Identification of Novel Proteins in Foam Nests of the Japanese Forest Green Tree Frog, Rhacophorus arboreus

Authors: Shigeri, Yasushi, Nakata, Makoto, Kubota, Hiroshi Y., Tomari, Naohiro, Yamamoto, Yoshihiro, et al.

Source: Zoological Science, 38(1) : 8-19

Published By: Zoological Society of Japan

URL: https://doi.org/10.2108/zs200113

ZOOLOGICAL SCIENCE **38**: 8–19 (2021)

Identification of Novel Proteins in Foam Nests of the Japanese Forest Green Tree Frog, *Rhacophorus arboreus*

Yasushi Shigeri[1]*, Makoto Nakata[2], Hiroshi Y. Kubota[3], Naohiro Tomari[4], Yoshihiro Yamamoto[4], Koichi Uegaki[5], Yoshikazu Haramoto[6], Chloe Bumb[1,7], Yoshie Tanaka[1], Tomoya Kinumi[8], and Hidetoshi Inagaki[9]

[1]*Department of Chemistry, Wakayama Medical University, 580 Mikazura, Wakayama 641-0011, Japan*
[2]*Peptide Institute, Inc., 7-2-9 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan*
[3]*Department of Zoology, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan*
[4]*Kyoto Municipal Institute of Industrial Technology and Culture, 134 Chudoji-minamimachi, Shimogyo-ku, Kyoto 600-8813, Japan*
[5]*Department of Applied Biological Chemistry, Faculty of Agriculture, Kindai University, Nara 631-8505, Japan*
[6]*Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan*
[7]*Faculty of Pharmacy, University of Strasbourg, 74 Route du Rhin CS60024, 67401, Illkirch, Cedex, France*
[8]*Research Institute for Material and Chemical Measurement, National Metrology Institute of Japan (NMIJ), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba C-3, Ibaraki 305-8563, Japan*
[9]*Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan*

Foam nests of frogs are natural biosurfactants that contain potential compounds for biocompatible materials, Drug Delivery System (DDS), emulsifiers, and bioremediation. To elucidate the protein components in the foam nests of *Rhacophorus arboreus*, which is an endemic Japanese frog species commonly seen during the rainy season, we performed amino acid analysis, SDS-PAGE electrophoresis, and matrix-assisted laser desorption/ionization mass spectrometry using intact foam nests. Many proteins were detected in these foam nests, ranging from a few to several hundred kDa, with both essential and non-essential amino acids. Next, we performed transcriptome analysis using a next-generation sequencer on total RNAs extracted from oviducts before egg-laying. The soluble foam nests were purified by LC-MS and analyzed using Edman degradation, and the identified N-terminal sequences were matched to the transcriptome data. Four proteins that shared significant sequence homologies with extracellular superoxide dismutase of *Nanorana parkeri*, vitelline membrane outer layer protein 1 homolog of *Xenopus tropicalis*, ranasmurfin of *Polypedates leucomystax*, and alpha-1-antichymotrypsin of *Sorex araneus* were identified. Prior to purification of the foam nests, they were treated with both a reducing reagent and an alkylating agent, and LC-MS/MS analyses were performed. We identified 22 proteins in the foam nests that were homologous with proteinase inhibitors, ribonuclease, glycoproteins, antimicrobial protein and barrier, immunoglobulin-binding proteins, glycoprotein binding protein, colored protein, and keratin-associated protein. The presence of these proteins in foam nests, along with small molecules, such as carbohydrates and sugars, would protect them against microbial and parasitic attack, oxidative stress, and a shortage of moisture.

Key words: Edman degradation, oviduct, proteome analysis, *Rhacophorus*, transcriptome analysis

INTRODUCTION

Several types of biosurfactants have been reported, such as alkyl chain phospholipids, DNA surfactant, lipopeptides, peptides, and proteins. Generally, they are amphiphiles possessing both hydrophilic and hydrophobic properties (Lu et al., 2007). Foam nests are known as biosurfactant-containing structures, and frogs, such as Leptodactylidae from Central and South America, and Rhacophoridae from

* Corresponding author. E-mail: yshigeri@wakayama-med.ac.jp
doi:10.2108/zs200113

Africa to Asia, create foam nests as biocompatible incubators for their fertilized eggs and embryos to protect them from environmental stress and microbial and parasitic attacks (Cooper and Kennedy, 2010; Cooper et al., 2017). Oke et al. (2008) reported the structure and physicochemical properties of an unusual blue protein, designated ranasmurfin, isolated from the foam nest of the Malaysian tree frog, *Polypedates leucomystax* (Rhacophoridae). Ranasmurfin contains zinc and has a 26 kDa homodimeric structure with little sequence similarity to any known proteins. Although its function has not been well clarified, it is thought to perform foam stabilization and to protect the eggs and embryos from harmful UV irradiation and microorganisms (McMahon et al., 2006; Cooper et al., 2017). Fleming et al. (2009) reported six novel proteins in the 10–30 kDa range, designated ranaspumin (Rsn-1 to Rsn-6) that predominate in the foam nests of the tungara frog, *Engystomops pustulosus* in the family Leptodactylidae of the tropical Americas. Among them, four proteins (Rsn-3 to Rsn-6) were proven to be lectins, while Rsn-1 was structurally similar to proteinase inhibitors of the cystatin family. Rsn-2 had a unique amino acid sequence with a charged hydrophilic C-terminus and a non-polar N-terminal sequence, suggesting an amphiphilic character. Rsn-2 actually reduced the surface tension of water more effectively than control proteins (lysozyme and BSA) (Brandani et al., 2017). Hissa et al. (2008) reported the novel and predominant protein Lv-ranaspumin (Lv-RSN-1) derived from the foam nests of *Leptodactylus vastus*, which is endemic to South America (Hissa et al., 2008). Lv-RSN-1, a 23.5 kDa monomeric protein, showed 40% similarity to Rsn-1, possessing a structure with 11 α-helices and two small antiparallel β-strands, and displaying a natural surfactant activity (Hissa et al., 2014). Moreover, they revealed that the foam nest from *L. vastus* contained more than 100 proteins. Among them, Lv-RSN-1 was about 45% of the total protein in the foam nest (Hissa et al., 2016).

The Japanese green tree frog, *Rhacophorus arboreus* (Amphibia, Anura, Rhacophoidae), is a frog species that is endemic to Japan (Fig. 1A), with some populations of this frog having protected status as a “special natural monument” (Okada and Kawano, 1924; Kaneko and Matsui, 2004). Although the frogs usually live in mountainous forests, they can be found near ponds during the breeding season from April to July for mating. During their mating, the female frogs excrete fluids from the cloaca and stir them with their hind legs to form a sponge-like foam, into which they lay their eggs, while the male frogs hold on to the backs of the females and release sperm to fertilize the eggs in the foam nest (Fig. 1B). The foam nest, which is hung from a tree branch over a pond, hardens superficially, protecting the eggs from desiccation and predation (Kusano et al., 2006). After summer arrives, the tadpoles emerge from their eggs, leave the foam nest, and drop down into the pond. Subsequently, the foam nest melts, tears, and drops into the pond, where it disperses. Most studies on *R. arboreus* have been focused on its reproductive system, ecology, and infection by microorganisms (Kasuya et al., 1992; Muto and Kubota, 2009, 2011; Haridy et al., 2014; Yokoe et al., 2016), while the biochemical and structural components of the foam nest of *R. arboreus* have been an enigma. In the present study, we focused on the protein components in the foam nest of *R. arboreus*. Using LC-MS/MS (liquid chromatography-tandem mass spectrometry) and next-generation sequencing (NGS) as well as conventional biochemical techniques, such as amino acid analysis and Edman degradation, we made a global analysis of the *R. arboreus* foam nest to clarify its protein components at the genetic and amino acid levels.

MATERIALS AND METHODS

Collection of *R. arboreus* and the foam nest

Although *R. arboreus* is listed in the International Union for the Conservation of Nature (IUCN) Red List of threatened species as of least concern (Kaneko and Matsui, 2004), it is not designated as a protected species in Shiga Prefecture, Japan. Therefore, we contacted the Shiga Prefectural Office and were granted permission to collect a small number of them. At a house pond in Hieidaira (Shiga Prefecture, Japan), several females that were not in amplexus were captured during the breeding season. They were kept in a bucket and spontaneously prompted to lay eggs without male frogs. After egg-laying, the foam nests and eggs were separated and stored at –80°C until the day of the analyses.

Analyses of intact foam nests

Amino acid analysis

Foam nest (wet weight 10.02 mg) from *R. arboreus* was hydrolyzed with hydrochloric acid (6 M, 500 μL) at 110°C for 24 h. The resulting hydrolysate was evaporated under nitrogen gas, dissolved in 1 mL dilution buffer (0.2 M sodium citrate buffer, pH 2.2), and examined by automated amino acid analysis. Amino acid analysis was quantitatively performed using an L-8800 amino acid analyzer (Hitachi High-Tech Science, Tokyo, Japan), ninhydrin reagent L-8500 set, and amino acid mixture standard solution type H (FUJIFILM Wako Pure Chemical, Osaka, Japan). Separation was performed on an ion-exchange column according to the manufacturer’s instructions. Spectral absorbances of amino acid derivatives were detected at 440 nm for proline and 570 nm for most other amino acids except for proline.

SDS-PAGE

Proteins in the foam nest were analyzed by SDS-PAGE, as follows. SDS-PAGE sample buffer including 1 mM DTT (dithiothreitol) was added to the foam nest (2 μL or 5 μL), and then heated at 95°C for 10 min. Aliquots were examined by SDS-PAGE and stained with Coomassie blue using 10–20% gradient gels (Oriental Instruments, Sagamihara, Japan), and compared to pre-stained protein markers (M.W. 6500–200,000, Nacalai Tesque, Kyoto, Japan).

MALDI-MS analysis

Next, the foam nest that was dissolved in the solution 0.1% TFA + 99.9% acetonitrile was directly analyzed by matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry (MALDI-TOF MS). The MALDI-TOF MS spectra were acquired using sinapinic acid as a matrix and a Microflex AI mass spectrometer (Bruker Daltonics, Yokohama, Japan) with a pulsed nitrogen laser (337 nm) in the linear positive ion mode. Sinapinic acid, trifluoroacetic acid (TFA), and all other reagents were obtained from FUJIFILM Wako Pure Chemical or Nacalai Tesque.

