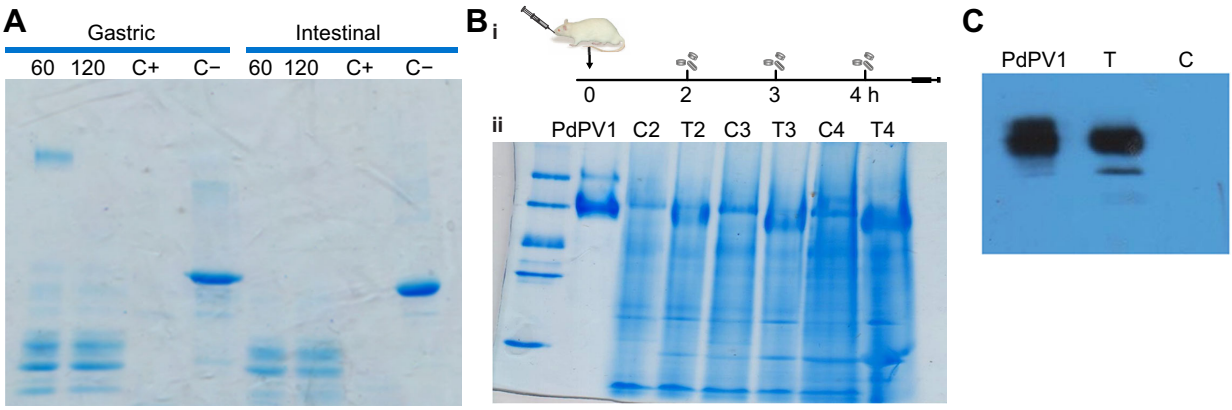


Fig. 6. *In vitro* and *in vivo* digestibility of PdPV1. (A) Simulated gastrointestinal digestion analyzed by SDS-PAGE. Gastric phase: PdPV1 after incubation with pepsin in SGF for 60 and 120 min; C+ and C–, positive and negative controls of BSA digestion with and without pepsin, respectively. Intestinal phase: PdPV1 incubated with trypsin for 60 and 120 min; C+ and C–, positive and negative controls of BSA digestion with (C+) and without (C–) addition of trypsin. (B) *In vivo* digestibility of PmPV1; (i) diagram of experimental protocol showing oral administration of PmPV1 and feces collection times; (ii) analysis of fecal proteins by Native PAGE. PdPV1: purified PdPV1; C2, C3, C4 and T2, T3 and T4: control and feces collected after 2, 3 and 4 h, respectively. (C) Immune detection by western blot of PdPV1 in feces collected after 4 h mice were gavaged with buffer (C) or with 0.6 mg PdPV1 (T).

DISCUSSION

During Ampullariidae transition from water to land, *Pomacea* eggs faced the harsh environmental conditions of the terrestrial environment, acquiring novel defenses, including highly stable, non-digestible oligomeric perivitellins that became the most abundant proteins in the PVF (Dreon et al., 2003, 2010; Ituarte et al., 2008, 2018; Pasquevich et al., 2014, 2017; Ip et al., 2019, 2020). In this study we isolated and characterized the main perivitellin of *P. diffusa* eggs, which unveiled some evolutionary trends within the genus. We found that, as in other *Pomacea*, PdPV1 is also a highly stable carotenoprotein that remains completely folded even at high temperatures, extreme pH values and high concentrations of chaotropes. Furthermore, this remarkable structural stability allows PdPV1 to be undigestible, withstanding the passage through the gastrointestinal tract if ingested by mice, indicating that it could reach the digestive epithelium in a fully active conformation.

