

Two Novel Heat-Soluble Protein Families Abundantly Expressed in an Anhydrobiotic Tardigrade

Ayami Yamaguchi¹, Sae Tanaka¹, Shiho Yamaguchi¹, Hirokazu Kuwahara¹, Chizuko Takamura², Shinobu Imajoh-Ohmi², Daiki D. Horikawa^{1‡}, Atsushi Toyoda³, Toshiaki Katayama⁴, Kazuharu Arakawa⁵, Asao Fujiyama^{3,6}, Takeo Kubo¹, Takekazu Kunieda^{1*}

1 Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan, **2** Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **3** Center for Genetic Resource Information, National Institute of Genetics, Mishima, Shizuoka, Japan, **4** Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **5** Institute for Advanced Biosciences, Keio University, Fujisawa, Kanagawa, Japan, **6** Principles of Informatics Research Division, National Institute of Informatics, Tokyo, Japan

Abstract

Tardigrades are able to tolerate almost complete dehydration by reversibly switching to an ametabolic state. This ability is called anhydrobiosis. In the anhydrobiotic state, tardigrades can withstand various extreme environments including space, but their molecular basis remains largely unknown. Late embryogenesis abundant (LEA) proteins are heat-soluble proteins and can prevent protein-aggregation in dehydrated conditions in other anhydrobiotic organisms, but their relevance to tardigrade anhydrobiosis is not clarified. In this study, we focused on the heat-soluble property characteristic of LEA proteins and conducted heat-soluble proteomics using an anhydrobiotic tardigrade. Our heat-soluble proteomics identified five abundant heat-soluble proteins. All of them showed no sequence similarity with LEA proteins and formed two novel protein families with distinct subcellular localizations. We named them Cytoplasmic Abundant Heat Soluble (CAHS) and Secretary Abundant Heat Soluble (SAHS) protein families, according to their localization. Both protein families were conserved among tardigrades, but not found in other phyla. Although CAHS protein was intrinsically unstructured and SAHS protein was rich in β -structure in the hydrated condition, proteins in both families changed their conformation to an α -helical structure in water-deficient conditions as LEA proteins do. Two conserved repeats of 19-mer motifs in CAHS proteins were capable to form amphiphilic stripes in α -helices, suggesting their roles as molecular shield in water-deficient condition, though charge distribution pattern in α -helices were different between CAHS and LEA proteins. Tardigrades might have evolved novel protein families with a heat-soluble property and this study revealed a novel repertoire of major heat-soluble proteins in these anhydrobiotic animals.

Citation: Yamaguchi A, Tanaka S, Yamaguchi S, Kuwahara H, Takamura C, et al. (2012) Two Novel Heat-Soluble Protein Families Abundantly Expressed in an Anhydrobiotic Tardigrade. PLoS ONE 7(8): e44209. doi:10.1371/journal.pone.0044209

Editor: John R Battista, Louisiana State University and A & M College, United States of America

Received: May 11, 2012; **Accepted:** July 30, 2012; **Published:** August 28, 2012

Copyright: © 2012 Yamaguchi, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding: This work was supported in part by a Grant-in-Aid for Young Scientists (B) 21710223, and a Grant-in-Aid for Scientific Research on Priority Areas, Comparative Genomics 20017010 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kunieda@biol.s.u-tokyo.ac.jp

‡ Current address: Universite Paris Descartes, Inserm U1001, Paris, France

Introduction

Water is essential for life and most animals cannot survive without water. Some organisms, including tardigrades, however, are able to tolerate an almost complete loss of water by entering a metabolically inactive state, referred to as anhydrobiosis, and they can resume their activity upon rehydration [1], [2]. Dehydrated tardigrades showed extraordinary tolerance against various physical extremes including exposure to space [3–6], but the molecular basis of these tolerant abilities is totally unknown. The anhydrobiotic ability was observed in several species belonging to four animal phyla; arthropods, nematodes, rotifers and tardigrades. In anhydrobiotic arthropods and nematodes, trehalose has long been suggested to have an important role in desiccation tolerance because it accumulates in large amounts (~15%–20% of body weight) upon desiccation [7–9]. In contrast, accumulation of trehalose was much less in tardigrades, varying from 0% to at most

2.9% (less than 1% in most species) [10–12], which suggests that tardigrades have other factors to tolerate dehydration. Another candidate molecule is the late embryogenesis abundant (LEA) protein family. LEA proteins were originally identified as abundant proteins in maturing plant seeds and their expression was significantly induced by desiccation in anhydrobiotic animals other than tardigrades [13–15]. The LEA proteins maintain their solubility even after heat-treatment, and are proposed to prevent protein-aggregation by interfering close association of damaged proteins as ‘molecular shield’ in a dehydrated condition [16–18]. Although the presence of LEA-like transcripts and protein were detected by expressed sequence tag (EST) analyses or proteomics of tardigrades [19–22], the induced expressions by desiccation and their biochemical property including heat-solubility have not been clarified and thus, their relevance to desiccation tolerance is obscure in tardigrades.

