Identification of four SUMO paralogs in the medaka fish, *Oryzias latipes*, and their classification into two subfamilies.

Daisuke Seki[1, 2], Shinpei Obata[1], Tsuyoshi Shirozu[1], Takeshi Kitano[1] and Hisato Saitoh[1, 2, 3]

[1]Department of Biological Sciences, Graduate School of Science and Technology, and
[2]Global COE (Centers of Excellence) Program, Global Initiative Center for Pulsed Power Engineering, Kumamoto University, Kumamoto, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

[3]To Whom correspondence should be addressed:
Email: hisa@kumamoto-u.ac.jp,
Full Postal Address: Department of Biological Sciences, Graduate School of Science and Technology,
Kumamoto University, Kumamoto, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
Phone: +81-96-342-3450, Fax: +81-96-342-3450

Abstract At least four paralogs of the small ubiquitin-related modifier (SUMO) exist in humans, but there is limited information about SUMO paralogs from other vertebrate species. We isolated the four cDNA encoding proteins, similar to human SUMOs, from the medaka fish, *Oryzias latipes*: *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3, and *Ol*SUMO-4. The amino acid sequences of *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 are 89–94% identical, but they share only 45% identity with *Ol*SUMO-1. Phylogenetic analysis, transient expression of *Ol*SUMOs in cultured cells and *in vitro* binding of *Ol*SUMOs with two different SUMO-interacting proteins demonstrated that the medaka SUMO paralogs can be grouped into two subfamilies, *Ol*SUMO-1 and *Ol*SUMO-2/3/4, respectively. Furthermore, this is the first report of all four *Ol*SUMO transcripts being expressed in medaka embryos, implying that they have a role in fish development. The study will improve understanding of the relationship between structural and functional diversity of SUMO paralogs during vertebrate evolution.

Keywords Posttranslational modification; Small ubiquitin-related modifier (SUMO); medaka fish

Introduction

Post-translational modifications are an important mechanism by which structures and functions of cellular proteins are controlled. Among post-translational protein modifications, sumoylation is a unique type in which the small ubiquitin-related modifiers (SUMOs) are covalently conjugated to lysine residues in a wide variety of target proteins in eukaryotic cells. Sumoylation is important in regulating numerous cellular processes, including transcription, epigenetic gene control, genomic instability, and protein degradation (Geiss-Friedlander and Melchior 2007; Wilson and Heaton 2008; Wang and Dasso 2009). The SUMO modification pathway is regulated markedly not only by multiple enzymes involving SUMO proteases (SENPs), SUMO-activation E1 enzyme (Aos1/Uba2), SUMO-conjugation E2 enzyme (Ubc9) and SUMO-E3 ligases, such as the protein inhibitor of the activated STAT (PIAS) family of proteins and Ran-binding protein 2/nucleoporin 358kDa (RanBP2/Nup358), but also by diverse SUMO-interacting proteins that recognize conjugated SUMO moieties via SUMO-interacting motifs (SIMs), also known as SUMO-binding domains (Geiss-Friedlander and Melchior 2007; Wilson and Heaton 2008; Wang and Dasso 2009).

SUMOs are highly conserved from yeast to humans. At least three paralogs have been reported in human and mice: SUMO-1/SMT3C, SUMO-2/SMT3A and SUMO-3/SMT3B. SUMO-2 and SUMO-3 are more closely related to each other (95% amino acid identity) than they are to SUMO-1 (≈50% identity). Although SUMO-1 and SUMO-2/3 can be equally conjugated to a subset of proteins, several lines of evidence indicate that SUMO-1 and SUMO-2/3 are conjugated to different proteins, and represent unique signals regulating different cellular functions (Saitoh and Hinchey 2000; Tatham et al. 2001; Rosas-Acosta et al. 2005; Vertegaal et al. 2006). Intriguingly, in humans but not in mice, there is another SUMO paralog, designated as SUMO-4, which differs from SUMO-1/2/3 in that it not only seems to be expressed mainly in the kidney, lymph node and spleen, but is also unable to form covalent modification with substrates because of a unique proline residue at position 90 (Pro-90) (Guo et al. 2004; Owerbach et al. 2005). To date, there is limited information on a SUMO paralog that is similar to human SUMO-4 in other vertebrate species, and when the structural and functional diversification of SUMO paralogs occurred during vertebrate evolution remains uncertain. Thus, it is important to identify and investigate other examples of vertebrate SUMOs.

Here we report the isolation of four cDNAs of medaka SUMO paralogs, termed: *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4. Medaka, *Oryzias latipes*, is a small egg-laying freshwater teleost fish with several advantages for biological experiments (Ozato et al. 1986; Wada et al. 1995; Ishikawa et al. 2000; Loosli et al. 2000; Kasahara et al. 2007; Shiraishi et al. 2008). Our data, including sequence comparison and quantitative analysis of transcripts during medaka embryogenesis, protein expression and subcellular localization in cultured cells, and comparison of binding affinities to paralog-specific SUMO-binding proteins, suggest a possible classification of medaka SUMOs into two subfamilies, *Ol*SUMO-1 and *Ol*SUMO-2/3/4, and imply the absence of a functionally important proline residue in medaka SUMO paralogs which corresponds to the Pro-90 in human SUMO-4, arguing the divergence and/or specialization of structure and function of human SUMO-4 during vertebrate evolution.