PdPV1 is composed of six polypeptide chains similar to its orthologs PcOvo, PsSC and PmPV1 (Pasquevich et al., 2017; Ituarte et al., 2018, 2019) with only moderate sequence similarities among them. It has been suggested that they were generated by an ancient gene duplication (Sun et al., 2012). Phylogenetic analysis of the four perivitellins shows each PdPV1 subunit grouped separately in one of six different clades, each containing the corresponding orthologous subunit of each species perivitellin. In addition, each orthologous group presents low sequence divergence, reinforcing the suggestions by Ituarte et al. (2018) that after gene duplication the advantage of having multiple copies of this gene may enhance the perivitelline synthesis during egg production, thus preventing additional gene variations. Moreover, the high expression level of these genes, almost exclusively restricted to the albumen gland of *P. diffusa* (Ip et al., 2020), further supports this hypothesis. The perivitellin family tree agreed with those based on morphology and MT COI genetic analyses (Hayes et al., 2015), as the subunits belonging to PsSC and PdPV1 (*bridgesii* clade) and PcOvo and PmPV1 (*canaliculata* clade) grouped separately. This was further shown using antibodies: while *bridgesii* clade perivitellins are recognized by an anti-PsSC polyclonal antibody, anti-PcOvo antibodies reacted weakly. These results provide the first evidence of structural differences between perivitellins from *bridgesii* and *canaliculata* clades, probably associated but not restricted to their different glycosylation patterns. In this regard, PdPV1 glycosylation pattern is quite similar to PsSC (Ituarte et al., 2010) but differs from those of perivitellins from the *canaliculata* clade.

For evolutionary trends to be distinguished, enough species within a group must have been studied. Thus, the characterization of PdPV1 disclose characteristics common to all apple snail carotenoproteins. They are all high molecular weight homologous oligomers formed by six subunits of 25–35 kDa each, with similar amino acid sequences. As anti-predator molecules, they can be regarded as anti-nutritive proteins that lower the nutritional value of eggs as kinetically stable and non-digestible storage proteins that survive the passage through the gut of predators unaffected (Pasquevich et al., 2017). However, some clade-specific differences also arise: the *bridgesii* clade with non-invasive species (*P. scalaris* and *P. diffusa*) features carotenoproteins with lectin activity that are absent in the ortholog perivitellins of the *canaliculata* clade that include the invasive species *P. canaliculata* and *P. maculata*, highlighting a clade-related association between invasiveness and type of egg defense. Assuming that PdPV1 and PsSC have similar roles, we can say they have a dual defensive function: as lectin altering predator gut morphophysiology and as non-digestible protein lowering the nutritional value of eggs. In this regard, both PdPV1 and PsSC lectin activity (Ituarte et al., 2018) seem directed towards ganglioside-containing structures. The similarities between PdPV1 and PsSC in their lectin activity and primary structure, together with the lack of homology with any other lectin, led us to speculate that these perivitellins have a novel carbohydrate recognition module and could probably be a novel lectin family, although this needs further structural analysis. Lectins are essential humoral effectors in the innate immune system and play a defensive role against microorganisms (Malham et al., 2002), a possible role for these lectins as previously suggested for PsSC (Ituarte et al., 2012). After ingestion, PsSC binds to enterocytes, inducing a modification of its glycosylation pattern and eventually altering the normal morphophysiology of the rodent digestive tract, thus having adverse nutritional consequences in potential egg consumers (Ituarte et al., 2018). Considering that the action of lectins is predominantly triggered after recognizing their specific glycan ligands (Tateno, 2014) and that both PsSC and PdPV1 bind to similar gangliosides, it is expected that binding of *Pomacea* PV1 lectins to cell surface glycolipids may be the first step of their noxious action on gut, a well-known defensive strategy in plants

(Peumans and Van Damme, 1995). To the best of our knowledge, the massive accumulation in eggs of a hyperstable, non-digestible lectin, combined with protease inhibitors and its ability to adversely affect gut morphophysiology, is unique among animal defensive systems against predation.

The high structural stability of PdPV1 together with its specificity for gangliosides usually over-expressed in certain types of tumors, opens a new research avenue for the study of PdPV1 potential biomedical applications as a tumor marker.