Here, to elucidate the molecular basis of tardigrade anhydrobiosis, we utilized the heat-soluble property characteristic of LEA proteins and searched for major heat-soluble proteins from an anhydrobiotic tardigrade, *Ramazzottius varieornatus*. We previously showed *R. varieornatus* have high tolerant ability against desiccation [23] and its genome sequences have been determined by our group. Thus this species is a suitable model for molecular analysis of tardigrade tolerant abilities. Our heat-soluble proteomics identified five abundant heat-soluble proteins forming two novel protein families with distinct subcellular localizations. No LEA proteins were detected. Both protein families were conserved among tardigrades, but not found in other phyla. Water-deficient conditions induced conformational changes of proteins in both families to an α -helices as LEA proteins. Two conserved repeats of 19-mer motifs in CAHS proteins were capable to form amphiphilic stripes in α -helices, suggesting their roles as molecular shield in water-deficient condition. Tardigrades might have evolved novel protein families different from LEA proteins and this study revealed novel repertoire of major heat-soluble proteins in anhydrobiotic animals.

Materials and Methods

Animals

We used a strain, YOKOZUNA-1, which was established from a single individual of *R. varieornatus*. The same strain was used for our ongoing genome sequencing project of *R. varieornatus*. Tardigrades were reared on agar plates by feeding algae, *Chlorella vulgaris* purchased from Chlorella Industry Co., Ltd. (Japan) as described previously [23].

Heat-soluble proteomics

Tardigrades were starved for 2 days to eliminate food contamination, and then 500 individuals were homogenized well manually using disposable plastic pestle (Kontes) in 40 μ l ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.76 mM KH_2PO_4 , pH = 7.4) with protease inhibitors, Complete (Roche). 'Pre-heat' soluble protein fraction was collected after centrifugation at 12,000 rpm for 20 min and heated at 92°C for 15 min. The heat-soluble fraction was separated as supernatants by centrifugation at 12,000 rpm for 20 min. Proteins were analyzed by SDS-PAGE and detected with Flamingo Fluorescent Gel Stain (BioRad).

Protein identification

Heat-soluble protein bands were excised and digested in trypsin. Tryptic peptides were separated by nanoflow LC and analyzed using a MALDI-TOF/TOF 4800 proteomics analyzer (Applied Biosystems). Corresponding peptide sequences were retrieved from 6 frame translations of our in-house draft genome assembly of YOKOZUNA-1 strain using MASCOT software (Matrix Science). The draft genome assembly used for this study covered more than 90% of estimated genome size and will be published elsewhere. Rapid amplification of cDNA ends (RACE) was performed to determine full-length cDNA sequences using the First-Choice RLM-RACE Kit (Invitrogen). The nucleotide sequences reported in this study have been submitted to DDBJ with accession numbers AB650497-AB650501.

Sequence analyses

Subcellular localizations and signal peptides were predicted by TargetP and SignalP programs, respectively [24], [25]. Sequences of other tardigrade homologs were retrieved from NCBI databases of EST and transcriptome shotgun assembly (TSA). EST

sequences for CAHS and SAHS gene homologs were assembled using Codon Code Aligner (Codon Code Corp.). Accession numbers used in these assemblies were listed in Tables S1, S2, and S3. ESTs coding actin or elongation factor 1 alpha (EF1 α) were retrieved by querying the sequences of *Drosophila melanogaster*; NP_727048 for actin and AAF58608 for EF1 α . Accession numbers for ESTs of actin and EF1 α were listed in Table S4. Multiple alignments and sequence Logo analyses were performed using Geneious software (Biomatters Ltd.).

Heat-solubility assay of recombinant proteins

Recombinant proteins were expressed in *Escherichia coli* BL21(DE3) strain using pET system (Novagen) in fusion with a His-tag at the N-terminus. Signal peptides were excluded for the secretory proteins. After bacterial cells were lysed in PBS with protease inhibitors by sonication, the soluble fraction was collected and served as pre-heat sample. After heat-treatment, heat-soluble and heat-insoluble fractions were separated by centrifugation and analyzed by SDS-PAGE.

Subcellular localization

To utilize the ease in visualization of subcellular components and high transfection efficiency, we used adherent mammalian cells. For expression of green fluorescent protein (GFP)-fused proteins, coding sequences were inserted into pAcGFP1-N1 (Clontech). Human embryonic kidney 293T cells were transfected with the construct using X-tremeGENE 9 reagent (Roche). After 24 to 48 hours incubation in Dulbecco's Modified Eagle Medium with 10% fetal bovine serum, cells were stained with Hoechst 33342 (Lonza) and MitoTracker Red CMXRos (Lonza), to visualize nuclear DNA and mitochondria respectively. Fluorescent signals were observed by confocal microscopy LSM710 (Carl Zeiss).

Immunoblot analysis

Transfected cells and culture medium were separated by centrifugation and subjected to SDS-PAGE. GFP-fusion protein was reacted with anti-GFP rabbit polyclonal antibody (Invitrogen) and detected by chemiluminescent reagent LumiGLO (Cell Signaling Tech.).

Circular Dichroism (CD) spectroscopy

His-tagged protein was expressed using a pET system (Novagen) and purified with Ni-NTA agarose (Invitrogen). After dialysis to remove the imidazole, the CD spectra were measured using an AVIV 202SF CD spectrometer.