Materials and Methods

Fish samples

The orange-red variety of medaka, the FLFII strain, was selected for the experiments. Fish embryos were maintained in ERM (17 mM NaCl, 0.4 mM KCl, 0.27 mM $CaCl_2 2H_2O$, 0.66 mM $MgSO_4$, pH 7.0) at 26℃under a 14 hours light and 10 hours dark cycle. Developmental stages of the embryos were determined according to the description by Iwamatsu (Iwamatsu et al. 1994).

Database analysis and construction of the phylogenic tree

BLAST searches for DNA and protein identities were conducted using the medaka transcription database (www.blast.ddbj.nig.ac.jp/top-j.html and www.ensembl.org/Oryzias_latipes/index.html) and the Medaka Genome Initiative (www.park.itc.u-tokyo.ac.jp/K-medaka/MGI2/MGI.html). Analyses of the predicated protein sequences were conducted using BLAST. Phylogenic tree are generated using ClustalW server at the DDBJ (clustalw.ddbj.nig.ac.jp/top-j.html) with standard setting. The GeneBank protein sequences and accession numbers used in these analyses were as follows: human SUMO-1, P63165; human SUMO-2, P61956; human SUMO-3, P55854; human SUMO-4, BAH05006; human ubiquitin, P62988; human Aos1, AAD23902; human Uba2, CAG33037; human Ubc9, CAA05359; human PIAS1,

O75925; human PIAS2, O75928; human PIAS4, AAH04389; human RanBP2/Nup358, P49792; human SENP1, Q9P0U3; human SENP3, Q9H4L4; human SENP5, Q96HI0; human SENP6, Q9GZR1; human SERNP7, Q9BQF6; human Ran-binding protein 2/Nucleoporin 358kDa (RanBP2/Nup358), P49792; human MBD1-containing chromatin-associated factor 1 (MCAF1), Q6VMQ6; human Ring finger protein 4 (RNF4), P78317; human thymine DNA glycosylase (TDG), Q13569; human histone H3-K9 methyltransferase SETDB1, Q15047.

RNA extraction, cDNA cloning, and sequence analysis

Total RNA was extracted from 0-, 3-, 6- and 9-dpf embryos of FLFII medaka using ISOGEN (Nippongene), and the first strand cDNA was synthesized from the total RNA by oligo(dT) priming with SuperScript™ III First-Strand Synthesis SuperMix (Invitrogen). cDNAs encoding the open reading frames of *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 were amplified by PCR using the following primers: *Ol*SUMO-1 forward (fw) 5’-CGCACACAGTCAGGATAAAC-3’/reverse (rv) 5’-AAAACATCAGAAATTGTGGCT-3’, *Ol*SUMO-2 fw 5’-ACACTAGCCACAGCAGCAG-3’/rv 5’-AGGGATGTGGAAAGAAAACAGT-3’, *Ol*SUMO-3 fw 5’-TCCCGTCAATTCACCAGAC-3’/rv 5’-GGTCTGAAGGTGGTCACTTAAT-3’, and *Ol*SUMO-4 fw 5’-AGCGCCAAAAGAGTGACG-3’/rv 5’-GCAGCATGTGTGGCTGA-3’, respectively.

Real-time reverse transcription-polymerase chain reaction (RT-PCR)

To quantitatively compare the amount of transcripts of medaka SUMO paralogs, real-time RT-PCR was performed using LightCycler 350S (Roche) and LightCycler FastStart DNA Master^{PLUS} SYBR Green I (Roche) according to the manufacturer’s protocols. The sequences of the primer used in this assay were as follows: *Ol*SUMO-1 fw 5'-GAGGCGACAAGAAAGATGGA-3'/rv 5'-TTTGGGGTTTGGTTATCTGC-3', *Ol*SUMO-2 fw; 5'-AAAACGAGCACATCAACCTG-3'/ rv 5'-GGCTGTCCATCAAATCGAAAT-3', *Ol*SUMO-3 fw 5'-CAAGATGGGTCGGTGGTC-3'/rv 5'-CAAGCTGTGCAGGTGTATCC-3', *Ol*SUMO-4 fw 5'-TAGCAGGTCAGGATGGATCT-3'/rv 5'-GCAGGTGTGTCTGTCTCATTT-3', and medaka β-actin (the GeneBank accession

number; S74868) fw 5'-TCCACCTTCCAGCAGATGTG-3'/ rv 5'-AGCATTTGCGGTGGACGAT-3'. We also used the transcripts derived from the GAPDH, RPL7 and 18srRNA genes as internal controls; the relative ratios of the SUMO paralogs were almost the same (data not shown).