Purification by LC-MS and Edman degradation analyses of proteins in foam nests

The foam nest (wet weight about 98 mg) was dissolved in 400 μL trifluoroethanol (TFE) and passed through a 0.45-μm filter. The resulting solution (10 μL) was analyzed with LC-MS, monitoring the absorbance at 220 and 280 nm. LC (Nexela HPLC system, Shimadzu, Kyoto, Japan) conditions were as follows: solvent A (0.1% TFA); solvent B (0.1% TFA in acetonitrile); 20–60% linear gradient of solvent B in solvent A at a flow rate of 1.5 mL/min for 30 min;

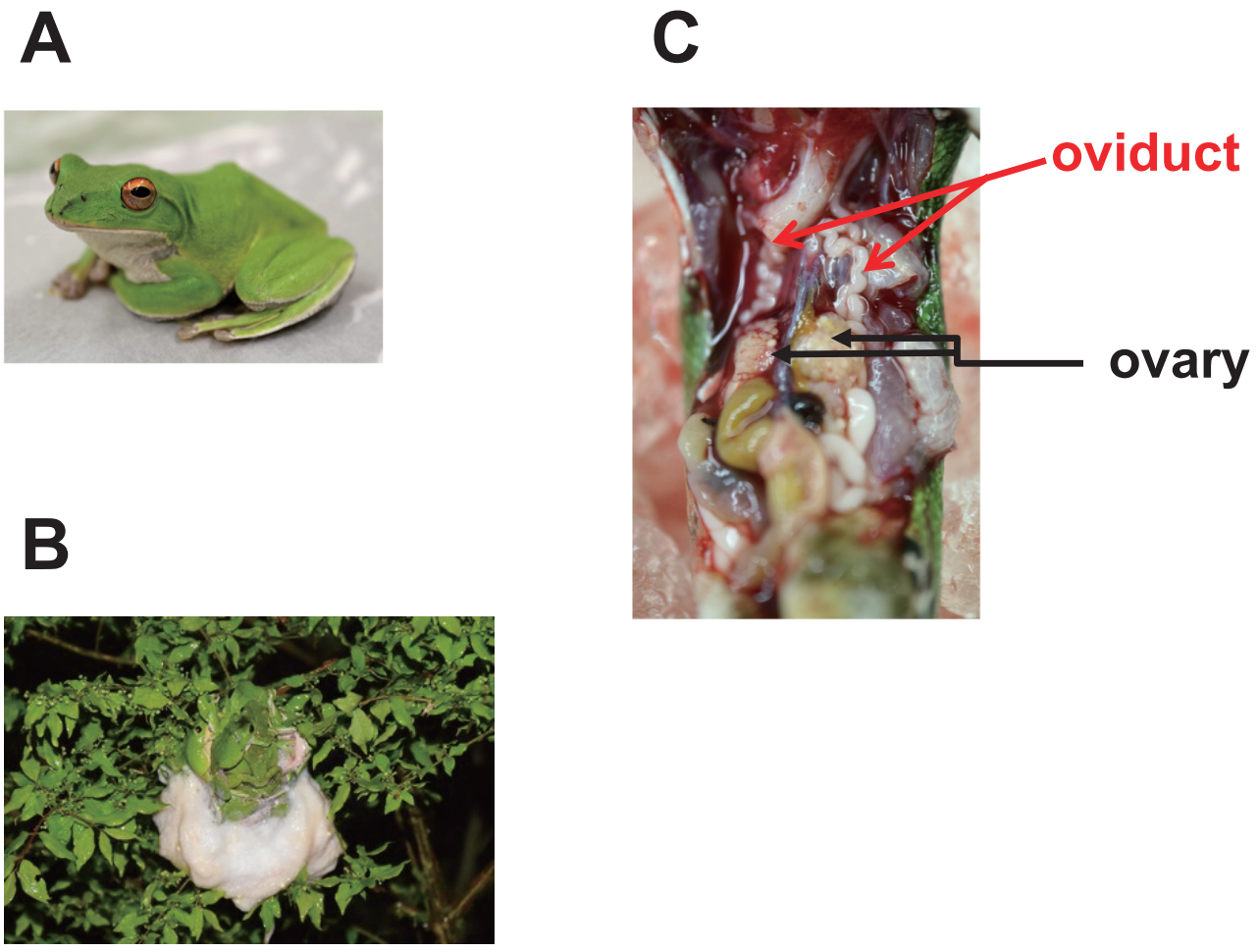

Fig. 1. **(A)** The Japanese green tree frog, *Rhacophorus arboreus*. **(B)** The female *R. arboreus* lays several hundred eggs in her foam nest on branches hanging over water. **(C)** Oviducts were removed from the abdomen of *Rhacophorus arboreus*.

column, Imtakt Cadenza CD-C18 (6 mm I.D. × 250 mm, Imtakt, Kyoto, Japan); column temperature, 40°C. The molecular masses of the separated protein fractions were verified using a LCMS-8050 triple quadrupole mass spectrometer (Shimadzu) in the single MS mode. The N-terminal amino acid sequences of the purified proteins were determined by automated Edman degradation using a PPSQ-33A protein sequencer (Shimadzu).

Transcriptome analysis (RNA-seq analysis)

The frog oviduct is known to be responsible for the production of the foam fluid (Coe, 1974; Kabisch et al., 1998; Furness et al., 2010), and total RNAs were extracted from *R. arboreus* oviducts before egg-laying (Fig. 1C). A TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA) was utilized to construct the library. The next-generation sequencing (NGS) was performed by Macrogen Japan (Kyoto, Japan) using a next-generation sequencer (NovaSeq6000, Illumina). The adapter and low quality sequences were eliminated using Trimmomatic 0.36 (Bolger et al., 2014) and FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The transcriptome data were assembled and reconstructed using Trinity (v.2.1.1) software (Grabherr et al., 2011; Haas et al., 2013). To estimate the relative expression levels of the oviduct transcripts, all input reads were mapped and counted using Bowtie 2 (Langmead and Salzberg, 2012). After construction of a FASTA format file including the amino acid sequences derived from major transcripts, the file was utilized for a MASCOT search. The original sequence reads were submitted to DDBJ under accession no. DRA010399. The project and sample ID in this article are PRJDB10049 and SAMD00232153. To identify the proteins in the foam nests, and to confirm their amino acid and nucleotide sequences based on the transcriptome data, the NCBI database (http://www.ncbi.nlm.nih.gov/), and Xenbase (http://www.xenbase.org/entry/), including genomic information for *Xenopus tropicalis* and *Xenopus laevis* (Hellsten et al., 2010; Session et al., 2016), were utilized. The assembled nucleotide sequences in this study were submitted to DDBJ/EMBL/GenBank Transcription Shotgun Assembly (TSA) under accession numbers ICQQ01000001–01000022.

LC-MS/MS analyses

LC-MS/MS analyses were performed as follows. A foam nest (wet weight 30 mg) was dissolved in a solution containing 50 μL NH_4HCO_3 (1 M, pH 8.5), 96 mg urea, 10 μL TCEP (Tris(2-carboxyethyl) phosphine HCl, 1 M), and 160 μL H_2O. Although it was a sticky solution at first, it changed to a smooth solution after incubation at 37°C for 1 h for the reductive reaction. Iodoacetamide (500 mM, 30 μL) was added and the solution was incubated at room temperature for 45 min for the alkylation reaction. Subsequently, 500 μL H_2O and trypsin (10 μg, porcine pancreas for proteome research, FUJIFILM Wako Pure Chemical) or lysyl endopeptidase (20 μg, mass spectrometry grade, FUJIFILM Wako Pure Chemical) were added and the solution was incubated overnight at 37°C for the protein digestion. After the addition of 1 μL TFA and filtration (0.45-μm filter), the

solution (5 μL) was used in the LC-MS/MS experiments. Determination of protein structures with MS/MS experiments was performed on a Bruker maXisII Q-TOF mass spectrometer (Bruker Daltonics, Yokohama, Japan) interfaced with an LC pump (Nexela HPLC system, Shimadzu) and XBridge peptide C18 column (3.5 μm, 2.1 mm I.D. × 150 mm, Waters, MA, USA) at a flow rate of 0.2 ml/min using a gradient mode from solution A (0.1% formic acid) to 50% solution B (0.1% formic acid in acetonitrile) for 120 min. Mascot database search software (Matrix Science, Tokyo, Japan) was utilized to identify proteins.

RESULTS

Analyses of intact foam nests

To identify the component proteins of *R. arboreus* foam nests, conventional and advanced analyses, such as SDS-PAGE, automated amino acid analysis, and MALDI-MS were performed using intact foam nests. As shown in Fig. 2A, the foam nest from *R. arboreus* showed many protein bands, with sizes ranging from a few to several hundred kDa, on an SDS-PAGE gel. Next, we performed amino acid analysis of the foam nests. Due to the acid hydrolysis process in the amino acid analysis, tryptophan was generally disrupted and cysteine was recovered as cystine. Moreover, after deamidation, asparagine and glutamine were converted to aspartic acid and glutamic acid, respectively. As shown in Table 1, the foam nests contained the essential (His, Ile, Met, Leu, Lys, Phe, Thr, Val) and non-essential amino acids (Ala, Arg, Cys, Gly, Pro, Ser, Tyr, Asp or Asn, Gln or Glu), as do frog meat of *Rana ridibunda* (Cagiltay et al., 2014) and human venous blood (Velazquez et al., 1977). However, the content of cysteine in the foam nests was about four times larger than that of frog meat (*R. ridibunda*). Figure 2B illustrates the MALDI-MS spectrum from the foam nests, with more than nine major protonated peaks as follows (*m/z* 4257, 5030, 5329, 6147, 6863, 7224, 10482, 12612, 14224, etc.).

Purification of intact foam nest proteins and Edman degradation analyses

Since insoluble components were found in the foam nests after freeze-thawing several times, the foam nests were dissolved in TFE and the resulting solution was applied to reversed-phase HPLC. Figure 3 shows an HPLC chromatogram from the foam nest solution monitoring the absorbance at 280 nm. Seven major peaks (Fraction 0–Fraction 6) were collected and analyzed by Edman degradation to identify the N-terminal amino acid sequences of the proteins present in each fraction. As a result, a 14-amino acid peptide (GPPHGYKEKKLYAI), 12-amino acid peptide (LDIGVDNGGPQG), 14-amino acid peptide (LDIGVDNGGPQGTW), 20-amino acid peptide (LDIGVDNGGPQGTWGKLDSC), 11-amino acid peptide (GDILRGCPFPS), and 11-amino acid peptide (YDGYLKLKVPK) were identified as N-terminal amino acid sequences of proteins in Fractions 1, 2, 3, 4, 5, and 6, respectively (Table 2). Since reversed-phase HPLC

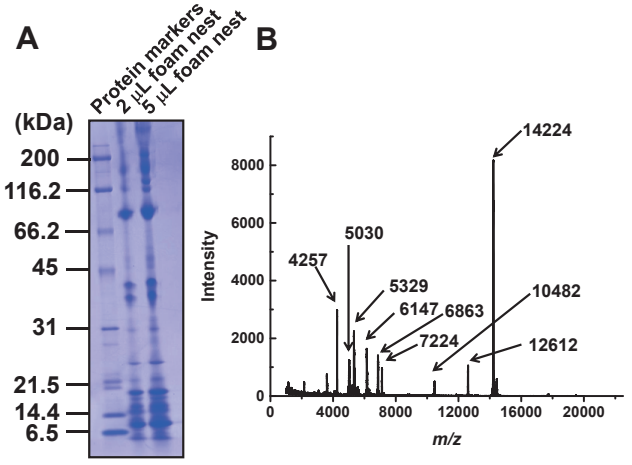

Fig. 2. (A) Electrophoretic analysis of foam nests from *Rhacophorus arboreus*. **(B)** MALDI-TOF MS spectrum of foam nests from *Rhacophorus arboreus*.