Results

Novel abundant heat-soluble proteins in tardigrades

Because LEA proteins maintain their solubility after heat treatment and are believed to have an important role in anhydrobiosis, we searched for major heat-soluble proteins from an anhydrobiotic tardigrade, *R. varieornatus*. The soluble protein fraction of the tardigrade lysate served as the pre-heat sample and was heated at 92°C for 15 min. Heat-soluble proteins were recovered as supernatant after centrifugation and analyzed by gel electrophoresis. As shown in Fig. 1A, three major bands, termed B1 to B3, were detected selectively in the heat-soluble fraction, though most other proteins disappeared after heat-treatment even in the insoluble fraction, possibly due to the severe aggregation caused by heat, which prevented separation of the proteins by gel electrophoresis. These heat-soluble bands were detected with

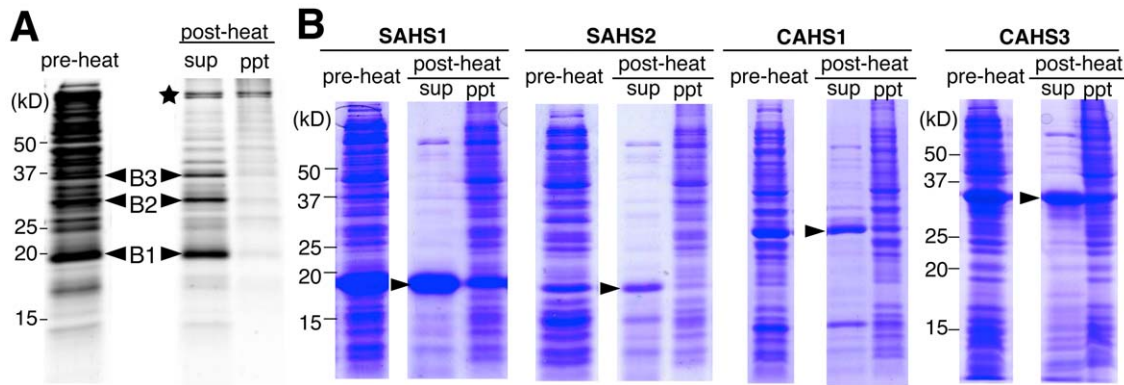


Figure 1. Heat soluble proteomics. (A) The soluble fraction of tardigrade lysate ('pre-heat') was heated and separated to heat-soluble fraction (sup) and insoluble precipitate (ppt) by centrifugation. Arrows indicate the heat-soluble bands detected selectively in the heat-soluble fraction, named B1, B2, and B3. Star indicates the partially heat-soluble bands detected in both the supernatant and precipitate after heat treatment. (B) Heat solubility of bacterial lysate expressing SAHS and CAHS proteins. Arrows indicate each SAHS or CAHS protein.
doi:10.1371/journal.pone.0044209.g001

similar intensity in pre-heat condition, suggesting that the proteins in these bands were fairly heat-soluble.

Mass spectrometry of these bands was separately analyzed and corresponding peptide sequences were retrieved from the draft genome database. Based on the retrieved sequences, we applied the RACE method and determined full length cDNA sequences for a total of five proteins. The estimated molecular weight of each protein coincided well with the mobility on the gel, suggesting the lack of post-translational modifications (Table S5). Two proteins identified from band B1 were similar to each other (37% identity). Although no Pfam motifs were found, a BLASP search against a protein database revealed low similarity with mammalian fatty acid binding proteins (FABPs) (Table S5). In contrast to mammalian FABPs, which are cytosolic and contain lipocalin motifs [26], the newly identified proteins contained no motif and instead contained a secretory signal-peptide at N-terminus (Fig. 2A). Thus, these proteins formed a FABP-related, but distinct secretory protein family. We designated this protein family the Secretory Abundant Heat Soluble (SAHS) protein family. Three other heat-soluble proteins identified from bands B2 and B3 also shared several conserved sequences and formed another protein family without secretory signals (Fig. 2B). These proteins showed no similarity with known sequences in BLASTP search and no known motifs were found in Pfam database, suggesting that they formed a novel protein family. We named this group the Cytoplasmic Abundant Heat Soluble (CAHS) protein family. Notably, no LEA proteins were detected in mass spectrometry of the predominant heat-soluble bands, suggesting that LEA proteins are not major heat-soluble proteins in the tardigrades.

Because our initial heat-soluble proteomics was performed using the lysate of tardigrades, certain internal solutes of tardigrades such as sugars might have contributed to the heat-solubility of the identified proteins. To examine this possibility, we expressed recombinant proteins in *E. coli* for two members of each SAHS/CAHS protein family as representatives and subjected the bacterial lysate to a heat solubility assay. After heat treatment, bacterially expressed SAHS and CAHS proteins were recovered mainly in the heat-soluble fractions (Fig. 1B). These results suggest that the heat-solubility of SAHS and CAHS proteins does not require the presence of other solutes or post-translational modifications.