Construction of expression plasmids

To generate pEGFP-*Ol*SUMO-1/2/3/4gg and pEGFP-*Ol*SUMO-1/2/3/4g mammalian expression plasmids, the DNA fragments of medaka SUMO paralogs were amplified by PCR using the following oligonucleotide primers: *Ol*SUMO-1gg fw 5’-GAGAATTCATGTCAGACACGGAGAC-3’/rv 5’-ATGTCGACTTATCCGCCGGTCTGTTC-3’, *Ol*SUMO-1g fw 5’-GAGAATTCATGTCAGACACGGAGAC-3’/rv 5’-ATGTCGACTTAGCCGGTCTGTTCTTG-3’, *Ol*SUMO-2gg fw 5’-GAAGAATTCATGGCAGACGAGACG-3’/rv 5’-TAGTCGACTTAACCTCCAGTCTGCTGTT-3’, *Ol*SUMO-2g fw 5’-GAAGAATTCATGGCAGACGAGACG-3’/rv 5’-TAGTCGACTTATCCAGTCTGCTGTTGGAA-3’, *Ol*SUMO-3gg fw 5’-ATAGAATTCATGTCGGAGGAGAAGCCA-3’/rv 5’-ATGTCGACTTACCCTCCAGTCTGCTG-3’, *Ol*SUMO-3g fw 5’-ATAGAATTCATGTCGGAGGAGAAGCCA-3’/rv 5’-ATGTCGACTTATCCAGTCTGCTGCTG-3’, *Ol*SUMO-4gg fw 5’-GAGGAATTCATGGCTGATGAAAAACCAAAG-3’/rv-5’TAGTCGACTTAGCCCCCCGTCTG-3’, *Ol*SUMO-4g fw 5’-GAGGAATTCATGGCTGATGAAAAACCAAAG-3’ (the *EcoRI* and *SalI* sites are underlined for the fw and rv primers, respectively). PCR fragments were digested with *EcoRI-SalI* and inserted into *EcoRI-XhoI*-digested pEGFP-C2 (Clontech) or pGEX-4T-1 (Amersham Pharmacia Biotech). pEGFP-human SUMO-1 and SUMO-3 plasmid constructs as described previously (Saitoh et al. 1998; Saitoh and Hinchey 2000; Uchimura et al. 2006; Uwada et al. 2010).

Cell culture, transfection, and immunoblotting

Hela cells (maintained in Dulbecco’s modified Eagale’s medium containing 10% fetal bovine serum and antibiotics at 37^{o}C in a 5% CO_2 incubator) were transfected with the appropriate pEGFP-expression plasmids using GeneJuice (TaKaRa). After 24 hours the

cells were washed with PBS and lysed in two pellet volumes of 3xSDS sample buffer. Proteins were separated on SDS-PAGE followed by Western blotting using anti-GFP antibody (Santa Cruz). For detection of GFP signals under the microscope, the cells transfected with GFP-constructs were grown on coverslips and fixed with 4% paraformaldehyde. The cells were permeabilized with 0.2% Triton X-100, mounted under coverslips, and analyzed using Biorevo BZ-9000 (Keyence).

Glutathione S-transferase(GST)-pulldown assay

pGEX, pGEX-human SUMO-1/3, pGEX-*Ol*SUMO-1/2/3/4, pET28-RanBP2-IR and pET28-MCAF1 plasmids were introduced into *E. coli* BL21(DE3) and Rosetta(DE3) strains. The recombinant GST, GST-human SUMO-1/3, GST-*Ol*SUMO-1/2/3/4, $(His)_6$-RanBP2-IR and $(His)_6$-MCAF1 fusion proteins were expressed as described previously (Uchimura et al. 2006; Uwada et al. 2010). A GST-pulldown assay was carried out as described previously (Uchimura et al. 2006; Uwada et al., 2010). The proteins were separated on SDS-PAGE followed by immunoblot analysis using anti-$(His)_6$-tag antibody (Roche) and anti-GST antibody (Santa Cruz).

Results

The SUMO pathway components in the medaka data base

For a comprehensive analysis of the SUMO modification pathway in lower vertebrate species, we used human genes encoding various components in the SUMO pathway to search for orthologs in medaka protein and DNA databases. Our initial search identified at least four distinct medaka SUMO genes: *OlSUMO-1*, *OlSUMO-2*, *OlSUMO-3* and *OlSUMO-4*, which contain multiple small introns (Fig.1A). In addition, there were at least: one gene similar to the human Aos1 (SUMO-E1) gene, one gene similar to human Uba2 (SUMO-E1) gene, and a gene similar to the human Ubc9 (SUMO-E2) gene. At least five distinct genes encoding genes similar to human de-sumoylation protease SENPs, genes similar to several human SUMO-E3s, including RanBP2/Nup358, PIAS1, PIAS2 and PIAS4, and genes similar to currently identified human SUMO-interacting/-binding proteins, such as MCAF1, RNF4, TDG and SETDB1 were also detected (Table 1). These results strongly suggest the existence of a protein-conjugation pathway in medaka analogous to the human SUMO pathways.