Conclusions

This study contributes to our knowledge on the reproductive biology of snails. Particularly, it revealed clade-specific defensive strategies in poisonous *Pomacea* eggs, some with carotenoproteins with lectin activity and others with neurotoxins as the major effectors. The occurrence of different defensive strategies is related to the species' phylogenetic position, and to their invasiveness. By contrast, the presence of highly stable, non-digestible storage carotenoproteins massively accumulated in the egg emerges as a common characteristic of the whole genus *Pomacea*. Another common feature is that PV1s provide eggs with coloration, a presumed aposematic signal to deter predators from ingesting their noxious eggs. More comparative work on snail perivitellins from these and other clades is still needed to fully appreciate the unique defenses of ampulllariids and their evolutionary trends.

Acknowledgements

M.S.D. is member of Carrera del Investigador Científico CICBA, Argentina. H.H. is a member of Carrera del Investigador Científico CONICET, Argentina. T.R.B. is a PhD fellow from CONICET, Argentina. We wish to acknowledge the Protein-glycan Interaction Resource of the CFG and the National Center for Functional Glycomics (NCFG) at Beth Israel Deaconess Medical Center, Harvard Medical School. We thank LNLS - Brazilian Synchrotron Light Laboratory for access to their facilities and partial financial support. We also thank Mario Raúl Ramos for his help with figure design.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: T.R.B., M.S.D., H.H.; Methodology: T.R.B., M.S.D., J.-W.Q., H.H.; Formal analysis: T.R.B., M.S.D., J.-W.Q., H.H.; Investigation: T.R.B., M.S.D., H.H.; Resources: M.S.D., H.H.; Data curation: M.S.D., J.-W.Q.; Writing - original draft: T.R.B., M.S.D., J.-W.Q., H.H.; Visualization: H.H.; Supervision: M.S.D., H.H.; Project administration: H.H.; Funding acquisition: M.S.D., H.H.

Funding

This work was supported by grants from the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICBA to M.S.D.) and Agencia Nacional de Promoción Científica y Técnica (PICT 2014-0850 and 2015-0661 to H.H. and M.S.D., respectively), Argentina. Financial support was also provided by LNLS - Brazilian Synchrotron Light Laboratory (projects SAXS1-17746 and SAXS2-18846).

Data availability

The datasets obtained from SAXS experiments are available in the SASBDB repository (https://www.sasbdb.org/data/SASDG75/)

Supplementary information

Supplementary information available online at
https://jeb.biologists.org/lookup/doi/10.1242/jeb.231878.supplemental

References

Baldo, B. A. and Uhlenbruck, G. (1974). Studies on the agglutinin specificities and blood group O(H)-like activities in extracts from the Molluscs *Pomacea paludosa* and *Pomacea urceus*. *Vox Sang.* **27**, 67-80. doi:10.1159/000466721

Bourne, R. and Bourne, R. (2010). ImageJ. *Fund. Dig. Imag. Med.* **9**, 185-188. doi:10.1007/978-1-84882-087-6_9

Cadierno, M. P., Dreon, M. S. and Heras, H. (2018b). Validation by qPCR of reference genes for reproductive studies in the invasive apple snail *Pomacea canaliculata*. *Malacologia* **62**, 163-170. doi:10.4002/040.062.0105

Cadierno, M. P., Saveanu, L., Dreon, M. S., Martín, P. R. and Heras, H. (2018a). Biosynthesis in the albumen gland-capsule gland complex limits reproductive effort in the invasive apple snail *Pomacea canaliculata*. *Biol. Bull.* **235**, 1-11. doi:10.1086/699200

Dreon, M. S., Ceolín, M. and Heras, H. (2007). Astaxanthin binding and structural stability of the apple snail carotenoprotein ovorubin. *Arch. Biochem. Biophys.* **460**, 107-112. doi:10.1016/j.abb.2006.12.033