Table 1. Amino acid composition of foam nest.

Amino acid	Number of amino acid residues		
	Foam nest	Frog meat (*Rana ridibunda*)*	Venous blood (human)**
Ala (A)	39.6	136.9	139.9
Cys (C)	57.5	14.3	99.9
Asp + Asn (D + N)	125.0	90	37.4
Glu + Gln (E + Q)	111.9	105.5	168.4
Phe (F)	33.9	40.4	19.7
Gly (G)	90.0	92.5	128
His (H)	18.0	28.4	38.1
Ile (I)	59.1	54.4	21.8
Lys (K)	90.4	78.2	44.5
Leu (L)	70.9	57.3	29.4
Met (M)	14.6	30.3	17.1
Pro (P)	51.7	28.7	62.3
Arg (R)	21.0	35.7	19.1
Ser (S)	58.1	77.7	48.6
Thr (T)	66.8	37.0	66.7
Val (V)	50.4	64.4	42.3
Tyr (Y)	40.9	28.2	16.7
Total number of amino acid residues	1,000	1,000	1,000

*(Cagiltay et al., 2014)
**(Velazquez et al., 1977)

Table 2. N-terminal amino acid sequences of proteins in foam nest determined by Edman degradation and molecular weight estimation by LC-MS analyses.

	N-terminal sequences of proteins in the foam nest by Edman degradation	Molecular mass by LC-MS analyses
Fraction 0	Not identified	
Fraction 1	GPPHGYKEKKLYAI	19486
Fraction 2	LDIGVDNGGPQG	12600
Fraction 3	LDIGVDNGGPQGTW	
Fraction 4	LDIGVDNGGPQGTWGKLDSC	
Fraction 5	GDILRGCPFPS	14222
Fraction 6	YDGYLKLKVPK	

was directly connected to mass spectrometry, multiply charged and protonated molecule peaks from each protein could be detected. The *m*/*z* values of these ions allowed us to estimate that the molecular weight of the proteins included in Fraction 1, Fraction 2, and Fraction 5 were approximately 19486, 12600, and 14222, respectively (Fig. 3).

Transcriptome analysis (RNA-seq analysis)

Transcriptome analysis with a next-generation sequencer was performed using total RNAs extracted from oviducts before egg-laying. The obtained nucleotide sequences were assembled and reconstructed using Trinity software, and relative expression levels were estimated using Bowtie 2. At first, using N-terminal amino acid sequences obtained

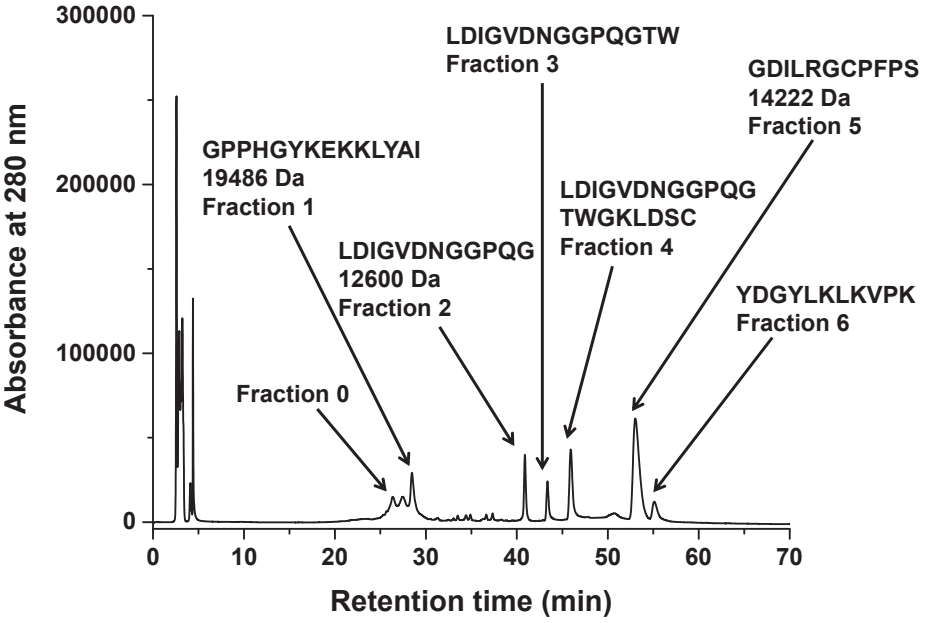

Fig. 3. Purification of foam nest constituents from *Rhacophorus arboreus* by reversed-phase HPLC monitoring the absorbance at 280 nm. For each fraction, the N-terminal amino acid sequence determined by Edman degradation and the predicted molecular mass calculated from the result of LC-MS are shown.

Table 3. Amino acid sequences of proteins in foam nest identified by Edman degradation, transcriptome analysis, and protein BLAST search.

Contig number (Accession number)	Homologous proteins and identity (%) by protein BLAST search, and amino acid sequences of identified proteins
(1) comp84488_c0_seq1 (ICQQ01000001)	extracellular superoxide dismutase (Cu-Zn)-like (*Nanorana parkeri*), 53.9% identity
MHCVSFILLVYLSCAVQA**GPPHGYKEKKLYAI**CNLQPDTKLEEADPEITGQITFKQSFPHGKLEAYFHIKGFPLTPSESTPRAINVHTYGDLS DGCHSVGGHYNPEFVDHPHHPGDFGNFLVQNGEIQQHRSNLAANIYGTTSIIGRSVVISKYADDLGTGHDEGSLEHGNSGPPLACCVI GISNKDSWEKLVPEPSKK	
Estimated monoisotopic mass without signal sequence: 19725	
(2) comp83308_c0_seq2 (ICQQ01000002)	vitelline membrane outer layer protein 1 homolog (*Xenopus tropicalis*), 50.0% identity
MLLSWIVTLFLLHTKLAHC**LDIGVDNGGPQGTWGKLDSC**PAGTLARGFELLVESFQGPLDDTALNGIRLFCAKPNEHVAVAYITSKIGNWGT WSASQWCPHGNIISFSLQVELFGILDNTAANNIQFLCSDGTILTGYGGPWGSYGGWSQACPDGIRGITAKVEDPQGKEKDDTALNDVLFRCY	
Estimated monoisotopic mass without signal sequence: 17627	
(3) comp81644_c2_seq1 (ICQQ01000003)	ranasmurfin (*Polypedates leucomystax*), 30.0% identity,
MKAILLLALFGLSFCIG**GDILRGCPFPS**SDFSDAKDCLELALKTNAKFKEHTGKFICAYLEKDKHKSKFKKAAEEYDAILDCLGCEGDKIDILTD KADILQTAFNKLLGFWEIFGPFSQTLCHLFELIPELPVKDVLGGLGGLGGLGG	
Estimated monoisotopic mass without signal sequence: 14299	
(4) comp79296_c0_seq1 (ICQQ01000004)	alpha-1-antichymotrypsin (*Sorex araneus*), 21.7% identity
MKALLLLCASAALVHLAAS**YDGYLKLKVPK**HYPPLDSNDYIYIIIIFHIDALDKLYNALENAKSSVPPLTDITVYISKDWHPPNKNIPTVPVD VSSPDAALDTLNAFVESKSSGKITKLVTGVTSHTAAIILNVINALKVPSPGGTGWTKLLDGEYKKVENTDLGVTLIEIPYSVYVSILYAIPSAG KVEAVGQSLIPATIEALRKSLTLQSVKVEVPRATLYTGFFMAVEKEGELDSTTYGYKVSAQFP	
Estimated monoisotopic mass without signal sequence: 25062	

Boxed sequences are signal peptide sequences predicted by SignalP 5.0 (Armenteros et al., 2019) and bold and underlined sequences are N-termini identified by Edman degradation.

Table 4. Proteins in foam nest identified by LC-MS/MS and transcriptome analysis.

Contig number search (Accession number)	Homologous proteins and identities (%) obtained by protein BLAST
The foam nest was treated with trypsin for LC/MS/MS analyses	
comp84262_c0_seq1 (ICQQ01000005)	ovostatin (*Xenopus tropicalis*), 31.6%
comp82631_c0_seq1 (ICQQ01000006)	IgGFc-binding protein (*Xenopus tropicalis*), 41.5%
comp84036_c0_seq1 (ICQQ01000007)	hemocytin (*Xenopus tropicalis*), 35.5%
comp83999_c1_seq3 (ICQQ01000008)	IgGFc-binding protein (*Xenopus tropicalis*), 57.5%
comp83288_c0_seq3 (ICQQ01000009)	WAP four-disulfide core domain protein 3-like (*Xenopus tropicalis*), 29.2%
comp82261_c0_seq1 (ICQQ01000010)	BPI fold-containing family B member 4 precursor (human), 29.3%
comp83288_c0_seq2 (ICQQ01000011)	antileukoproteinase precursor (human), 33.8%
comp83686_c0_seq2 (ICQQ01000012)	ovostatin (*Xenopus tropicalis*), 41.7%
comp83288_c0_seq1 (ICQQ01000013)	WAP four-disulfide core domain protein 3-like (*Xenopus tropicalis*), 31.9%
comp83308_c0_seq2 (ICQQ01000002)	vitelline membrane outer layer protein 1 homolog (*Xenopus tropicalis*), 50.0%
comp81644_c2_seq1 (ICQQ01000003)	ranasmurfin (*Polypedates leucomystax*), 30.0%
comp83850_c0_seq1 (ICQQ01000014)	ficolin-1 (*Lithobates catesbeianus*), 64.2%
comp84464_c0_seq3 (ICQQ01000015)	alpha-2-macroglobulin-like protein 1 isoform 2 (human), 42.1%
comp83677_c1_seq2 (ICQQ01000016)	onconase precursor (*Rana pipiens*), 56.7%
comp79294_c0_seq2 (ICQQ01000017)	hylaserpin S2 (*Hyla simplex*), 32.9%
comp76542_c0_seq1 (ICQQ01000018)	mucin-5AC precursor (human), 45.5%
comp83857_c2_seq2 (ICQQ01000019)	serine protease inhibitor 2 (*Rana chensinensis*), 53.4%
comp82467_c0_seq1 (ICQQ01000020)	beta-microseminoprotein isoform X5 (*Xenopus tropicalis*), 43.9%
comp83603_c0_seq1 (ICQQ01000021)	keratin-associated protein 10-4-like (*Xenopus tropicalis*), 49.0%
comp79314_c0_seq3 (ICQQ01000022)	fucolectin (*Xenopus tropicalis*), 59.6%
The foam nest was treated with lysyl endopeptidase for LC/MS/MS analyses	
comp82261_c0_seq1 (ICQQ01000010)	BPI fold-containing family B member 4 (*Xenopus tropicalis*), 29.3%
comp79296_c0_seq1 (ICQQ01000004)	alpha-1-antichymotrypsin (*Sorex araneus*), 21.7%

by Edman degradation and the assembled data of nucleotide sequences, the proteins in Fraction 1, Fractions 2–4, Fraction 5, and Fraction 6 were proven to be identical to comp84488_c0_seq1 (199 amino acids), comp83308_c0_seq2 (184 amino acids), comp81644_c2_seq1 (148 amino acids), and comp79296_c0_seq1 (250 amino acids), respectively, as