*Isolation of four different SUMO transcripts: Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3, *and Ol*SUMO-4

To confirm that all four *Ol*SUMO genes can be transcribed in medaka, we designed primer pairs and used them to amplify the cDNA fragments using total RNA prepared from fertilized medaka embryos (see Materials and methods section for details). The cDNA fragments for all of the primer pairs, corresponding to full-length *Ol*SUMO-1/2/3/4, amplified efficiently; we then cloned them and determined their DNA sequences. The GeneBank accession numbers of *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 are as follows: GQ463435, GQ463436, GQ463438 and GQ463437, respectively. We found three sequences identical to GQ463435, GQ463437, GQ463438 in the Ensemble's medaka gene database (www.ensembl.org/Oryzias_latipes/index.html) and have deposited them as SUMO-1, SUMO-2 and SUMO-4, respectively. We thus gave the nomenclatures for GQ463435, GQ463436, GQ463438 and GQ463437 as *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4, respectively. Although it appears that designation in the database does not consider any biological criteria such as homology search and phylogenetic alignment (see below), we followed the *Ol*SUMO nomenclature (as above) to avoid confusion.

The amino acid sequences deduced from each cDNA are shown in Fig. 1B. Comparisons of the amino acid sequences showed that *Ol*SUMOs are only 16–17% similar to ubiquitin. The similarities of *Ol*SUMO-1 versus *Ol*SUMO-2, *Ol*SUMO-1 versus *Ol*SUMO-3, *Ol*SUMO-1 versus *Ol*SUMO-4, *Ol*SUMO-2 versus *Ol*SUMO-3, *Ol*SUMO-2 versus *Ol*SUMO-4 and *Ol*SUMO-3 versus *Ol*SUMO-4 were 48, 46, 48, 92, 89 and 94%, respectively. Thus, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 proteins are highly homologous, and they are approximately 50% identical to *Ol*SUMO-1. Phylogenetic analysis showed that *Ol*SUMO-1 had the highest similarity to human SUMO-1 (*Hs*SUMO-1), and *Ol*SUMO-3 was closely related to human SUMO-3 (*Hs*SUMO-3); *Ol*SUMO-2 and *Ol*SUMO-4 were equally related to human SUMO-2 (*Hs*SUMO-2) (Fig. 1C). The relationship of human SUMO-4 to *Ol*SUMO-2/4 is no closer than that of human SUMO-2 to *Ol*SUMO-2/4. It should be noted that a proline at 90 amino acid residue in human SUMO-4, which appears critical for this paralog's function (Guo et al. 2004; Owerbach et al. 2005), is not conserved in either *Ol*SUMO paralog.

OlSUMO proteins have potential to serve in the SUMOylation pathway

To confirm that the proteins encoded by the four medaka SUMO genes attach to other proteins similarly to SUMO proteins from other organisms, cDNAs of the *Ol*SUMOs were expressed in Hela cells to see whether *Ol*SUMO proteins could conjugate to cellular proteins. We first generated the constructs that lack the C-terminal amino acids from the highly conserved di-glycine (gly-gly) residues of each *Ol*SUMO (Fig. 1B). The generated fragments, *Ol*SUMO-1gg, *Ol*SUMO-2gg, *Ol*SUMO-3gg and *Ol*SUMO-4gg, were fused to the C-terminus of an enhanced green fluorescent protein, generating EGFP-*Ol*SUMO1gg, EGFP-*Ol*SUMO2gg, EGFP-*Ol*SUMO-3gg and EGFP-*Ol*SUMO-4gg fusion proteins, respectively. EGFP-tagged *Ol*SUMOgg constructs were then expressed in Hela cells, and total cell lysates prepared from the transfected cells were analyzed by Western blot with anti-GFP antibody. Non-conjugated forms of EGFP-*Ol*SUMOgg proteins that migrate at around 45 kDa (EGFP~27kDa+SUMO~18kDa) were clearly identified (*arrowheads* in Fig. 2). In addition, all EGFP-*Ol*SUMOgg proteins form a number of conjugates that migrate at higher molecular mass than the 45 kDa-monomer band. It should be noted that the signal intensities at higher molecular mass varied among the cells expressing different constructs, suggesting functional heterogeneity among *Ol*SUMO paralogs. Removal of the C-terminal invariant gly residues from the EGFP-*Ol*SUMOgg constructs, termed EGFP-*Ol*SUMO-1g, EGFP-*Ol*SUMO-2g, EGFP-*Ol*SUMO-3g and EGFP-*Ol*SUMO-4g, completely removed the high molecular mass bands. Thus these data indicate that all *Ol*SUMO proteins have the potential to conjugate to other proteins, and suggest that activation of the gly residue at the C-terminal end is critical for transfer of SUMO to target proteins, arguing that all *Ol*SUMO paralogs are logically active in mammalian cultured cells.