Dreon, M. S., Frassa, M. V., Ceolin, M., Ituarte, S., Qiu, J.-W., Sun, J., Fernandez, P. E. and Heras, H. (2013). Novel animal defenses against predation: a snail egg neurotoxin combining lectin and pore-forming chains that resembles plant defense and bacteria attack toxins. *PLoS ONE* **8**, e63782. doi:10.1371/journal.pone.0063782

Dreon, M. S., Heras, H. and Pollero, R. J. (2003). Metabolism of ovorubin, the major egg lipoprotein from the apple snail. *Mol. Cell. Biochem.* **243**, 9-14. doi:10.1023/A:1021616610241

Dreon, M. S., Ituarte, S. and Heras, H. (2010). The role of the proteinase inhibitor ovorubin in apple snail eggs resembles plant embryo defense against predation. *PLoS ONE* **5**, e15059. doi:10.1371/journal.pone.0015059

Dreon, M. S., Ituarte, S., Ceolin, M. and Heras, H. (2008). Global shape and pH stability of ovorubin, an oligomeric protein from the eggs of *Pomacea canaliculata*. *FEBS J.* **275**, 4522-4530. doi:10.1111/j.1742-4658.2008.06595.x

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. and Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Anal. Chem.* **28**, 350-356. doi:10.1021/ac60111a017

Giglio, M. L., Ituarte, S., Pasquevich, M. Y. and Heras, H. (2016). The eggs of the apple snail *Pomacea maculata* are defended by indigestible polysaccharides and toxic proteins. *Can. J. Zool.* **94**, 777-785. doi:10.1139/cjz-2016-0049

Goldstein, I. and Hayes, C. (1978). The lectins: carbohydrate binding proteins of plants and animals. *Adv. Carbohydr. Chem. Biochem.* **35**, 127-340. doi:10.1016/S0065-2318(08)60220-6

Hammersley, A. P. (2016). FIT2D: a multi-purpose data reduction, analysis and visualization program. *J. Appl. Crystallogr.* **49**, 646-652. doi:10.1107/S1600576716000455

Hayes, K. A., Burks, R. L., Castro-Vazquez, A., Darby, P. C., Heras, H., Martin, P. R., Qiu, I.-W., Thiengo, S. C., Vega, I. A., Wada, T. et al. (2015). Insights from an integrated view of the biology of apple snails (Caenogastropoda: Ampullariidae). *Malacologia* **58**, 245-302. doi:10.4002/040.058.0209

Hayes, K. A., Cowie, R. H. and Thiengo, S. C. (2009). A global phylogeny of apple snails: gondwanan origin, generic relationships, and the influence of outgroup choice (Caenogastropoda: Ampullariidae). *Biol. J. Linn. Soc.* **98**, 61-76. doi:10.1111/j.1095-8312.2009.01246.x

Heras, H., Frassa, M. V., Fernández, P. E., Galosi, C. M., Gimeno, E. J. and Dreon, M. S. (2008). First egg protein with a neurotoxic effect on mice. *Toxicon* **52**, 481-488. doi:10.1016/j.toxicon.2008.06.022

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradivic, Z. and Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037-1049. doi:10.1093/nar/gkh253

Ip, J. C. H., Mu, H., Chen, Q., Sun, J., Ituarte, S., Heras, H., Van Bocxlaer, B., Ganmanee, M., Huang, X. and Qiu, J.-W. (2018). AmpuBase: a transcriptome database for eight species of apple snails (Gastropoda: Ampullariidae). *Gigascience* **19**, 179. doi:10.1186/s12864-018-4553-9

Ip, J. C. H., Mu, H., Zhang, Y., Heras, H. and Qiu, J.-W. (2020). Egg perivitelline fluid proteome of a freshwater snail: insight into the transition from aquatic to terrestrial egg deposition. *Rapid Commun. Mass Spectrom.* **34**, e8605. doi:10.1002/rcm.8605

Ip, J. C. H., Mu, H., Zhang, Y., Sun, J., Heras, H., Chu, K. H. and Qiu, J.-W. (2019). Understanding the transition from water to land: insights from multi-omic analyses of the perivitelline fluid of apple snail eggs. *J. Proteomics* **194**, 79-88. doi:10.1016/j.jprot.2018.12.014