SAHS and CAHS are conserved among tardigrades

Comparison of cDNA sequences and genome sequences revealed the genomic structure of SAHS and CAHS proteins (Fig. S1). In both protein families, all of their splice sites were not conserved among family members, suggesting that most introns have been inserted after the divergence of paralogous members during evolution (Fig. 2A, 2B, S1). A TBLASTN search against the EST or TSA databases in NCBI retrieved significantly similar sequences from two other tardigrade species, *Hypsibius dujardini* and *Milnesium tardigradum*. The assembly of corresponding EST sequences revealed 6 SAHS and 5 CAHS genes in *H. dujardini*. In this species, the total number of ESTs coding SAHS proteins was 100 and was much higher than those for actin and EF1 α , which were 15 and 64 respectively (Table S4). The number of ESTs for CAHS was 23 and also higher than that of actin. These results suggested that SAHS and CAHS genes were abundantly expressed in *H. dujardini* as well. Although 5 CAHS genes were found in the TSA database of *M. tardigradum*, no SAHS gene was found in the transcriptomic database of this species at present. Phylogenetically, *R. varieornatus* has closer relationship with *H. dujardini* than *M. tardigradum* and thus, the abundant expression of SAHS genes might be a common characteristic in the clade containing *R. varieornatus* and *H. dujardini*. Comparison of these homologous genes revealed three and four conserved regions for SAHS and CAHS proteins, respectively (Fig. 2A, 2B). Detailed inspection of conserved regions revealed two repeats of 19-mer peptides in the CAHS-c2 region (Fig. 2B). We designated these as CAHS-motifs. Each CAHS motif comprised two octapeptides connected by a tripeptide (Fig. 2C). The consensus sequence of the octapeptide was $\Phi\psi\psi\Omega\Phi\Omega(\Phi\psi)\Omega\Phi$, hydrophobic residue; ψ , basic residue; Ω , acidic and amide residue), which was distinct from the consensus 11-mer motif of LEA proteins, $\Phi\Phi\Omega X\Phi\psi\Omega\psi\Phi X\Omega$ [27], [28]

Distinct subcellular localization of CAHS and SAHS

Eukaryotic cells contain membrane-delineated compartments. Computer prediction of subcellular localization using TargetP program suggested that five proteins were targeted to three different locations; SAHS1 and SAHS2 are secretory; CAHS1 and CAHS2 are cytoplasmic; and CAHS3 is mitochondrial (Table S5). The prediction score for CAHS3 was low and its predicted mitochondrial localization was uncertain. To elucidate subcellular localizations of CAHS and SAHS proteins experimentally, we