Biochemical distinctions among OlSUMOs

To improve understanding of biochemical and physiological properties of *Ol*SUMO paralogs, we tested subcellular localization of four *Ol*SUMO paralogs. This was carried out by transient transfection assay using EGFP-*Ol*SUMOs. A marked concentration of green fluorescence at the nuclear rim was observed in many of the EGFP-*Ol*SUMO-1-transfected Hela cells shown (Fig. 3A). In contrast, a mild concentration of the signal at the nuclear rim in the EGFP-*Ol*SUMO-2/3/4-transfected

Hela cells implies that *Ol*SUMO-1 and *Ol*SUMO-2/3/4 have different requirements in their subcellular localization at the nuclear rim. It should be noted that we observed some punctuate concentrations in the nucleus of both the EGFP-*Ol*SUMO-1 and EGFP-*Ol*SUMO-2/3/4 transfected cells, indicating the possibility of colocalization of *Ol*SUMO-1 and *Ol*SUMO-2/3/4 in the nuclear punctuate structures in interphase cells.

Next, we conducted a GST-pulldown assay to compare the affinity of two types of SIM-containing SUMO-interacting proteins, RanBP2 and MCAF1, to the *Ol*SUMO paralogs. RanBP2, a SIM-containing protein localized at the nuclear pore, binds more readily to human SUMO-1 than it does to SUMO-2/3 (Saitoh et al. 1998; Song et al. 2004; Hecker et al. 2006), and MCAF1, a SIM-containing protein involved in nuclear/chromatin function, binds more readily to human SUMO-2/3 than it does to SUMO-1 (Hecker et al. 2006; Uchimura et al. 2006; Sekiyama et al. 2008; Uwada et al. 2010). The $(His)_6$-RanBP2-IR protein was detected in the GST-*Ol*SUMO-1-pulldown fraction, but not in GST-*Ol*SUMO2/3/4-pulldown fractions (Fig. 3B). In contrast, the $(His)_6$-MCAF1 protein was detected in the GST-*Ol*SUMO-2/3/4-pulldown fractions, but not in the GST-*Ol*SUMO-1-pulldown fraction. These results indicate that *Ol*SUMO1and *Ol*SUMO2/3/4 paralogs can be distinguished by their subcellular localization, and differentiated biochemically by SUMO-1-specific and SUMO-2/3-specific SIM-containing proteins, supporting the notion that *Ol*SUMO paralogs can be grouped into two subfamilies.

Quantitative analysis of transcripts of OlSUMOs during medaka embryogenesis

To elucidate functional heterogeneity or differential regulation among *Ol*SUMO paralogs, or both, we assessed the relative amounts of *Ol*SUMO transcripts in medaka embryos at different developmental stages using real-time RT-PCR. All *Ol*SUMO transcripts were reasonably abundant in 0-dpf embryos (stage 8~9), 3-dpf embryos (stage 29), 6-dpf embryos (stage 36), and 9-dpf embryos (stage 39) (Fig. 4). While the expression levels of *Ol*SUMO-2/4 in 0-dpf embryos and *Ol*SUMO-2/3/4 in 3-dpf embryos were significantly higher than they were in the others, all of the *Ol*SUMO transcripts seemed to undergo small changes in abundance through embryogenesis. To compare the relative amounts of the subfamilies at each embryonic stage, we integrated the values of the *Ol*SUMO-2/3/4 transcripts and compared them with those of *Ol*SUMO-1. The levels of the transcripts from the *Ol*SUMO-2/3/4 subfamily were

297 constantly higher than those of *Ol*SUMO-1 at any of embryonic stages (Fig. 4 and data
298 not shown). The *Ol*SUMO-2/3/4 transcripts were approximately 8-fold more abundant
299 than the *Ol*SUMO-1 transcript, especially at 3-dpf. These data indicate that transcripts
300 of all four SUMO paralogs exist during medaka embryogenesis, and that the amount of
301 *Ol*SUMO-2/3/4 transcript is greater than that of *Ol*SUMO-1 at any stage of
302 development.
303
304 **Discussion**
305 By searching the medaka genome and EST databases, we discovered gene loci
306 and cDNA sequences for multiple components of the SUMO modification system, four
307 SUMO paralogs: *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4; one E1 (one
308 *Ol*Aos1 and one *Ol*Uba2); one E2 (*Ol*Ubc9); three *Ol*PIAS families of E3s;
309 *Ol*RanBP2/Nup358; another class of E3; and a five-gene family encoding putative
310 SUMO proteases (*Ol*SENPs). We also found several mammalian orthologs of
311 downstream effector proteins from the SUMO modification pathway, including MCAF1
312 (Uchimira et al. 2006), TDG (Baba et al. 2005), SETDB1 (Ivanov et al. 2007), RNF4
313 (Häkli et al. 2005), and others (data not shown). These SUMO pathway components in
314 the medaka database suggest that there is a SUMO modification pathway in medaka.
315 Given that we and others also found multiple SUMO paralogs and SUMO modification
316 enzymes in the zebrafish (*Danio reio*) database (data not shown; Nowak and
317 Hammerschmidt 2006; Yuan et al. 2009), our search implies sumoylation as an
318 important signaling mechanism in fishes.
319 To date, in humans, four SUMO paralogs, *Hs*SUMO-1 to *Hs*SUMO-4, have
320 been identified; *Hs*SUMO-1, *Hs*SUMO-2 and *Hs*SUMO-3 can act as protein modifiers,
321 whereas SUMO-4 seems to be expressed only in restricted tissues and may not have the
322 ability to be conjugated to other proteins (Guo et al. 2004; Owerbach et al. 2005;
323 Geiss-Friedlander and Melchior 2007; Wilson and Heaton 2008), indicating that
324 *Hs*SUMO-4 constitutes a subgroup that is distinct from *Hs*SUMO-1/2/3. Our
325 biochemical studies using SIM-containing proteins, RanBP2 and MCAF1, support the
326 notion of two SUMO subfamilies in medaka (Fig. 3B). With regard to their subcellular
327 localization, we also found subfamily-specific properties (Fig. 3A). These findings
328 suggest that *Ol*SUMOs can be grouped into two subfamilies; *Ol*SUMO-1 and
329 *Ol*SUMO-2/3/4 subfamilies. In addition, all *Ol*SUMOs are expressed throughout