Ituarte, S., Brola, T. R., Dreon, M. S., Sun, J., Qiu, J.-W. and Heras, H. (2019). Non-digestible proteins and protease inhibitors: implications for defense of the colored eggs of the freshwater apple snail *Pomacea canaliculata*. *Can. J. Zool.* **566**, 558-566. doi:10.1139/cjz-2018-0210

Ituarte, S., Brola, T. R., Fernández, P. E., Mu, H., Qiu, J.-W., Heras, H. and Dreon, M. S. (2018). A lectin of a non-invasive apple snail as an egg defense against predation alters the rat gut morphophysiology. *PLoS ONE* **13**, e0198361. doi:10.1371/journal.pone.0198361

Ituarte, S., Dreon, M. S., Ceolín, M. and Heras, H. (2008). Isolation and characterization of a novel perivitellin from the eggs of *Pomacea scalaris* (Mollusca, Ampullariidae). *Mol. Reprod. Dev.* **75**, 1441-1448. doi:10.1002/mrd.20880

Ituarte, S., Dreon, M. S., Ceolin, M. and Heras, H. (2012). Agglutinating activity and structural characterization of scalarin, the major egg protein of the snail *Pomacea scalaris* (d'Orbigny, 1832). *PLoS ONE* **7**, e50115. doi:10.1371/journal.pone.0050115

Ituarte, S., Dreon, M. S., Pasquevich, M. Y., Fernández, P. E. and Heras, H. (2010). Carbohydrates and glycoforms of the major egg perivitellins from *Pomacea* apple snails (Architaenioglossa: Ampullariidae). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **157**, 66-72. doi:10.1016/j.cbpb.2010.05.004

Jayaprakash, N. G. and Surolia, A. (2017). Role of glycosylation in nucleating protein folding and stability. *Biochem. J.* **474**, 2333-2347. doi:10.1042/BCJ20170111

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. and Battistuzzi, F. U. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549. doi:10.1093/molbev/msy096

Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem* **193**, 265-275.

Malham, S. K., Lacoste, A., Gélébart, F., Cueff, A. and Poulet, S. A. (2002). A first insight into stress-induced neuroendocrine and immune changes in the octopus *Eledone cirrhosa*. *Aquat. Living Resour.* **15**, 187-192. doi:10.1016/S0990-7440(02)01173-7

Manning, M. and Colón, W. (2004). Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. *Biochemistry* **43**, 11248-11254. doi:10.1021/bi0491898

Matsukura, K., Okuda, M., Cazzaniga, N. J. and Wada, T. (2013). Genetic exchange between two freshwater apple snails, *Pomacea canaliculata* and *Pomacea maculata* invading East and Southeast Asia. *Biol. Invasions* **15**, 2039-2048. doi:10.1007/s10530-013-0431-1

Merril, C. R. (1990). Gel-staining techniques. *Methods Enzymol.* **182**, 477-488. doi:10.1016/0076-6879(90)82038-4

Orr, J., Adamson, G. and Lindgren, F. (1991). Preparative ultracentrifugation and analytic ultracentrifugation of plasma lipoproteins. In *Analyses of Fats, Oils and Lipoproteins* (ed. E. Perkins), pp. 525-554. Champaign, Illinois: AOCS press.

Pasquevich, M. Y., Dreon, M. S. and Heras, H. (2014). The major egg reserve protein from the invasive apple snail *Pomacea maculata* is a complex carotenoprotein related to those of *Pomacea canaliculata* and *Pomacea scalaris*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **169**, 63-71. doi:10.1016/j.cbpb.2013.11.008

Pasquevich, M. Y., Dreon, M. S., Qiu, J.-W., Mu, H. and Heras, H. (2017). Convergent evolution of plant and animal embryo defences by hyperstable non-digestible storage proteins. *Sci. Rep.* **7**, 15848. doi:10.1038/s41598-017-16185-9