Table 5. Amino acid sequences of proteins identified in foam nest by LC-MS/MS analyses and transcriptome analysis.

Contig number	Accession number	Amino acid sequences of identified proteins

comp84262_c0_seq1 (ICQQ01000005)

MIRRESRLLCALLLLGVLGGATSQPSNDPEPLIITDPLIDPVPVASCRSPTTRTKSPGPKCNYYMVVPFKLTSGCDGVVCTWLKCEVPVTVKVTMEYYFEKNIIVSKEDTTEY
FECTDFTVPQVEKQEAIYFIYEAVGDNVKVIARKPAVIDVPRTACLAQKNKPLYKPGDTVQLRLPCFDENLKQVVKNFTKFTVADPTGIRVFQELNPTIDHGVATLSFVTSDLSREGYW
GIQFEDDKGQIQYSSFQLQLYELPRFDWNTKCPNTISVLDKKAEISGSAQYVYGKPVPGTVVVRCCRTTPKYGRNANCFRGVDDICVEFTKELDENGNFTEQLDLDVFKLPFSGLDN
SITCDMTMKEDGTGILVPSYCYMSITNKLAVLLLDRTIIGTYYNKVNVLFVATLTDEKNNPLAYKEITVQIDGVDVKYVTTDANGRAECEIDTSQYYKANITLTATYSPSDLCYGEDSFG
SYYAYSVRSYPYDEVILYRYYSQSKSQIRARPLTGVLPCHLNHRVQLDYSFTKEGVGKDATTINIYYIIISRNIIAKHGDISVDLSEGLKGSTYIEFYADYKVATEAKVLVFVATAKDIISDT
VSLEIGLCFENEVNLAFTKPIVAPGATVERRITAEPGSFCGLRAKDASFNLIAQYDEFNANDIHSRVRSYLFGLQYGDVDYKDPEPPCIDGNTENFCNGVYVKPTSARTDRDSNQD
FEEYSLLLFSNLDKRRPKVCGEETEPVYPKYYASSIETIGRVPDAGSFADAGSAKRSIDTVRDDFRDIFGYDTTIIGSGGESIVSSKIPQTITNWESDAFCISSTKGIGLTIDPAKTT
TIQPFFVEASGPPSVTRGEKVDFSITAANYLNKCTQIKAEVENSENYEVTFPDGVQEKCACSQQRVSWTVSFTFLRIGTVNLTMKASTTKVSDTCDGSQDDAEPLRSDTIIQGIDVEAE
GIKREAASSHLVTVKDGTAAVEISVREPNDLVPDSFRCIVNIMGDPVALSVENFDNLLSTQTSCCDQIQAELLSIIFLERYLKTFGELSEDQSKKMLAKFASGYMNLLSCRQYDG
GYSWYRSSYTSNLWGTVSSLRTIHFLSDYVYIDDSVPTNGLIYISRFQDLKSGCLNPETPSIIYQDHYNASIQITAMAAPILQELSGNYSSAGPLLEGVNKCLKNVYFERQNTLTQCYLF
NAAASRGDWDTWQSYYIILLGKAIYNGDTCYWESDNSVRPVIYHYFPRSSGDILSTAIVLISLTKYPDPGQDLINQMASIAKWLSGQLTRTGGWYSSQATSAAIEALTLFGRIINTGET
NAQVSLQRDGIEVFSVTIDRSNKFIVQSKEISPCFGEYKFIVTGTASPLLQLATFYNTKIATGEDSAYIISAQASSPHCFGGVARLVNIKISLRFNGASGGYSRTYLRVDIFTGYQPDY
SIQNKIPNVREYKVKGRKALIVIEIFDTKTIEFTIPIYLRSRVTTFQTAYIVAEDIDTGETGIADYRYPCNDELTKQ

Estimated monoisotopic mass without signal sequence: 164015

comp82631_c0_seq1 (ICQQ01000006)

RRYKGQIRGHKLDVLDADMSGTIALQLSLILSCLVGVAFPQSECQSKVCNADEVCAVQDGNEICVSAFGSYCWIVGDPHYKPFDGAQYDFEGNCYYTISKSTSQDDGLIPFNIEARN
ENRGNSKVSFVGLVRLYVYNNKIVFSRQQRGFAQVNNQISRLPLSLDNKVTITQSGHFAYVATDFLKLYYDWNSVMIIQLSKQYLGKVEGMCGSYQTLDGAYIQKESLLPQL
DAIAIGKSFATPESDTKCWHDCHGPCTQCPEDKKYETNCQIILAEKGPFSGCHNVVDPQIFYEYCVNDQCISNGAKYILCQSLHSYKAICQIKGAKIGEWRRQTACSMGCPANS
VYKLCGNPCPKTCNDDITQPCSMPCLETCECRQGYVLDGGKCVPQANCGCVSNGRIYSPNETFWDDPSCTRKCKCEPDTRTVVCFSETCKKYEKCAVVRGIQTCQSLSNGTCAAY
GDPHFRTFDKVNYDFQGVCVYLFSGLRRNGTGLEDFQVYIQNKNRGNRAVSYVHVVKVRVYDLEIILNSYYVDKIQVNGLFVNFPYVVGLDDISIYKSGFDAVIETDFGLRITYNWQ
GYLLVTVPGVYGGTLGGLCGNFDGNANNDFETPDQKLVPNPNAFGTRWTVEYDPPDCNTFNGFPCLKLKQIIAEHRNKKIGCGIMLDKNGPFKNCHAKVNPETLFNDCAFDACYYH
GRLDAICQVVSAYATICLAYGITVDPWRSDLFCDVLCPPNSHYRICAPDCQPTCASYGPPLGCTPGCAKGCVCDDGYLLSGNDCVTPDECGCTYEGLYYKSGSKFLTKECTESCT
CEIGVNLNCKVYSCALNETCSNENGYLTCIPLGPPRHCFVSGNSNYQTLDGVTYDFEGNCAYVLSASTPKPGLPSFAVYLNNERYGKECMAISKSVALEVYGHTLILQHNRGGFIT
VDGVLYNMPFKEPDGKIDAYSHSVNSFIVLDTGLILSYDHSFSIILTVPLFYQGSIEGLCGNYNNNSNDEFNLGDETLKEDISTFGKYWKVQPSYWDIDINCDSSSSICSPCSSQIKQ
ILGSDKYCGFLNKPNGPLSGCYAAINPDDYFKNCKKALCDLESEAKWREMLCQIIQSYVLACQAANANVLPWRNDTFCSLECGEHSQYSICAKGCSESCKMLTDPGQCPATCAEGC
ECDNEYFFDGEGCVMMENCGCYIDGRYYKDGQQYLTQGCSQLCTCQSGYVSCQSRSCSADEECLISNEALVCNRTSWTCSIDSGLAVTFDNQTYDLSSNYFDFSFYCEPLSSSW
FRVIAGIYSCGPEPLAVGKVFISTISDWVSIAREGKAWIGGSRINLDTGSPDLIQISSRLSIRKLTDGVQVHIGEHLQILLSKEGKLKLRTTDFYLGKLCGLCGNPAWISEFPNGLRGSFS
DLWIAEEFEECPRE

comp84036_c0_seq1 (ICQQ01000007)

DDQCELYNCMKSNDQFNSIMIQKNCEYQSQSDCGNGEDFVKLPGECCGKCIENQCVVTLENGTVLLLNEGQSVPSSDDQCELYNCMKSEEQFNTIVIQKSCEYQSQSDCGIGQLY
QKLLDQCCGKCLQSSCIYQTPEGETKLMAEGEILKSEDSCTEVTCQKINDQLVASLVTKTCIYNSEKDCTLGYIYQPVPNECCGNCVQNECVFTAINGDSKTLKAGETIAETSDICT
TYTCMETNGKFSTLTEKKTCEHNSAQDCQPDEEYQKSSDQCCGQCVPTKCLLMGVDEENQILEVGGTWTSPSNPCLSYGCTKSDERIYSMLLNKTCGAVNEKDCES
GYIEVSEDGCCSTCKAPRLCGVDYSTRILEEDGCQIEEEVPYCTGVCINMPVSMTNPSEAKEKRHQECLCCDVLQSSIKTVFLKCPGGTTVEHKYHYIDKCGCTTQLCEDK

comp83999_c1_seq3 (ICQQ01000008)