medaka embryogenesis (Fig. 4). When transiently expressed in Hela cells, all *Ol*SUMOs were incorporated into higher molecular mass regions, suggesting that they all have ability to be conjugate to cellular proteins *in vivo* (Fig. 2). An amino acid sequence alignment experiment suggested that all medaka SUMO paralogs do not contain a unique proline residue located at position 90 in *Hs*SUMO-4 (Fig.1). Thus we suppose that all *Ol*SUMOs appear distinct from *Hs*SUMO-4 and suggest the emergence of human SUMO-4 paralog after mammalian evolution.

The SUMO modification system is essential in most organisms including *S. cerevisiae*, *C. elegans*, *Arabidopsis thaliana* and mice (Geiss-Friedlander and Melchior 2007), and may play a critical role in some parasitic diseases (Cabral et al. 2008). Whether individual SUMO proteins are essential in organisms that have multiple SUMO paralogs remains unclear. Genetic studies have linked SUMO-1 haploinsufficiency to the cleft lip or palate condition in humans, indicating that SUMO-1 and SUMO-2/3 play a role in development (Alkuraya et al. 2006; Pauws and Stanier 2007). On the other hand, mice lacking SUMO-1 develop without any apparent abnormalities, implying that humans and mice may be different in their specific requirements for SUMO paralogs (Evdokimov et al. 2008; Zhang et al. 2008). In future experiments using the medaka system, it is important to investigate which developmental processes can be compensatory among SUMO paralogs, to elucidate whether more subtle phenotypic differences are involved in the development of tissues and organs lacking either SUMO paralogs, and to identify targets for SUMOylation by different *Ol*SUMO paralogs during embryogenesis and organogenesis. We believe our study provides a basis for such experiments.

Acknowledgments

We thank Ms. Maki Kuroki for help with extraction of RNA from medaka embryos. We appreciate the helpful discussions with the members of the Saitoh Laboratory . This work was supported by grants to Hisato Saitoh from the Naito Foundation, the Novartis Foundation, the Ministry of Education, Culture, Sport, Science and Technology (MEXT), and the Japan Society for the Promotion of Science (JSPS). Daisuke Seki is a Junior Research Associate of the global COE Program at the Global Initiative Center for Pulsed Power Engineering.

Table 1. Components of the medaka SUMO pathway.

Protein activity	Gene	Chromosome	Accession number [a]	Expression	
				(ESTs)[b]	(RT-PCR)[c]
SUMO	*OlSUMO-1*	2	ENSORLG00000000622	Yes	Yes
	OlSUMO-2	1	ENSORLG00000003279	Yes	Yes
	OlSUMO-3	22	ENSORLG00000013523	Yes	Yes
	OlSUMO-4	8	ENSORLG00000009773	Yes	Yes
E1	*OlSAE1*	13	ENSORLG00000001538	Yes	NT[d]
	OlSAE2	6	ENSORLG00000012575	Yes	NT
E2	*OlUBC9*	8	ENSORLG00000013514	Yes	NT
E3	*OlPias1a*	6	ENSORLG00000006728	Yes	NT
	OlPias1b	3	ENSORLG00000002077	Yes	NT
	OlPias2	9	ENSORLG00000017353	Yes	NT
	OlPias4	4	ENSORLG00000004970	Yes	NT
SENP	*OlSENP1*	7	ENSORLG00000007817	Yes	Yes
	OlSENP2	22	ENSORLG00000013689	Yes	Yes
	OlSENP5	17	ENSORLG00000009124	Yes	Yes
	OlSENP6	24	ENSORLG00000017309	Yes	Yes
	OlSENP7a	21	ENSORLG00000011710	Yes	Yes
	OlSENP7b	21	ENSORLG00000011701	Yes	Yes
SUMO- interacting proteins	*OlRanBP2 / OlNup358*	21	ENSORLG00000015523	Yes	NT
	OlMCAF1	1	ENSORLG00000004183	Yes	NT
	OlRNF4	Unknown	ENSORLG00000020573	Yes	Yes
	OlTDG	6	ENSORLG00000001329	Yes	NT
	OlSETDB1	16	ENSORLG00000010602	Yes	NT

a. Accession number, ensemble medaka database accession number.

b. Expression confirmed by the presence of EST in the database (www.ensembl.org/Oryzias_latipes).

c. Expression confirmed by RT-PCR using total RNA from 9- days after fertilization (daf) of the medaka embryo.

d. NT, not-tested by RT-PCR.

e. ND, not-determined.