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. and Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr. Int. Union Crystallogr.* **45**, 342-350. doi:10.1107/S0021889812007662

Peumans, W. J. and Van Damme, E. J. M. (1995). Lectins as plant defense proteins. *Plant Physiol.* **109**, 347-352. doi:10.1104/pp.109.2.347

Prokop, O., Uhlenbruck, G. and Kohler, W. (1968). A new source of antibody-like substances having anti-blood group specificity: a discussion on the specificity of helix agglutinins. *Vox Sang.* **14**, 321-333. doi:10.1111/j.1423-0410.1968.tb01722.x

Przeslawski, R. (2004). A review of the effects of environmental stress on embryonic development within intertidal gastropod egg masses. *Moll. Res.* **24**, 43-63. doi:10.1071/MR04001

Sanchez, J.-F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A. and Mitchell, E. P. (2006). Biochemical and structural analysis of *Helix pomatia* agglutinin. A hexameric lectin with a novel fold. *J. Biol. Chem.* **281**, 20171-20180. doi:10.1074/jbc.M603452200

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L. et al. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478-1488. doi:10.1038/emboj.2013.79

Sun, J., Zhang, H., Wang, H., Heras, H., Dreon, M. S., Ituarte, S., Ravasi, T., Qian, P.-Y. and Qiu, J.-W. (2012). First proteome of the egg perivitelline fluid of a freshwater gastropod with aerial oviposition. *J. Proteome Res.* **11**, 4240-4248. doi:10.1021/pr3003613

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886. doi:10.1016/S0006-3495(99)77443-6

Tateno, H. (2014). Evaluation of glycan-binding specificity by glycoconjugate microarray with an evanescent-field fluorescence detection system. In *Lectins, Methods in Molecular Biology (Methods and Protocols)*, Vol. 1200 (ed. J. Hirabayashi), pp. 353-359. New York: Humana Press. doi:10.1007/978-1-4939-1292-6_30

Uhlenbruck, G., Steinhausen, G. and Cheesman, D. F. (1973). An incomplete anti-B agglutinin in the eggs of the prosobranch snail *Pila ovata*. *Experientia* **29**, 1139-1140. doi:10.1007/BF01946768

Yusa, Y. (2001). Predation on eggs of the apple snail *Pomacea canaliculata* (Gastropoda: Ampullariidae) by the fire ant *Solenopsis geminata*. *J. Mollusc. Stud.* **67**, 275-279. doi:10.1093/mollus/67.3.275

Yusa, Y., Sugiura, N. and Ichinose, K. (2000). Predation on the apple snail, *Pomacea canaliculata* (Ampullariidae), by the Norway rat, *Rattus norvegicus*, in the field. *Veliger* **43**, 349-353.