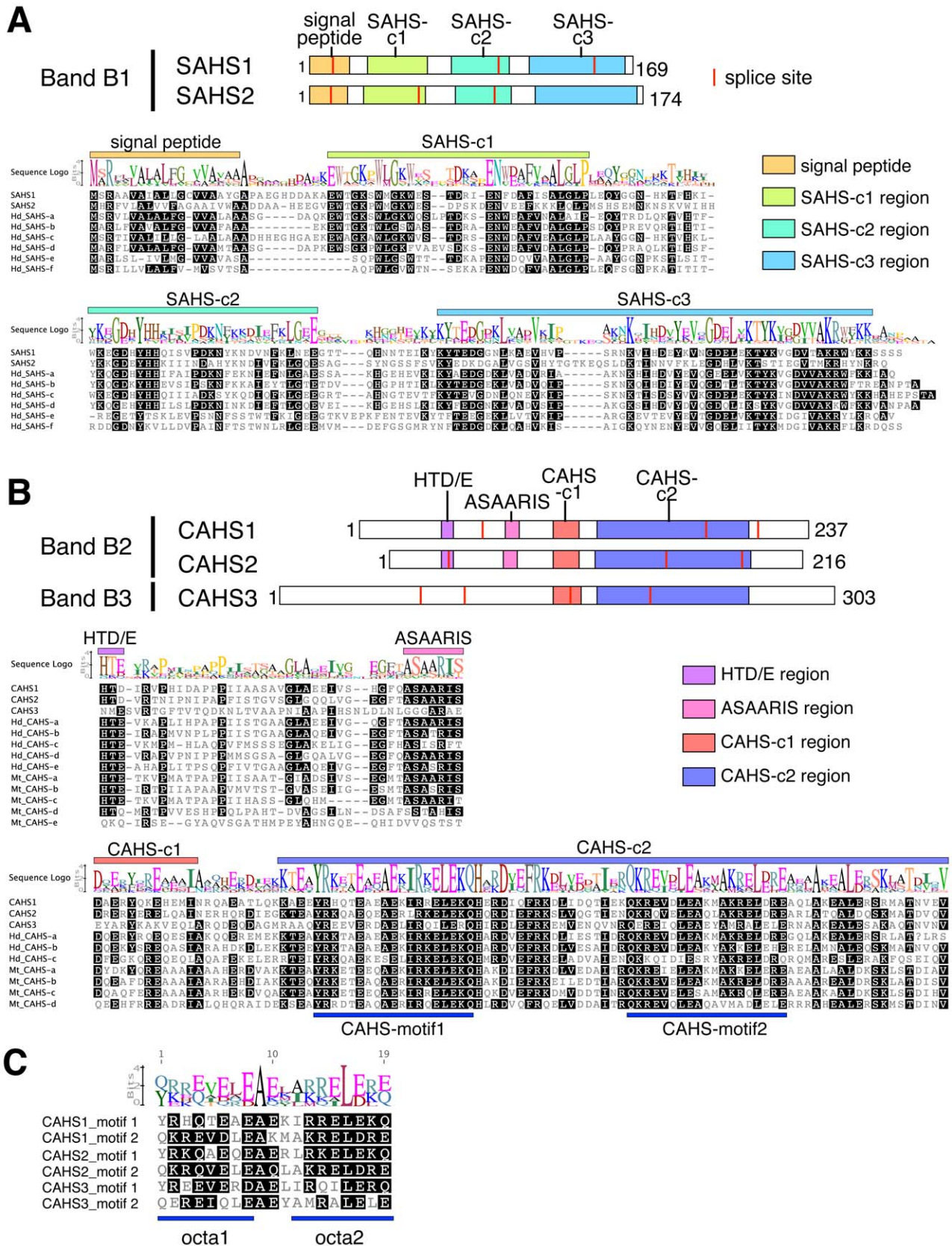


Figure 2. Gene structures and conserved motifs of SAHS and CAHS proteins among tardigrades. (A) Schematic representation of SAHS1 and SAHS2 proteins. Splice sites were not conserved between them. Multiple alignments of SAHS1, SAHS2 and 6 SAHS homologs of *H. dujardini* (Hd_SAHs) revealed 3 conserved regions. Hydrophobic secretory signals were found at the N-terminus in all SAHS homologs. (B) Schematic

representations of CAHS proteins and multiple alignments among CAHS homologs; 3 proteins from *R. varieornatus* and 5 homologs each from *H. dujardini* (Hd_CAHSs) and *M. tardigradum* (Mt_CAHSs). There are 4 conserved regions. Two repeats of 19 mer CAHS-motifs were conserved in the CAHS-c2 region. (C) Multiple alignments of 6 CAHS-motif sequences from *R. varieornatus*. Two octapeptides (octa1, octa2) were repeated with a tripeptide linker.

doi:10.1371/journal.pone.0044209.g002

chose SAHS1, CAHS1 and CAHS3 as representative members of three differently predicted locations, and expressed their GFP-fusion proteins in cultured cells. CAHS1-GFP was distributed mostly in the cytoplasm and weakly in the nucleus (Fig. 3A–D), whereas CAHS3-GFP localized only in the cytoplasm and no signals were observed in the nucleus or mitochondria (Fig. 3E–H). Thus, CAHS proteins localize mainly in the cytoplasm and certain members can also distribute in the nucleus. In contrast, SAHS1-GFP fusion proteins were detected in the culture medium rather than the cell bodies by immunoblot analysis (Fig. 3I), suggesting that SAHS1 is a secretory protein. Fluorescent signals for SAHS1-GFP were hardly detected, but in a small population of cells, weak intracellular signals were observed (Fig. 3J–L), possibly due to proteins remaining in secretory pathway such as the endoplasmic reticulum or Golgi apparatus. SAHS might contribute to protect these secretory organelles.

α -Helices induced by water-deficient conditions

As all SAHS and CAHS proteins were predicted to be intrinsically unstructured similar to LEA proteins by the FoldIndex program [29] (Fig. S2), we examined their secondary structures using CD spectroscopy. In hydrated conditions, SAHS1 had a minimum at 215 nm, which is the typical CD spectrum of β -structures (Fig. 4A). FABP is rich in β -structures [30]. Thus, SAHS1 and FABP probably share similar secondary structures rather than the predicted unstructured conformation. In contrast,

the CD spectrum of CAHS1 had a single deep minimum around 200 nm, similar to that of hydrated LEA proteins [31], suggesting that these proteins are unstructured in hydrated conditions (Fig. 4B). When water-deficient conditions were mimicked by increasing the desolvating agent of trifluoroethanol (TFE), the spectra of both SAHS1 and CAHS1 dynamically changed to the typical spectra of α -structures with minima at 206 nm and 221 nm (Fig. 4A, 4B) [31]. In contrast to SAHS1 which had robust β -structure and more than 50% TFE was required for the transition to α -helices, the CD spectrum of CAHS1 was drastically affected by the addition of only 10% of TFE, suggesting that CAHS1 is more sensitive to water availability. When amino acids of the 19-mer CAHS motif are aligned to α -helical structures, they are expected to form amphiphilic helices, with one hydrophobic stripe and a wider hydrophilic stripe, which can be further divided to four differently charged regions (Fig. 4C). The typical LEA motif is also expected to form amphiphilic helices, but its hydrophilic stripe is divided into three regions rather than four, and the charge distribution patterns were different between CAHS and LEA motifs (Fig. 4C, 4D).

Discussion

In this study, we identified five novel abundant heat-soluble proteins forming two novel protein families from an anhydrobiotic tardigrade, *R. varieornatus*. Although there are several genes