MSLFLRLCVLLALACGYCKAGSFGRTFVTTFLQNNKIGLLSTQSTALYINTYNPLTNVSVIVNNVVQNITLIREISTSLEINIPLIVDVKGIANSGNVVLIKSDKDISVTAINSKGLGSDT
TIVYPVNKLGVDYYVVTPSERFGDGIKEFAVVSTEEATTVTIYLKVSVTYKSEKYGPGKALTVNLPPWTAIQISTEDLSGTRIISLKPVAVLSGHACIGSLCNHVYEQLIPVNSWGTN
FIVTPTKKTKVDVAYVMASQNKTSVQYYSDIRFKNKTILNPGEMLQIEVLAMTTLSIQASAGVLVLFYGRNLVDISGQVLIMVPPITEYCSSYSITPISSVVNVVSIIALTSGTSEIKLNG
GKIPDVSWNQILDTTYSWTTKELSPSAQQPSFTASSSFLLLSYGFSLTSTISYGSIGVCKDGIQIFIPPSCNDVKCRAREKCQMVNGIPVCVALSEANCTAMGDPHYKTLDGRYYD
FQGTCTYTIAKTCGKDTSLLNFSVEAKNENRGNTKFSYVSHVTVKILGQQSIVINVMRNDFGFVKVNSRIQRLPLILNNTISIFQSGYFMNIVTDFGLRVLYDFDALLYVFISSSYFQN
VCGICGNYDENPSNDFATPSGTQATNLTVFGESWKVNDGDNNCWHDCKGECKPCPLDIQRNSLADQKCGLITAKTGPFMQCTAKINPKPFVDNCVYDVCMNDGSKYILCQS
LAAYASACQLNNITIGEWRALFRCPLQCPANSEYKLCGSACNPTCNDAAINQSSCSGPCLETCRCNDGLVMDGGQCIPRVKCGCIFEGKYYAPNETFWEDNKCEKQCICNPNTK
KVDCQITKCRPTETCALINGIQSCYPQVYGTCIMSTKPHFLTFDGVAYDFQGTCVYLLAGLCNSTRDLVNFQVQVLNENKASGLFSYPVEVWVSIYDLSVVISRRYPGQILVNGILTN
LPFIVNKGVDRVTIYKSSSNAVLKTSSHLKVSFNWDRWVAVTVPSSYGGQLCGLCGSFNGDINNEFVMKNGQIAANPSIFGQSWNVQTVPGCTDDGKANCSNISDLEQV
FRSSKSGCGILIDKSGPFRDCHALINPESRFKDCVFDNCFYKDRQDIMCKAIAKYVVSCQVAGGRVYPWRSDTFCSPDCPKNSHYEVCTSGCVPTCPSLSPPLGCNPQC
SEGCNCNDGFILSGADCVSIPVCGCMYNGVYYRLGEVFYTSSFCNQKCNCTEGGAVECKAYACGPNEECKVVDGVQKCQPVGSAQCSVVGDSHYMSFDGRAFDFQGNCTYIL
SKTVTNQDNLVPFAINVKKEKLGNGKVTVPKLVYFEVYGYNITLQYHVKGQVMINGVYYNIPINLEAGKIWVYQHGKRVIIHTDFGVKINYNLVYHVMVTVPGNYKGQLGGLC
GNYNGDKKDDFQSPTKIVTTDAAAFGASWKVQIPGVICDNGCGSSDNPCPPCDDKKIEIFRMDNYCGFIKKVGGPLSACYNITNPDIYFNNCIFDMCSGAGDGDTLCKNIESY
AAACQAAGVNIQPWRTDILCPLTCPANSKYKLCADVCSLTCASITEYVKCPTNCSEGCECNDGFFFDGNNCVSVDNCGCFKNGQYYQPNVKVLSSDCKQACTCHPIRGLL
CVSTACAPAEDCTIQDGVAKCINKDRCMSTQCRVMETCQDQNRKAVCAPNYTGTSWCWGDPHCHTFDGYNYDFQGNCTYILCKYDGNDSDLVPFTIEEKNNNRGIVSYVKTITI
YMYGFKISLTKGEFGKVKVNDILTNLPLTLLDGKISVGISGVNAIVRTDFGLQVSYDYSWIFVVELSSSYYGLIRGLFGNFNGNVKDELITSDNRTATSILDWAKSWKANDPDPFC
MDECPELNCATCEDAKKILYEADKKCGLIKDVAGPFKECYPVIKPDSFFDNCLYDVCIKDGANQIHCQALNAYSVACQNLGVNMVDWRTPSGCILPCPANSHYEFCGNACPATCT
NPTAPSQCTKKCVEGCQCNNGFILSVDKCVPTSACGCSYNGAYYQANEAFWADENCRLLCKCDPVLGIAVCMESSCKPSERCMVTNGVRGCKPTSYCTCTGSGDPHYISFDG
KVYDFMGTCIYLLVGVNSTDRSLVSFNVKVLNHNRGGNKAVSHTKTVTLDVYNTTFTLSFVYPRRILVNGVAISLPFNFQTNKVVAYISGTQGIITTDFDVTIRYDWDDYVAVTIP
STYTNAVQGLCGNYNLNATDDFNMKDGKTASNAIIFGDSWKVGDTPGCSPACIENCPVCTEAEEKSYKGDNYCGLIINPKGPFGLCLSVINSTSYFSDCVFDACQYKGLPSSFCK
AIGRYVAACQDAGVKIPEWRKPNFCSMPCPRNTHYELCGNPCHVTCYGLSSPTDCDLPCKETCNCDDGFLLSGHKCVPIRDCGCVYQDTYYQKNDVFYPQGQCNQRCQC
GADGIVQCQSEACEPKEECKVVNGIFGCLPKESGRCVCLGDSLYISFDGLRFDFGGTCTYIFAVVKVADPKLVPFMVVIENKSYGDNPILAVPRLVVVYAYGYRVTIEQDMSTIVKIN
REIVKLPWTLEDNSIVVNQEGTNVVLQTDFGLKMLYDTVYHVVLTIPSTYRGKMGGLCGNFNDNDEDDLQLPNGQVVTNITEFGASWKEKSSKPDCTDDPEGHPKCTSAQLRMF
SAPSSCGIITDPSGPFGVCHAHINPDDNFKHCIYDSCVVDGKDDILCKSLQAYAAACQSDGIFVGLWRNSSFCSMPCPANSHYEPCQVTCEQTCSSLTNPITCSDCCFEGCQCDA
GYVSDGNICVIPKNCGCVYNGRYLSEGESIVNADCSEQCTCQSGGVTCTAFSCQDKKQCNFVNGVYKCSKEESDCLFSRKQFITFDGLSGGPFGNGAVELTSLCNSNSLSWFR
VIADIESCVGKSDSVTRIHIFLPKGLVTISTSKEVWVNGLLKVLPITFDILTASAKDNVVSIQITADLNIELSDTGNVTIHVLKSLSEMVCGACGNFNGNSPDDLISLAGESVTNILELVK
SWTARDLNSCNA

Estimated monoisotopic mass without signal sequence: 324681

comp83288_c0_seq3 (ICQQ01000009)

IHQNCEVRCDDVVSEVEKYGQCPAPPGGSNNPPKTCDLKGTCKDDFECPGNQKCCPAICKKKCTSPAPPKPGSCPPVSTETCTSYAPQLCEFDSICPGTQKCCAIENC
GIGCVNKVSQVPKRGLCPGAPPRPYECKNSDAELDCQSDQDCDEKLKCCPSHCDMKCIPPVSEVRKGFCPVEDFGSSDTCTKFVGGNCFFDTECEGGQKCCNFDCQKTC
MKIQVPKCGKCPLLTPNLQCSFKLQKQQYRCYGDSDCAPKQKCCFYGCSKECRPAEEE

comp82261_c0_seq1 (ICQQ01000010)

MFKLWNTFLIFSLLINTQCLSAPPQTNYILRLSAEYFGTGQLTERSKQNMKLPLESGGRSLSAGPLDLLVVLNGVITGVTDVINALSGDFGDVRAVGGLPAGAPALGGLVAGVPAVG
GIYGSLLNLGGLLGGARALGGLDAAPDVGGVVGAVPVVGGVVGGAPNVGGLLAEIPIIGDLLGSLLNVVGLLGSQGSHGSHGSRSGGENDKFSKLLDSLLRGENIEGSPDKGDIID
AIKGITGLKVSHLTSPKVDLCFVPHTGVHVNVHTTLDLSLNYSLTKEQINLLIEVNVSSTAKLQLDTRGDSVFVVEGCQTTIGFLDVKSANPTPLALEVALYALFSDISPIVLCPLVHLV
VKSINNDLHEVKATAPYGTIGNIHYTTSGLPFISDKFIEIYLNAALIQKGKKLTTYPFDQNISLPPMNESAKVQVGLSANFLNSVLAGLQKNGHLNTVIKDFNIYAVDDKQKLVDLNISV
PLAPVLKLQSHQSLVKLTANIHIPKIGRLVVAADLRPRFVVKNNKLKIPVTAESLELIQSPSQDIGSDLPLQEWINSIFNLNYLPKLNTVLGKALPLPLMKCTDSNVKKKIIKNLLAVSC

Estimated monoisotopic mass without signal sequence: 60012

comp83288_c0_seq2 (ICQQ01000011)

MRVTLIAFTIILGTLFLQHVNALDRSGTCPAPDFSISLLLICMRLTHDTCSTDDGCPNREKCCSDGCRMTCKPPTEKIVEKTGQCPAPPGGSNNPPKTCDLKGTCKDDFECPGN
QKCCPAICKKKCTSPAPPKPGSCPPVSTETCTSYAPQLCEFDSICPGTQKCCAIENCGIGCVNKVSQVPKRGLCPGAPPRPYECKNSDAELDCQSDQDCDEKLKCCPSHCDMK
CIPPVSEVRKGFCPVEDFGSSDTCTKFVGGNCFFDTECEGGQKCCNFDCQKTCMKIQVPKCGKCPLLTPNLQCSFKLQKQQYRCYGDSDCAPKQKCCFYGCSKECRPAEEE

Estimated monoisotopic mass without signal sequence: 34200

comp83686_c0_seq2 (ICQQ01000012)