Table S1: **Peptide sequences employed to identify PdPV1 subunits.**

PdPV1-1 (PV1-A1) Pdi67619_c0_g1	**PdPV1-2** (PV1-A2) Pdi37843_c0_g1	**PdPV1-3** (PV1-A3) Pdi14773_c0_g1	**PdPV1-4** (PV1-A4) Pdi37785_c0_g1	**PdPV1-5** (PV1-B1) Pdi118877_c0_g1	**PdPV1-6** (PV1-B3) Pdi33234_c0_g1
.-	**XDEY(G)IMEINPPXXVSXQ**	**(S)QEYLLLDIIEASTQ**	.-	**EXXMHDLVV**	
TFIDSFPNK	VSVTPAEDMVDTMQK	SKPTPLQDMADLMEEIGVGWPK	SPQEIRDEMKDTEVLYSFK	QYDSEFK	MNLKDVNEITPVIQLEER
SVEDYKDMLKEVDLVYSFK	VSVTPAEDMVDTMQK	QNLDDASNTGIDVLGGPGNAR	SPQEIRDEMKDTEVLYSFK	NKISFVQILR	GQTLQQYEAELK
SVEDYKDMLK	VSVTPAEDMVDTMQK	IYFFINIPR	YYIIYEVKDSEK	YLAVVREPFLTR	
SVEDYKDMLK	VSEQEILNMLSPLEIKHK	HLIVVKQNDLNLEK	MYIYINIPLDEVDDFASK	VYISSGFPPK	
SKSVEDYKDMLK	EGVDIWGGPGEYHVKPEYLIR	ANGAYPPK	MVPVVDMFDLAK	TLTLIKYTPSLK	
TLKFEGEPTQGAISHVQAYGEKIK	ATGVTPHKHFFFVSSAR	TSVQYLTK	TKMVPVVDMFDLAK	TLKQYDSEFKR	
TLKFEGEPTQGAISHVQAYGEK	KVSEQEILNMLSPLEIKHK	TMEQYMSEK	TLPLEGLTTQQSR	TLKQYDSEFK	
FEGEPTQGAISHVQAYGEK	HFFFVSSARDEVEVIGR	TMEQYMSEK	MYIYINIPLDEVDDFASK	GLFVKELGTSK	
EVDLVYSFKVVGSSR	LDPVGYMK		GFMEEIKDSFQK	CLNTALKDYPK	
EVDLVYSFK	HFFFVSSAR		GFMEEIKDSFQK	GLFVKELGTSKYLAVVR	
INLPGDVISIAVGDLNDK			DEMKDTEVLYSFK	LEISVPPSK	
INLPGDVISIAVGDLNDKMR			DEMKDTEVLYSFK	KLEISVPPSK	
INLPGDVISIAVGDLNDKMR			ALGAPSYQIVVEVNSLNLR	NKISFVQILRLEN	
IKTFIDSFPNKGFYPLGR			AFYTNGSSPPK		
SKSVEDYKDMLK			YYIIYEVKDSEKSPQEIR		
MNESSYEK			KLEEVELKGK		
MNESSYEK			KLEEVELK		
			YYIIYEVK		

Subunit names in parenthesis correspond to the general nomenclature adopted for Ampullariid PV1s followed by the unigene IDs taken from Ip et al. (2018). N-terminal sequences (bold) and internal aminoacid sequences of PdPV1 subumits were obtained by Edman degradation and MS/MS, respectively.

Figure S1: **Subunit sequences of PdPV1.**

A

PdPV1-1 (Pdi67619_c0_g1)
*MLVATLILVALSAVFTNVYA*SCQYIILDVNKVKSKSVEDYKDMLKEVDLVYSFKVVGSSRLLFVVRM**NESS**YEKLSKINLPGDVISIAVGDLNDKMR
TMGVDWKKWDELPNA**NLTL**FERTLKFEGEPTQGAISHVQAYGEKIKTFIDSFPNKGFYPLGRTPFKAYFIISLSFRCHQVGVGSAYALNYLNGPGD
STTKVEFLTKV

PdPV1-2 (Pdi37843_c0_g1)
*MFVATSLLLALATLVATVASK*DEYLIMEINPPKKVSEQEILNMLSPLEIKHKFRVTGTTKLWIVIKLDPVGYMKLD**NITV**PGKVSVTPAEDMVDTM
QKFGVFWPQASLTEE**NITL**FQSHSYLTDVTKEQLTAMMVGYGEHMVETLKSHPYQFYRATGVTPHKHFFFVSSARDEVEVIGREGVDIWGGPG
EYHVKPEYLIRI