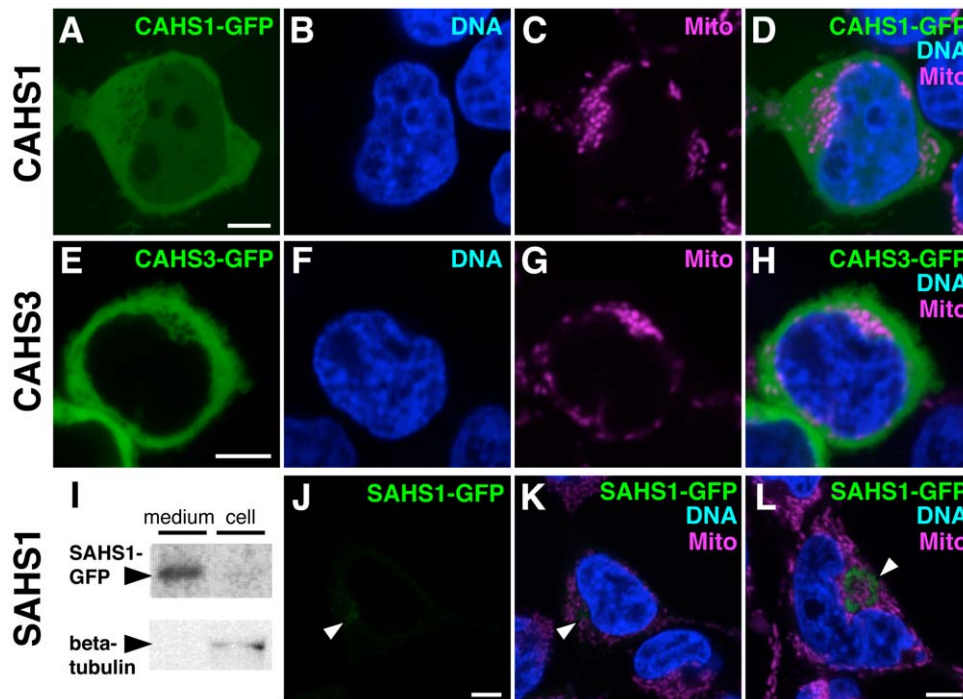


Figure 3. Distinct subcellular localization of CAHS and SAHS proteins. Pseudo-color fluorescent images of 293T cells expressing AcGFP fusion proteins of CAHS1 (A–D), CAHS3 (E–H) and SAHS1 (J–L). (A, E, J) AcGFP fusion proteins. (B, F) DNA. (C, G) mitochondria. (D, H, K, L) merged images. Bars indicate 5 μ m. (I) Immunoblot analyses using anti-GFP and anti- β -tubulin antibody to examine whether SAHS1-GFP was secreted into the culture medium (medium) or retained inside the cells (cell).

doi:10.1371/journal.pone.0044209.g003

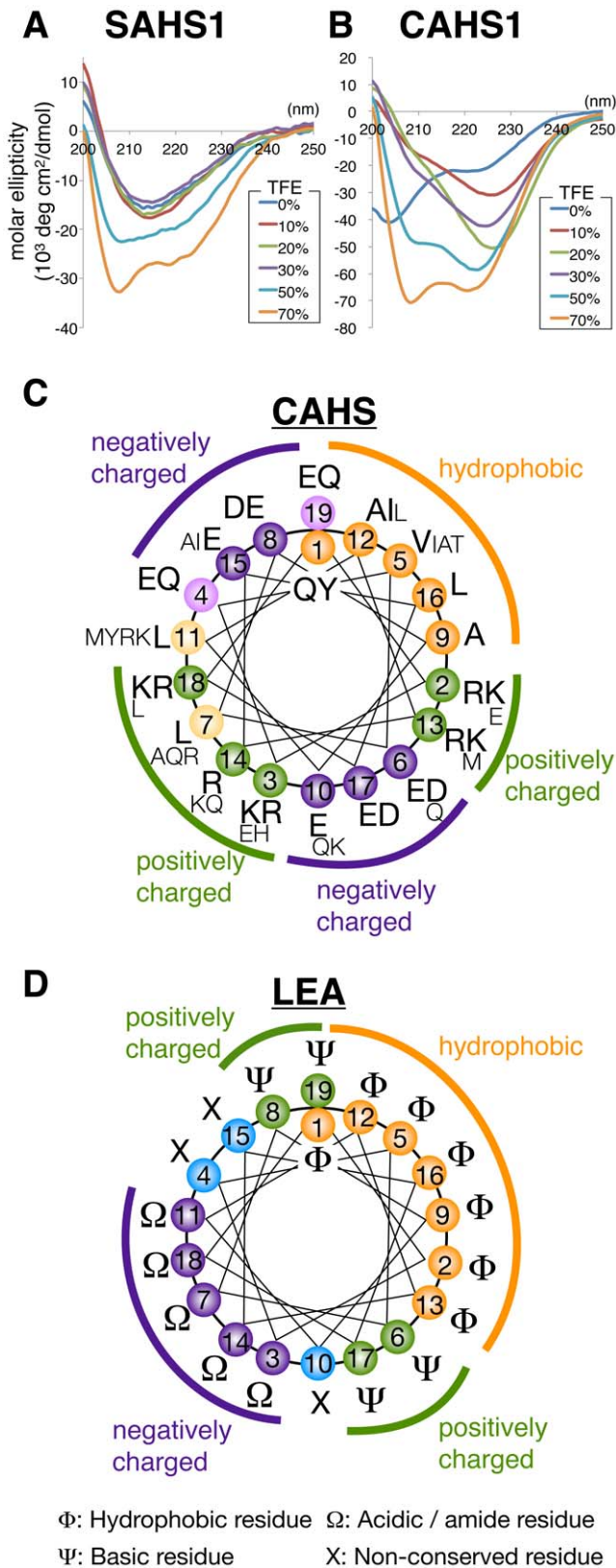


Figure 4. Dynamic change of secondary structures of SAHS and CAHS proteins. (A, B) CD spectra of SAHS1 and CAHS1 proteins in the absence or presence of various concentrations of the desolvating agent TFE, mimicking a water deficient condition. (C) Formation of 5 striped amphiphilic α -helices by CAHS repeats. The numbers show the positions in the repeat. Based on multiple alignments of 6 CAHS