MELRGGLLTLCLLAILPLCNAELQVSISIPSLLRSGETGKACLTVKRHTNPVDVQAVLEAGETNYTVINEQFPAGDIFKTYEIKAPVVKDATAMFLRVSAVSVNYNYTARRSVVIAPAG
NVAFIRTDKETYKGNQDVQIAVMAMDNNLRPVEETFPLLYVTDPSGNRLHQWVNQSTNGAFLFMRFKMLDEPDLGPYSIVAERQSGSSVSKAIKVEQYILPKFNVNLKAPSMVTIL
DKNVSMTVSARYTYDQPVPGKISGRLCRPPAYYYSGNACNRNPDGICVPFTGEMDSEGTFKSSVDLSTFQMDRSGYQMIFNVQATLTEQGTNVQVTETKTIVITSQLGRAYFDRK
FMLQHFKVGIPFAGEIIVENGLGEPEIGQLVELQVVGKTVQNLTSDENGRAKFVVDTSEFTTSGVSLQLIYKNQEQCYDGNWIVPSYSNDNLYIRRFFSRTNSFMQLHGPKEDLQP
GKKYNIEAEYTFGKDALAEGENTVKFVHMIIARSKIVDSGEHSVDVSKSRQGKFNIPYETDPDHVPVVQIAIQYMLKNEVVGDAINLNLENYFKNPVSANFSSEMGTPGSNVNLDV
TATPESACVTHVYDSSLQLLDPGQSFTAATVFNSLKFNQLHGYTIEGLNVDPPSPACIDVNKQILLDGLYWNPVNFPEEDDVSKIFRDIGMHLITDTPFKQPELCSQPFERQPMFFN
SQPGFLTSMADAGPRGTATATFMAAIDGGTPQITSMRTYFPELWLYGPTFVGTNGSTQVPLEVPGTITKWKSDVMCLSNNTGFGMTKYPANFTSFQKMFVDVALPDSIVRGETMV
LVASIANYLDKCTKVNVTLHSSNGYVAEPVDTDTVKCVCSQERVSYSWKINATAIGKSDVTVTAETIHIGESCEGSADLNEPSRKDTIVKSFNVEPEGIKQEVTKTDLVCVKGENS
VIQVNVNPAGRIIPGSMTAKVQVIGDLLGRALTNPEALIVQPSGCGEQNLAKLMPIVKAIEFLNLTGRLTEEIRQRAVQNMAIGYIRQMSYRNSDGSFSAFSRGKSSSWLTLLSFKTL
WAIKDYTFVDDNIFKQGVINLERLQDLDGRFKPVGTLFNSGLKGGADDDVGFTCAVVSFLSETAYAASPSLLRPALFYLDAASRKEQSLYNYAMLFYTFRVTGNVERANAMHEK
LKSMAVNEGDSLYWERKVRPKKPTPYLFSPPPPSAEMEITATVLKAMAYGETSSTVSKDKLNEMTKISTWLVRQQNIHGSYSTTSDTVAVLDALCDYGALVFQNDATNTVQLKS
GDQVIREFKVDQPSRILLQSEALPQVPGNYSMEVSGNGCVLLQTGVVFYTPFKQEDSAFIMSVTTFPENCKNGVAYSVDIQVNASYAGIQNKSNMVVMDIPLLSGYSMG
SNSLLKLQSTFPKAEVKNNHLIVYFDDMTNETVSFNVTQDMGSRVDNFQQRTVKVFDYYEKEENGASPLFHPCTL

Estimated monoisotopic mass without signal sequence: 159112

comp83288_c0_seq1 (ICQQ01000013)

MRVTLIAFTIILGTLFLPHVNAGDKPGTCPDAEFNPLNKWLVCGPILMSGDKCSNDDGCLNREICCSNGCWKSCKPPIEKIVVEKDGQCPAPPGGSNNVPKTCDQKDACKDDFH
CAGNEKCCPSLCSKKCVEPAAPKPGSCPFVNTDMCTSFAPQRCNFDSNCPDTKKCCFLSKCEVGCADVVSQERDGQCPAPPGGSINVPKTCDKADVCKIDAECPKGQKC
CRALCSNQCVAPAPPKPGSCPPVSTETCTSYAPQLCEFDSICPGTQKCCAIENCGIGCVNKVSQVPKRGLCPGAPPRPYECKNSDAELDCQSDQDCDEKLKCCPSHCDMKCIP
PVSEVRKGFCPVEDFGSSDTCTKFVGGNCFFDTECEGGQKCCNFDCQKTCMKIQVPKCGKCPLLTPNLQCSFKLQKQQYRCYGDSDCAPKQKCCFYGCSKECRPAEEE

Estimated monoisotopic mass without signal sequence: 45318

comp83850_c0_seq1 (ICQQ01000014)

MWALVTSVLCLMTSLTHTEETCPEIKLVGLEGSDKLAILRGCPGLPGPPGEIGEAGVRSEKAARNCKELLNSGNFLSGWYTINPTGAKSINVFCDMDTEGGGWIVFQRRSDGT
VNFFRDWKDYKRGFGNQLSEFWLGNDNIYTLTSSGTQQLRVDLIDFEYNPTFAAYASFAILGEQDNYKLKLGAFTRGSAGDSLTYHNNRPFSTKDKINHDDGSIRNCAAEYKGAW
WYRDCHESNLNGLYLRGKHDSYANGINWASGKGYYYSYKISEMKFRPV

Estimated monoisotopic mass without signal sequence: 28931

comp84464_c0_seq3 (ICQQ01000015)

MRKPMKTKVFSLVNDNLPAGHTLCLFYLTVLYGVSAAQPKLNYVVLLPAELHYPSTEKACLLLDRAVGISKYSITVTLEHSGGSVVLFPKEGCPKLPLSCCSFIVPPPSKGNSEVAT
VKIQATGDITFVETNTVLIQKINDGIVIQTDKAAYKPGDTVLFRIVTLNDNFLCSKTAYNVVELKDPKNNRIAQWLKTPTKMCIAELSFNISSDPILGTYTITVENSQMQTFIVREEVL
PRFEVTIEEPTKIYVTEPTFQLKVCGTYTYGKGVVGSVSVQLCNANTWNLWSWRGTSCTYLNGEADKTGCFSTMTQPTGLNEDYNIEISANFTERDTGVSFSAKKVIPVESSPETI
IYEEFATYYRRGYPFYIRLFVQDRNGLPVPDKKVRIFYGESNIYKYNYTDKSGKVTFILDTSDWINQVTVQPYLGDSDIVLPSRTVSPLYSEAKSSLRLEPFYRVIPCGNTET
VQVQYVINPLDLEEGTKSVNFYYTVVGKAGILSYETQAIQLRDSSSLNGTLRVPITFTQQFGAAPKMIGFLILKNGNVVADRLFIDVEVCFPNPATLWFKSSESSPKAEINLTLSS
SAGSVCALRAVDKSVQIQFQDKELTRELIFNLFPNSVRGGYPDEVNEDKFEQCFFSSLFLDVFSLFQDVGLKILTNMPIKKPEELQICPFIISADFEIFEDQQKEPEPELVPVSTSQDV
VRKFFPETWLWTLVEIGEINSTSITVTVPDTITKFNARTFCMGNSGFGLSSEVSLIVFKPFFLDLAMPYSVIQGETLTLKALVFNYLTQCLKVQVTLLSSPNFTIQDCKGCVYIKCI
CADESANFIWNITANKIGFVPLTVRAEAIKSTDTCAGKVVYVPPTGNLDILEKPLLVKPQGIKKELAQSMFICLKDPDNNIEKPFSLEVPKNLVKDSESAYVSVIGDVLGTALQNLDR
LITMPSGCGEQNMLTMAPIIYVLGYLSATGQLEASLKDKALSYLQSGYQRELSYKHTDGSYSAFGESDGEGGTWLTAFVMKCFYQAKKYIFIDDQILDQALKWLASGQNNNGCFT
NRGRLIHTTMKGGVNDDVSLCAYITAALLERGGSVQPIKLDLALSNLRAKVDGISNPYTLALLAYTFTLANDIVTRQKLLSRLNLLSQSSGTDLYWTYTSSSSEGSLSADVELTAYVL
LALISGSAPTLEELTTASRIVSWLSKKQNPYGGFSSTQDTVVAIQALAEYTKVTFNPNSNLEVTVSANEEFLYRFMVDNTNRLLLQKGPLTTIPGNYQLNISGNGCVLIQIVSKYNVL
PLPEKLPFTISAKIDNCGIKYGQYGQLIVIVRYDGPRKTTNMVLIEVELLSGFSVTEDTKSLLLRSPFVKRVDFSKGLITIYLDELNHETQRYYIDLVQDLLVYDLKPVNVNIFDYYQ
KEESATTTYSPCPSPIRKS

comp83677_c1_seq2 (ICQQ01000016)

MFRQSSLFLIFGVVLSLAHLSSCQNWETFQKKHLTSNLKVDCDKEMRKALFDCKSRNTFIYSLPGPVQALCRGIAGPRNTLSRDMFTLIDCIEKSKHCQYKDKTSKNVICITCE
HELPVHFVGVGSC

Estimated monoisotopic mass without signal sequence: 11879

comp79294_c0_seq2 (ICQQ01000017)

MYSIYISENPSASPLTCAGYTAAFVSPSSRNKPITMAALKVFFLLLSVIPLISCNPVPDGQPILLDRRCPIGAVWSCVSPCVPGKKCLVVCEFKCRCTQIGYVIQRDHCVPALPTELKTV

comp76542_c0_seq1 (ICQQ01000018)

MKFQIILWMWILHTRLSKSEDDTRKCYQSEVPNGYQSEQYGQSRSLSNRNEVVKLKENCMCSFWGNFNYRTFDGNIFRFPGMCSYLLSSHCGDVFEDFNIQIQQGLVDMLPSI
IRLSVKLSEVLIEVVDRVPMIKGQRLQLPYNANGIFIEKEDNILRIRSKIILELAWADENYLELTLNKKYKGKVCGLCGNFNGEANDDISFAGAKVDPLLFGELYKLRKPGEVCKDSK
PESTAKCSIYRTESLKNLKDSKWATCNKLVNPDPFVEAFMMDMCLCKDSDRLSTFCLCSTLTQYSRQCALAGGKPPEWRKQTICYKSCTNNLIYKECGIPCTSTCSNPGRQYFCD
GACIPKCTCPEGHVVDDITNSGCVRIEKCSCIHNNKTYPPGSYYTTRCSSCICSDGYWKCEEIPCPSVCSVEGGSHITTFDKTQYIFNGKCTYVLTKPCQTDAFLVTMDIKKCPITE
APTCLSRVLLILKEASYEFEIKFDSSNFPYVTYDIDAMTKANVTVCWATSFYMIVYTNLGFYLEVQFTPIMQLYIIMDQSYKGKLCGLCGNYNFIQKDDFISSSGVLESNAIDFANTWK
THYQCRNIVPDHTDPCMYGIDIEKYALYWCQLLLDLKGPFSPCFTLVSPSQYYQNCLFDTCSSKRAEDQLCAVLSSYAWVCAANGIVLTGWRVDVCSSYAKSCPETFVYSYN
VTTCMPTCRSLAEPNTTTNITFSTVDGCICPEETYLNDDDGICVKKEECPCYYKGVPVHFDESLHVGQKICTCNKGKLNCIQPPIPECPPPNIYFDCDKAGVYANGVQCSQSCDTT
KPKCYSPLCVSGCLCPKGLLLTYNGTCVMEEQCPCVHNGEHYQPGETVKMHCETCTCRNRSWVCDSQPDKDICTVYGQGHYSTFDNKFFTSSGHCQYTLVQDYCNLDN
PRRGTFKIIAENIQCGMSESTCSVAITLFLGAHILFLSDGNVDVVARPSNVPIPYLIRQMGLYLVVETNIGLVLIWDQLTTIFIHLDVRYKGKVCGLCGNFDGNSENDFLTRSQCLVED
VKEFIESWKASANCLDVYTTKDPCVIHPHRMPWAQKMCHIIISNVFAKCHAEVNPVKYYEDCVKDTCSCDSGGDCNCYCTAVAAYAQACNKVCVCIEWRSPTVCPLFCEGYDNEL
CQWHYTACHVPCLKTCKNTEGICQSYIQGLEGCYPKCPKNKPYFNEEIMQCVADCGCLDNHSNHYNLGEKIESCSACERCICSENGIQCRYDSNECGCEYQGEIIQIGETVQL
DEGSYGCTLVKCGINGTEITPCYEP