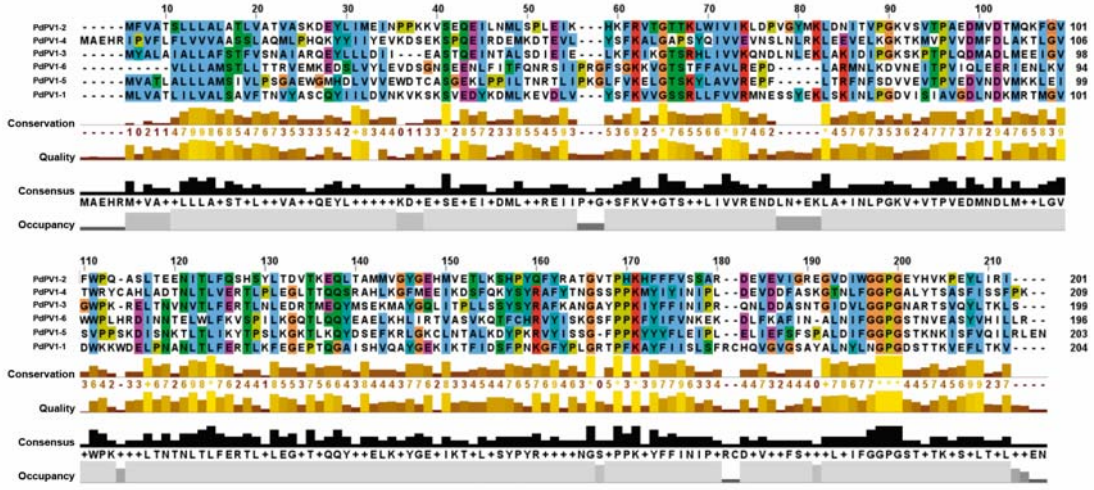
PdPV1-3 (Pdi14773_c0_g1)
*MYALAIALLAFSTFVSNAIAR*QEYLLLDIIEASTQEINTALSDIEIELKFKIKGTSRHLIVVKQNDLNLEKLAKIDIPGKSKP**T**PLQDMADLMEEIGVGW
PKRELTNV**NVTL**FERTLNLEDRTMEQYMSEKMAYGQLITPLLSSYSYRAFKANGAYPPKIYFFINIPRQNLDDASNTGIDVLGGPGNAR**T**SVQYLT
KLS

PdPV1-4 (Pdi37785_c0_g1)
*MAEHRIPVFLFLVVVAASSLAQ*MLPHQKYYIIYEVKDSEK**S**PQEIRDEMKD**T**EVL**Y**SFKALGAPSYQIVVEVNSLNLRKLEEVELKGKTKMVPVVD
MFDLAKTLGVTWRYCAHLADT**NLTL**VERTLPLEGLTTQQ**S**RAHLKGFMEEIKDSFQKYSYRAFYT**NGSS**PPKMYI**Y**INIPLDEVDDFASKGTNLFG
GPGALY**T**SASFISSFPK

PdPV1-5 (Pdi118877_c0_g1)
*RQLATSSHIANLSSLEDCSRRQSRMVATLALLLAMSIVLPSGAE*WGMHDLVVVEWDTCASGEKLPPILT**NRTL**IPKGLFVKELGTSKYLAVVREPFL
TRF**NFSD**VVEVTPVEDVNDVMKKLEI**S**VPPSKDIS**NKTL**TLIKY**T**P**S**LKGKTLKQYDSEFKRLGKCLNTALKDYPKRVYISSGFPPKYYYFLEIPLELIEF
SFSPALDIFGGPGSTKNKISFVQILRLEN

PdPV1-6 (Pdi33234_c0_g1)
*VLLLAMSTLLTTRVEM*KEDSLV**Y**LEVDSGNSEENLFITFQ**NRSI**IPRGF**S**GKKVGT**S**TFFAVLREPDLARMNLKDVNEITPVIQLEERIENLKVWWPL
HRDINNTELWLFKVSPILKGQTLQQYEAELKHLIRTVASVKQTFCHRVYISKGSFPPKFYIFVNKEKDLFKAFINALNIFGGPGSTNVEASYVHILLR

B



A) Deduced aminoacid sequences of the six PdPV1 subunits. Putative signal sequences are in italics and underlined. Potential phosphorilation sites are in bold underlined, potential N- and O-glycosylation sites are in bold and on black squares respectively. A conserved sequence is marked in bold red. B) Multiple sequence alignment of PdPV1 subunits.