References

Alkuraya FS, Saadi I, Lund JJ, Turbe-Doan A, Morton CC, Maas RL (2006) SUMO1 haploinsufficiency leads to cleft lip and palate. Science 313: 1751

Baba D, Maita N, Jee JG, Uchimura Y, Saitoh H, Sugasawa K, Hanaoka F, Tochio H, Hiroaki H, Shirakawa M (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. Nature 435: 979-982

Cabral FJ, Pereira OS Jr, Silva CS, Guerra-Sá R, Rodrigues V (2008) *Schistosoma mansoni* encodes SMT3B and SMT3C molecules responsible for post-translational modification of cellular proteins. Parasitol Int 57: 172-178

Evdokimov E, Sharma P, Lockett SJ, Lualdi M, Kuehn MR (2008) Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. J Cell Sci 121: 4106-4113

Geiss-Friedlander R, Melchior F (2007) Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol 8: 947-956

Guo D, Li M, Zhang Y, Yang P, Eckenrode S, Hopkins D, Zheng W, Purohit S, Podolsky RH, Muir A, Wang J, Dong Z, Brusko T, Atkinson M, Pozzilli P, Zeidler A, Raffel LJ, Jacob CO, Park Y, Serrano-Rios M, Larrad MT, Zhang Z, Garchon HJ, Bach JF, Rotter JI, She JX, Wang CY (2004) A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. Nat Genet 36: 837-841

Häkli M, Karvonen U, Jänne OA, Palvimo JJ (2005) SUMO-1 promotes association of SNURF (RNF4) with PML nuclear bodies. Exp Cell Res 304: 224-233

Hecker CM, Rabiller M, Haglund K, Bayer P, Dikic I (2006) Specification of SUMO1-

and SUMO2-interacting motifs. J Biol Chem 281: 16117-16127

Ishikawa Y (2000) Medakafish as a model system for vertebrate developmental genetics. BioEssays 22: 487-495

Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG, Sadofsky MJ, Zhou MM, Rauscher FJ 3rd (2007) PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Mol Cell 28: 823-837

Iwamatsu T (1994) Stages of normal development in the medaka, *Olizias latipes*. Zool Sci 11: 825-839

Kasahara M et al (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447: 714-719

Loosli F, Koster RW, Carl M, Kuhnlein R, Henrich T, Mucke M, Krone A, Wittbrodt J (2000) A genetic screen for mutations affecting embryonic development in medaka fish (*Orizias laptipes*). Mech Dev 97: 133139

Nowak M, Hammerschmidt M (2006) Ubc9 regulates mitosis and cell survival during zebrafish development. Mol Biol Cell 17:5324-36

Owerbach D, McKay EM, Yeh ET, Gabbay KH, Bohren KM (2005) A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. Biochem Biophys Res Commun 337: 517-520

Ozato K, Kondoh H, Inohara H, Iwamatsu T, Wakamatsu Y, Okada TS (1986) Production of transgenic fish: introduction and expression of chicken delta-crystallin gene in medaka embryos. Cell Differ 19: 237-244

Pauws E, Stanier P (2007) FGF signalling and SUMO modification: new players in the aetiology of cleft lip and/or palate. Trends Genet 23: 631-640

Rosas-Acosta G, Russell WK, Deyrieux A, Russell DH, Wilson VG (2005) A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. Mol Cell Proteomics 4: 56-72

Saitoh H., Hinchey J (2000) Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. J Biol Chem 275: 6252-6258

Saitoh H, Sparrow DB, Shiomi T, Pu RT, Nishimoto T, Mohun TJ, Dasso M (1998) Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. Curr Biol 8: 121-124

Sekiyama N, Ikegami T, Yamane T, Ikeguchi M, Uchimura Y, Baba D, Ariyoshi M, Tochio H, Saitoh H, Shirakawa M (2008) Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. J Biol Chem 283: 35966-35975

Shiraishi E, Yoshinaga N, Miura T, Yokoi H, Wakamatsu Y, Abe S, Kitano T (2008) Mullerian inhibiting substance is required for germ cell proliferation during early gonadal differentiation in medaka (*Oryzias latipes*). Endocrinology 149: 1813-1819

Song J, Durrin LK, Wilkinson TA, Krontiris TG, Chen Y (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proc Natl Acad Sci USA 101: 14373-14378

Tatham MH, Jaffray E, Vaughan OA, Desterro JM, Botting CH, Naismith JH, Hay RT (2001) Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. J Biol Chem 276: 35368-35374

Uchimura Y, Ichimura T, Uwada J, Tachibana T, Sugahara S, Nakao M, Saitoh H (2006) Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. J Biol Chem 281: 23180-23190

Uwada J, Tanaka N, Yamaguchi Y, Uchimura Y, Shibahara KI, Nakao M, Saitoh H (2010) The p150 subunit of CAF-1 causes association of SUMO2/3 with the DNA replication foci. Biochem Biophys Res Commun 391: 407–413