Estimated monoisotopic mass without signal sequence: 142628

comp83857_c2_seq2 (ICQQ01000019)

MYTTWALAVIATIVLGFPFIFTVEVGICDLPVDEGSCDKYVHRYYYDNEQGTCNKFLYGGCEGNANRFSTLKNCEEACKPARICGLPKIIGPCKALERRFYYNNKKKMCKLFPYG
GCKGNANNFLSKEDCERFCKPAH

comp82467_c0_seq1 (ICQQ01000020)

MKQLVIMVIGASILAGCHAGCFISLPESPKTGVRPTGCRYMGVIHKLGTEWRTKGCMDCICRLDGSLTCCPAYSIPVEYDQENCKLVFDLKSCSYNVIPIKDPTLICEVRAAVA

Estimated monoisotopic mass without signal sequence: 10463

comp83603_c0_seq1 (ICQQ01000021)

MSLVTGIVLLLCCTAVCGKLGTCPDDVDYPRCEDSMLIFQHECTTVDQDCDGERKCCFSGCRRRCLLPLGAKNGDCPYFNHAKCIHIKPSPYECHDDNQCQGTDRC
CYYGCRMQCKATLTVKPGQCPIPLTLCVSDTLPPPLCKSDSECEGKKKCCIKFCRRECYDPRP

Estimated monoisotopic mass without signal sequence: 17131

comp79314_c0_seq3 (ICQQ01000022)

MTFLHTVTLLWILCGVSAHLKDYENVALRGRATQSSIIYDEKYGYLSAAINAIDGNADPDHSHGSCSHTHNIYLSPWWRVDLLKSYKVAYITITNRGDCCSDRLNGAEILVGDSLS
DNGNNNPRCGLITSIPAGGTQSFPCNGMVGRYVNIVLRGKQDYLQLCEVQVWATECSESP

Estimated monoisotopic mass without signal sequence: 17369

Boxed sequences are signal peptide sequences predicted by SignalP 5.0 (Armenteros et al., 2019). Overlapping identified proteins in Table 3 are not shown in Table 5.

shown in Table 3. Comp84488_c0_seq1, comp83308_c0_seq2, comp81644_c2_seq1, and comp79296_c0_seq1 showed 53.9%, 50.0%, 30.0%, and 21.7% sequence identities with extracellular superoxide dismutase (Cu-Zn)-like (antioxidant defense in *Nanorana parkeri*), vitelline membrane outer layer protein 1 homolog (antimicrobial barrier in *X. tropicalis*), ranasmurfin (blue-colored protein protecting from UV in *P. leucomystax*), and alpha-1-antichymotrypsin (proteinase inhibitor in *Sorex araneus*), respectively. Moreover, we utilized web-based software (SignalP 5.0, https://services.healthtech.dtu.dk/service.php?SignalP-5.0) to predict the presence of signal peptides and the location of their cleavage sites in proteins (Armenteros et al., 2019). As shown in Table 3, the 18-amino acid peptide (MHCVSFILLVYLSCAVQA) in comp84488_c0_seq1, 19-amino acid peptide (MLLSWIVTLFLLHTKLAHC) in comp83308_c0_seq2, 17-amino acid peptide (MKAILLLALFGLSFCIG) in comp81644_c2_seq1, and 19-amino acid peptide (MKALLLLCASAALVHLAAS) in comp79296_c0_seq1 were confirmed to be signal peptides in accordance with the data of Edman degradation. The estimated monoisotopic mass without signal sequence of comp84488_c0_seq1, comp83308_c0_seq2, comp81644_c2_seq1, and comp79296_c0_seq1 was 19725, 17627, 14299, and 25062, respectively, as shown in Table 3.

LC-MS/MS analyses

Since aggregation of the foam nest occurred after freeze-thawing and cysteine content became somewhat higher, reductive alkylation was initially done in the sample preparation process prior to protease digestion. Reductive alkylation is known to be an important step for identifying proteins with abundant disulfide bonds in high yield and for high sequence coverage (Suttapitugsakul et al., 2017). Moreover, two proteases (trypsin and lysyl endopeptidase) with different cleavage sites are frequently utilized to increase the accuracy of the protein identification. Although the foam nests produced a viscous solution initially, it became a smooth solution after the reductive alkylation (data not shown). As shown in Table 4, protein identification was conducted using the database searching software Mascot. The LC-MS/MS dataset was matched to transcriptome sequence data and the search results for the tryptic digest peptides revealed the presence of 20 proteins. Two of these identified proteins were identical to comp83308_c0_seq2, and comp81644_c2_seq1, as determined by Edman degradation. By performing a homology search (NCBI database and Xenbase), 20 proteins were tentatively classified into eight groups with the following expected activities. (i) proteinase inhibitors (ovostatin, WAP four-disulfide core domain protein 3, antileukoproteinase, alpha-2-macroglobulin-like 1, hylaserpin, and serine protease

inhibitor 2), (ii) ribonuclease (onconase), (iii) glycoproteins (hemocytin, ficolin, mucin, and fucolectin), (iv) antimicrobial protein (BPI fold-containing family B member 4 precursor), (v) immunoglobulin-binding proteins (IgGFc-binding protein, beta-microseminoprotein), (vi) glycoprotein-binding protein (vitelline membrane outer layer protein), (vii) blue-colored protein (ranasmurfin), and (viii) keratin-associated protein. The LC-MS/MS datasets from the lysyl endopeptidase digest revealed that one more protein, comp82261_c0_seq1, exhibited 29.3% identity with BPI fold-containing family B member 4 of *X. tropicalis*; it was also identified and classified as an antimicrobial protein (BPI fold-containing family B member 4), along with comp79296_c0_seq1, which was confirmed by Edman degradation. The amino acid sequences of identified proteins containing putative signal sequences of secretory proteins from foam nests examined by LC-MS/MS analyses and NGS data are summarized in Table 5.

DISCUSSION

Among the foam nest producing frogs of the Leptodactylidae and Rhacophoridae families, there is considerable diversity in their egg-laying sites and nest forms. In the case of *L. vastus* (Leptodactylidae) and the tungara frog, *Engystomops pustulosus* (Leptodactylidae), these foam nests are usually found as floating structures at the edge of temporary pools or puddles (Cooper et al., 2017). In contrast, the foam nest of *R. arboreus* (Rhacophoridae) is generated as a floating form in the air that is then attached to branches or leaves of trees in ponds surrounded by forest (Kusano et al., 2006). Therefore, it can be assumed that the foam nests would contain different components; by utilizing several kinds of analytical techniques we were able to identify 22 novel proteins in the foam nest of *R. arboreus* in this study.

The foam nest from *L. vastus* that was analyzed by SDS-PAGE was reported to possess many protein bands ranging from a few to several hundred kDa, and a dense protein band of approximately 20 kDa (Lv-RSN-1) (Hissa et al., 2008). On the other hand, the foam nest from *E. pustulosus* contained six abundant protein bands in the 10–35 kDa range (Rsn-1–6) and some fainter bands with higher molecular mass (Fleming et al., 2009). In the case here of *R. arboreus*, we detected a similar protein profile to that of *L. vastus*, with sizes ranging from a few to several hundred kDa, suggesting the diversity of foam nests produced by amphibians.

Using an amino acid analyzer, the foam nest from *R. arboreus* was found to contain both essential and non-essential amino acids, and the content of cysteine was about four times higher than that in frog meat of *R. ridibunda*. Cysteine residues and the resulting disulfide bonds were thought to be important for the formation of foam nests for the following reasons. First, hydrophobins, which have been well known as the most surface-active proteins for the attachment of fungi to surfaces, possess eight highly conserved cysteine residues that form four disulfide bonds. These disulfide bonds stabilize an amphipathic tertiary structure that results in surfactant activity (Berger et al., 2019). Second, ranasmurfin, a blue-colored protein mainly isolated from the foam nest of *P. leucomystax*, is a dimeric protein stabilized by three internal disulfide bonds (Cys 4–Cys 62, Cys 17–Cys 65, and Cys 37–Cys 101) (Oke et al., 2008). Third, the novel and predominant protein, Lv-RSN-1, derived from the foam nest of *L. vastus*, was also stabilized by four disulfide bonds (Cys 18–Cys 67, Cys 38–Cys 114, Cys 125–Cys 168, Cys 146–Cys 207) (Hissa et al., 2014). Fourth, Rsn-2, which was isolated from the foam nest of *E. pustulosus*, reduced surface tension of water the most effectively among the Rsn proteins (Rsn-1–6) (Fleming et al., 2009). The surfactant activity of Rsn-2 was eliminated or significantly reduced when some cysteine residues of Rsn-2 were converted by site-directed mutagenesis to other amino acids that could not form disulfide bonds (Cooper et al., 2017). Fifth, after treatment with reductive alkylation prevented the formation of disulfide bonds, the viscosity of the foam nest from *R. arboreus* significantly decreased. Taken together, these results suggest that cysteine residues and formation of disulfide bonds might be one of the important factors in maintaining foam nest integrity for *R. arboreus*, as well as for other amphibians.