repeats, major amino acids at each position are shown in larger size and minor ones in smaller size. (D) Formation of 4 striped α -helices of conserved LEA motifs. Φ, aliphatic residue; Ψ, basic residue; Ω, acidic or amide residue; X, non-conserved residue. doi:10.1371/journal.pone.0044209.g004

containing LEA-motifs in our draft genome database, no LEA proteins were detected by mass spectroscopy analysis of major heat-soluble bands. These results suggested that tardigrade LEA proteins are not heat-soluble or their expression level is much lower than those of CAHS or SAHS proteins. In any cases, this study established CAHS and SAHS proteins as major heat-soluble proteins in tardigrades. Proteins from both families become α -helical structure upon water deficient conditions. In particular, CAHS motifs form amphiphilic α -helices as LEA proteins do. These results suggested that in the tardigrade, the newly identified SAHS and CAHS proteins could support desiccation tolerance as abundant proteins in a similar manner to LEA proteins.

According to subcellular localization, CAHS protein may contribute to protection of cytoplasm and nucleus, while SAHS protein may protect extracellular components. No accumulation of SAHS1-GFP was observed around the plasma membranes, suggesting that SAHS1 has low affinity to membrane and membrane protection is not the major role of SAHS proteins. Our results did not exclude the possibility that SAHS proteins accumulate in a dehydration dependent manner and protect the membranes. None of CAHS and SAHS proteins localized in mitochondria, implying that mitochondria might be protected by other molecular species.

Reflecting the differences in primary sequences, the α -helices of CAHS motifs had one extra negatively-charged region compared to those of LEA motifs. As a result, each stripe became narrow and thus may prevent too strong hydrophobic or electro-static interaction with other molecules, possibly making CAHS proteins better stuffing molecules. Comparative proteomics using two-dimensional electrophoresis between hydrated and dehydrated tardigrades, revealed no significant reproducible changes (AY, unpublished data), raising the possibility that tardigrades constitutively expressed the factors necessary for anhydrobiosis. Weak interactive ability of CAHS motif might be appropriate for constitutive abundant expression.

Recent search for thermostable proteins from monogonont rotifer, *Brachionus manjavacas*, identified a LEA protein and vitellogenin as major heat soluble proteins in diapausing eggs of the rotifer, but no CAHS/SAHS proteins were found in the search [32]. CAHS and SAHS proteins were conserved among tardigrades, but were not found in other phyla, suggesting that tardigrades might have evolved novel heat-soluble proteins different from LEA proteins as major heat-soluble proteins. This study revealed a novel repertoire of major heat-soluble proteins in anhydrobiotic animals

Supporting Information

Figure S1 Genomic structures of SAHS and CAHS proteins. (A) Both SAHS1 and SAHS2 contained four exons. Colored boxes and red lines indicated conserved regions and splice sites, respectively. (B) CAHS proteins are composed of four or five exons and their splice sites were not conserved among them. (PDF)

Figure S2 Predicted intrinsically unstructured regions of SAHS and CAHS proteins. Y axis is foldIndex based on the net charges and residue hydrophobicity of the given sequence. The values beyond zero (green areas) show the tendency of a given

amino acid for being ordered and the values below zero (red areas) show the tendency of a residue for being intrinsically unstructured. (PDF)

Table S1 Assembly of SAHS family members of *Hypsibius dujardini*. (PDF)

Table S2 Assembly of CAHS family members of *Hypsibius dujardini*. (PDF)

Table S3 Summary of CAHS family members of *Milnesium tardigradum*. (PDF)

Table S4 Summary of ESTs coding housekeeping genes of *Hypsibius dujardini*. (PDF)