With MALDI-MS analyses of the *R. arboreus* foam nest, several ion peaks with *m*/*z* from several hundred to about 15,000 were observed, as shown in Fig. 2B. Among them, one ion peak (*m*/*z* 14,224) was the largest. In contrast, we performed LC-MS analyses with another soft ionization technique, ESI-MS (electrospray ionization mass spectrometry), to identify proteins in the foam nest. Consequently, proteins with molecular masses 19,486, 12,600, and 14,222 Da were detected in Fractions 1, 2, and 5, respectively. Transcriptome analysis revealed that the proteins in Fractions 1, 2, and 5 corresponded to the proteins encoded by comp84488_c0_seq1, comp83308_c0_seq2, and comp81644_c2_seq1, respectively. On the basis of the amino acid sequence, the theoretical molecular masses of comp84488_c0_seq1, comp83308_c0_seq2, and comp81644_c2_seq1 without signal sequence were larger than those of LC-MS data. A relatively likely possibility to explain the discrepancy was thought to be proteolysis, but other possibilities, such as post-translational modifications, could not be excluded. N-linked glycans are attached to the side chain amide group on an asparagine residue as one of possible post-translational modifications. A common sequence surrounding such modified asparagines is known to be Asn-Xaa-Ser/Thr, where Xaa may be any amino acid residue (Mathews et al., 2000). Unfortunately, there was no such sequence in the proteins encoded by comp84488_c0_seq1, comp83308_c0_seq2, and comp81644_c2_seq1. On the other hand, protein kinases, such as Ser/Thr protein kinases and Tyr kinases, are ATP-dependent enzymes that add a phosphorylation group to the -OH group of a Tyr, Ser, or Thr in the target proteins (Mathews et al., 2000). Phosphorylation sites of target proteins could be predicted using online web resources, such as NetPhos 3.1 server (http://www.cbs.dtu.dk/services/NetPhos/) and GPS 5.0 (http://gps.biocuckoo.cn/). These predictions suggested some phosphorylatable amino acid residues in the proteins encoded by comp84488_c0_seq1, comp83308_c0_seq2, and comp81644_c2_seq1. Since one phosphorylation site at each amino acid increases the molecular weight by 80, multiple phosphorylations might be expected in these proteins. In addition, comp81644_c2_seq1 (ICQQ01000003) might be an ortholog of ranasmurfin in *R. arboreus*. Ranasmurfin has been extracted from foam nests with a blue/green pigmentation produced by other frogs. This

coloration comes from two histidine sidechains coordinating the zinc to form the blue chromophore (Cooper et al., 2010, 2017). Moreover, it has a unique modification at Tyr and Lys residues for dimer formation, affecting the molecular weight (Oke et al., 2008). Although its function is not well understood, it has been thought to work as sunscreen to protect eggs and embryos. We isolated a similar protein, comp81644_c2_seq1, from the foam nests of *R. arboreus*, as indicated by LC-MS, Edman degradation, LC-MS/MS, and transcriptome analyses, as shown in Table 3 and Fig. 3. Although this novel protein showed 30.0% sequence identity with ranasmurfin, the color of *R. arboreus* foam nests is white, and comp81644_c2_seq1 might be a monomer, based on the results of LC-MS (Fig. 3). Comparison of the amino acid sequences of ranasmurfin and comp81644_c2_seq1 was further performed by protein BLAST. Ranasmurfin had three internal disulfide bonds (Cys 4–Cys 62, Cys 17–Cys 65, and Cys 37–Cys 101) and the amino acid sequences around the six cysteine residues were well-conserved in comp81644_c2_seq1, suggesting the stabilization by similar disulfide bonds. Moreover, Tyr 2 and Lys 31, which form an unusual linkage between subunits in ranasmurfin, were not conserved in comp81644_c2_seq1. Further studies are needed to clarify these differences between ranasmurfin and comp81644_c2_seq1 (ICQQ01000003).

Cooper et al. (2007) reported a stable surface structure of the foam nest from the tungara frog, *E. pustulosus*, consisting of surfactant proteins, lectins, and polysaccharides, and small-angle neutron reflection experiments on the nest foam showed a complex air-water interface layer of about 7.5 nm thickness. In this study, we isolated 22 novel proteins from the foam nest of *R. arboreus*. To make the foam nest an effective cradle for the *R. arboreus* eggs, protecting them from microbial and parasitic attack, oxidative stress, and a shortage of moisture, these proteins should be expected to function as superoxide dismutase (antioxidant defense), proteinase inhibitors, ribonuclease, glycoproteins, antimicrobial proteins and barrier, immunoglobulin-binding proteins, glycoprotein binding proteins, colored protein (protection from harmful UV rays), and keratin (a part of cytoskeleton)-associated proteins, as indicated by their sequence similarities. Further studies examining the protein expression and physiochemical properties of these 22 novel proteins, and elucidation of the true identity of the higher-molecular weight bands seen in SDS-PAGE analyses, are clearly needed for a better understanding of this protected species and its unusual egg-laying behavior.

ACKNOWLEDGMENTS

We are grateful to Dr. Leslie Sargent Jones (Appalachian State University, Retired) for her careful reading of our manuscript and Dr. Takeshi Kasama (Tokyo Medical and Dental University) for his technical assistance with MALDI-MS. This work was supported by JSPS KAKENHI Grant Number 18K05346 (to YS).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

YS designed this study and wrote the manuscript. MN, NT, and YY performed LC-MS and Edman degradation analyses. HK collected *R. arboreus* and isolated its organs. KU, YH, CB, and YT analyzed data. TK and HI performed transcriptome (RNA-seq) analysis, and LC-MS/MS analyses, respectively. All authors contributed to the final manuscript.

REFERENCES

Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotech 37: 420–423

Berger BW, Sallada ND (2019) Hydrophobins: multifunctional biosurfactants for interface engineering. J Biol Eng 13: 10

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120

Brandani GB, Vance SJ, Schor M, Cooper A, Kennedy MW, Smith BO, et al. (2017) Adsorption of the natural protein surfactant Rsn-2 onto liquid interfaces. Phys Chem Chem Phys 19: 8584–8594

Cagiltay F, Erkan N, Selcuk A, Ozden O, Devrim Tosun D, Ulusoy S, et al. (2014) Chemical composition of wild and cultured marsh frog (*Rana ridibunda*). Bulg J Agric Sci 20: 1250–1254

Coe M (1974) Observations on the ecology and breeding biology of the genus *Chiromantis* (Amphibia: Rhacophoridae). J Zool 172: 13–34

Cooper A, Kennedy MW (2010) Biofoams and natural protein surfactants. Biophys Chem 151: 96–104

Cooper A, Vance SJ, Smith BO, Kennedy MW (2017) Frog foams and natural protein surfactants. Colloids Surf A Physicochem Eng Asp 534: 120–129

Fleming RI, Mackenzie CD, Cooper A, Kennedy MW (2009) Foam nest components of the tungara frog: a cocktail of proteins conferring physical and biological resilience. Proc R Soc B 276: 1787–1795

Furness AI, McDiarmid RW, Heyer WR, Zug GR (2010) Oviduct modifications in foam-nesting frogs with emphasis on the genus *Leptodactylus* (Amphibia, Leptodactylidae). S Am J Herpetol 5: 13–29

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotech 29: 644–652

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protoc 8: 1494–1512

Haridy M, Tachikawa Y, Yoshida S, Tsuyuguchi K, Tomita M, Maeda S, et al. (2014) *Mycobacterium marinum* infection in Japanese forest green tree frogs (*Rhacophorus arboreus*). J Comp Path 151: 277–289

Hellsten H, Harland R, Gilchrist M, Hendrix D, Jurka J, Kapitonov V, et al. (2010) The genome of the western clawed frog *Xenopus tropicalis*. Science 328: 633–636

Hissa DC, Vasconcelos IM, Carvalho AF, Nogueira VL, Cascon P, Antunes A, et al. (2008) Novel surfactant proteins are involved in the structure and stability of foam nests from the frog *Leptodactylus vastus*. J Exp Biol 211: 2707–2711

Hissa DC, Bezerra GA, Birner-Gruenberger R, Paulino Silva L, Uson I, Gruber K, et al. (2014) Unique crystal structure of a novel surfactant protein from the foam nest of the frog *Leptodactylus vastus*. ChemBioChem 15: 393–398

Hissa DC, Bezerra WM, de Freitas CDT, Ramos MV, Lopes JL, Beltramini LM, et al. (2016) Frog foam nest protein diversity and synthesis. J Exp Zool A Ecol Genet Physiol 325: 425–433

Kabisch K, Herrmann HW, Klossek P, Luppa H, Brauer K (1998) Foam gland and chemical analysis of the foam of *Polypedates leucomystax* (Gravenhorst 1829) (Anura: Rhacophoridae). Russ J Herpetol 5: 10–14

Kaneko Y, Matsui M (2004) *Rhacophorus arboreus*. The IUCN Red List of Threatened Species. URL: https://www.iucnredlist.org/species/58973/11863497

Kasuya E, Kumaki T, Saito T (1992) Vocal repertoire of the Japanese treefrog, *Rhacophorus arboreus* (Anura: Rhacophoridae). Zool Sci 9: 469–473

Kusano T, Sakai A, Hatanaka S (2006) Ecological functions of the foam nests of the Japanese treefrog, *Rhacophorus arboreus* (Amphibia Rhacophridae). Herpetol J 16: 163–169

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359

Lu JR, Zhao XB, Yaseen M (2007) Biomimetic amphiphiles: biosurfactants. Curr Opin Colloid Interface Sci 12: 60–67

Mathews CK, van Holde KE, Appling DR, Anthony-Cahill SJ (2013) Biochemistry. 4th ed. Pearson Education Inc., Upper Saddle River, NJ

McMahon SA, Walsh MA, Ching RTY, Carter LG, Dorward M, Johnson KA, et al. (2006) Crystallization of ranasmurfin, a blue-coloured protein from *Polypedates leucomystax*. Acta Crystallogr Sect F 62: 1124–1126

Muto K, Kubota HY (2009) A novel mechanism of sperm motility in a viscous environment: corkscrew-shaped spermatozoa cruise by spinning. Cell Motil Cytoskeleton 66: 281–291

Muto K, Kubota HY (2011) Ultrastructural analysis of spermiogenesis in *Rhacophorus arboreus* (Amphibia, Anura, Rhacophoridae). J Morphol 272: 1422–1434

Okada Y, Kawano U (1924) On the ecological distribution of two new varieties of *Rhacophorus* in Japan. Zool Mag 36: 104–109

Oke M, Ching RTY, Carter LG, Johnson KA, Liu H, McMahon SA, et al. (2008) Unusual chromophore and cross-links in ranasmurfin: a blue protein from the foam nests of a tropical frog. Angew Chem Int Ed 47: 7853–7856

Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. (2016) Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature 538: 336–343

Suttapitugsakul S, Xiao H, Smeekens J, Wu R (2017) Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. Mol Biosyst 13: 2574–2582

Velazquez A, Reyes A, Chargoy J, Rosado A (1977) Amino acid and protein concentrations of human follicular fluid. Fertil Steril 28: 96–100

Yokoe M, Takayama-Watanabe E, Saito Y, Kutsuzawa M, Fujita K, Ochi H, et al. (2016) A novel cysteine knot protein for enhancing sperm motility that might facilitate the evolution of internal fertilization in amphibians. PLOS One 11: e0160445

(Received July 8, 2020 / Accepted September 8, 2020 / Published online December 7, 2020)