References

- Mohjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, et al. (2011) Survival in extreme environments - on the current knowledge of adaptations in tardigrades. *Acta physiologica* (Oxford, England) 202: 409–420.
- Schill RO (2010) Tardigrades. In: *Sleeping Beauties- Dormancy and resistance in harsh environments*, Cerdà/Clark/Lubzens (Eds.): *Dormancy of Cells and Organisms Topics in Current Genetics* (Stefan Hohmann, Series Editor). Springer, Heidelberg, pp. 133–146.
- Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P (2008) Tardigrades survive exposure to space in low Earth orbit. *Current biology* 18: R729–R731.
- Rebecchi L, Altiero T, Guidetti R, Cesari M, Bertolani R, et al. (2009) Tardigrade Resistance to Space Effects: First Results of Experiments on the LIFE-TARSE Mission on FOTON-M3 (September 2007). *Astrobiology* 9: 581–591.
- Persson D, Halberg K a., Jørgensen A, Ricci C, Mohjerg N, et al. (2011) Extreme stress tolerance in tardigrades: surviving space conditions in low earth orbit. *Journal of Zoological Systematics and Evolutionary Research* 49: 90–97.
- Rebecchi L, Altiero T, Cesari M, Bertolani R, Rizzo AM, et al. (2011) Resistance of the anhydrobiotic eutardigrade *Paramacrobiotus richtersi* to space flight (LIFE-TARSE mission on FOTON-M3). *Journal of Zoological Systematics and Evolutionary Research* 49: 98–103.
- Clegg JS (1965) The origin of trehalose and its significance during emergence of encysted dormant embryos of *Artemia salina*. *Comp Biochem Physiol* 14: 135–143.
- Madin KAC, Crowe JH (1975) Anhydrobiosis in nematodes: carbohydrate and lipid metabolism during dehydration. *J Exp Zool* 193: 335–342.
- Watanabe M, Kikawada T, Okuda T (2003) Increase of internal ion concentration triggers trehalose synthesis associated with cryptobiosis in larvae of *Polypedilum vanderplanki*. *J Exp Biol* 206: 2281–2286.
- Westh P, Ramløv H (1991) Trehalose accumulation in the tardigrade *Adorybiotus coronifer* during anhydrobiosis. *Journal of Experimental Zoology* 258: 303–311.
- Hengherr S, Heyer AG, Köhler H-R, Schill RO (2008) Trehalose and anhydrobiosis in tardigrades-evidence for divergence in responses to dehydration. *FEBS J* 275: 281–288.
- Jönsson KI, Persson O (2010) Trehalose in three species of desiccation tolerant tardigrades. *Open Zool J* 3: 1–5.
- Browne J, Dolan K, Tyson T, Goyal K (2004) Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*. *Eukaryotic Cell* 3: 966–975.
- Tunnacliffe A, Lapinski J, McGee B (2005) A Putative LEA Protein, but no Trehalose, is Present in Anhydrobiotic Bdelloid Rotifers. *Hydrobiologia* 546: 315–321.
- Kikawada T, Nakahara Y, Kanamori Y, Iwata K, Watanabe M, McGee B, Tunnacliffe A, Okuda T (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochem Biophys Res Commun* 348: 56–61.
- Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94: 791–812.
- Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, Tunnacliffe A (2003) Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J Biol Chem* 278: 12977–12984.
- Chakrabortee S, Boschetti C, Walton IJ, Sarkar S, Rubinsztein DC, Tunnacliffe A (2007) Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc Nat Acad Sci* 104: 18073–18078.
- Förster F, Liang C, Shkumatov A, Beisser D, Engelmann JC, et al. (2009) Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional protein clusters as well as RNA elements in tardigrades. *BMC genomics* 10: 469.
- Schokraie E, Hotz-Wagenblatt A, Warnken U, Mali B, Frohme M, et al. (2010) Proteomic analysis of tardigrades: towards a better understanding of molecular mechanisms by anhydrobiotic organisms. *PLoS one* 5: e9502.
- Mali B, Grohme M a, Förster F, Dandekar T, Schnölzer M, et al. (2010) Transcriptome survey of the anhydrobiotic tardigrade *Milnesium tardigradum* in comparison with *Hypsibius dujardini* and *Richtersius coronifer*. *BMC genomics* 11: 168.
- Förster F, Beisser D, Grohme MA, Liang C, Mali B, et al. (2012) Transcriptome Analysis in Tardigrade Species Reveals Specific Molecular Pathways for Stress Adaptations. *Bioinformatics and Biology Insights* 6: 69–96.
- Horikawa DD, Kunieda T, Abe W, Watanabe M, Nakahara Y, Yukuhiro F, Sakashita T, Hamada N, Wada S, Funayama T, Katagiri C, Kobayashi Y, Higashi S, Okuda T. (2008) Establishment of a rearing system of the extremotolerant tardigrade *Ramazzottius varieornatus*: a new model animal for astrobiology. *Astrobiology* 8: 549–556.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot Eng* 10: 1–6.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005–1016.
- Storch J, Corsico B (2008) The emerging functions and mechanisms of mammalian fatty acid-binding proteins. *Annu Rev Nutr* 28: 73–95.
- Dure III L (2001) Occurrence of a Repeating 11-Mer Amino Acid Sequence Motif in Diverse Organisms. *Prot Pept Lett* 8: 115–122.
- Shimizu T, Kanamori Y, Furuki T, Kikawada T, Okuda T, et al. (2010) Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides. *Biochemistry* 49: 1093–1104.
- Prilusky J, Felder C, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann J, Silman I, Sussman JL (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21: 3435–3438.
- Schievano E, Quarzago D, Spadon P, Monaco H, Zanotti G, Peggion E (1994) Conformational and binding properties of chicken liver basic fatty acid binding protein in solution. *Biopolymers* 34: 879–887.
- Tolletter D, Jaquinod M, Mangavel C, Passirani C, Saulnier P, Manon S, Teyssier E, Payet N, Avelange-Macherel MH, Macherel D (2007) Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *Plant Cell* 19: 1580–1589.
- Jones BL, Schneider DM, Snell TW (2012) Thermostable proteins in the diapausing eggs of *Brachionus manjavacas* (Rotifera). *Comparative biochemistry and physiology Part A, Molecular & integrative physiology* 162: 193–199.

Table S5 Summary of newly identified heat-soluble proteins from *Ramazzottius varieornatus*. (PDF)

Acknowledgments

We are grateful to Mariko Yokogawa and Ichiro Shimada (The University of Tokyo, Japan) for support regarding the CD spectroscopy.

Author Contributions

Conceived and designed the experiments: AY T. Kunieda. Performed the experiments: AY ST SY CT SIO. Analyzed the data: AY ST T. Kunieda. Contributed reagents/materials/analysis tools: AY HK DDH AT T. Katayama KA AF. Wrote the paper: AY T. Kubo T. Kunieda.