Vertegaal AC, Andersen JS, Ogg SC, Hay RT, Mann M, Lamond AI (2006) Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. Mol Cell Proteomics 5: 2298-22310

Wada H, Naruse K, Shimada A, Shima A (1995) Genetic linkage map of a fish, the Japanese medaka *Orizias laptipes*. Mol Mar Biol Biotechnol 4: 269-274

Wang Y, Dasso M (2009) SUMOylation and deSUMOylation at a glance. J Cell Sci 122: 4249-4252

Wilson VG, Heaton PR (2008) Ubiquitin proteolytic system: focus on SUMO. Expert Rev. Proteomics 5: 121-135

Yuan H, Zhou J, Deng M, Liu X, Bras ML, The HD, Chen SJ, Chen Z, Liu TX, Zhu J (2009) Small ubiquitin-related modifier paralogs are indispensable but functionally redundant during early development of zebrafish. Cell Res: in press

Zhang FP, Mikkonen L, Toppari J, Palvimo JJ, Thesleff I, Jänne OA (2008) Sumo-1 function is dispensable in normal mouse development. Mol Cell Biol 28: 5381-5390

Figure Legends

Figure 1. Identification and characterization of the medaka *OlSUMO* gene family.

A, schematic representation of organization of the medaka *OlSUMO* genes. *Black boxes* and *lines* indicate exons and introns, respectively. The numbers of chromosome on which *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 gene loci are located are shown on the *left side*.

B, amino acid sequence alignment of medaka *Ol*SUMO paralogs by the ClustalW method. The GeneBank accession numbers of *Ol*SUMO-1, *Ol*SUMO-2, *Ol*SUMO-3 and *Ol*SUMO-4 are as follows: GQ463435, GQ463436, GQ463438 and GQ463437, respectively. Potential sumoylation sites (ϕKXE) are *boxed*. The Ub domain present in all *Ol*SUMO paralogs is *outlined*. The arrowhead represents the potential processing site by SENPs that expose the highly-conserved gly-gly residues (*reverse type*) involved in SUMO conjugation. The asterisk denotes a proline at 90 amino acid residue in human SUMO-4, which appears critical for paralog function (*grey box*).

C, phylogenic relationship among medaka *Ol*SUMO-1/2/3/4, human SUMOs (*Hs*SUMO-1/2/3/4), *C. elegans* SUMO (*Ce*SUMO) and *D. melanogaster* SUMO (*Dr*SUMO). Phylogenic tree are generated using ClustalW server at the DDBJ (clustalw.ddbj.nig.ac.jp/top-j.html) with standard setting. Bootstrap values (1,000 replicates) are shown at the branches.

Figure 2. Ability of *Ol*SUMO paralogs to conjugate to cellular proteins.

Hela cells were transfected with EGFP vector (*lane 1*), pEGFP-*Hs*SUMO-1gg (*lane 2*), pEGFP-*Hs*SUMO-1g (*lane 3*), pEGFP-*Hs*SUMO-3gg (*lane 4*), pEGFP-*Hs*SUMO-3g (*lane 5*), pEGFP-*Ol*SUMO-1gg (*lane 6*), pEGFP-*Ol*SUMO-1g (*lane 7*), pEGFP-*Ol*SUMO-2gg (*lane 8*), pEGFP-*Ol*SUMO-2g (*lane 9*), pEGFP-*Ol*SUMO-3gg (*lane 10*), pEGFP-*Ol*SUMO-3g (*lane 11*), pEGFP-*Ol*SUMO-4gg (*lane 12*) or pEGFP-*Ol*SUMO-4g (*lane 13*). After 24 hours, total cell lysates were analyzed by immunoblotting using anti-EGFP antibody. The position of the EGFP-*Ol*SUMO monomer (~45 kDa) is indicated by an *arrowhead*. The high molecular mass bands are indicated by a *bracket* with an *asterisk*. Molecular mass standards are expressed in kilodaltons (kDa).

Figure 3. Distinct biochemical properties among *Ol*SUMO-1/2/3/4 proteins.

A, subcellular localization of *Ol*SUMOs. Hela cells were transfected with pEGFP-*Hs*SUMO-1gg, pEGFP-*Hs*SUMO-3gg, pEGFP-*Ol*SUMO-1gg, pEGFP-*Ol*SUMO-2gg, pEGFP-*Ol*SUMO-3gg and pEGFP-*Ol*SUMO-4gg. After 24 hours, the cells were fixed and EGFP signals were detected under the fluorescent microscope. DAPI-stained images are also indicated (*inset*).

B, interaction among *Ol*SUMOs and SIM-containing proteins. A bacterial lysate expressing recombinant $(His)_6$-RanBP2-IR (*upper panel*) or $(His)_6$-MCAF1 (*bottom panel*) was incubated with beads of GST-*Ol*SUMO-1gg (*lane 3*), GST-*Ol*SUMO-2gg (*lane 4*), GST-*Ol*SUMO-3gg (*lane 5*) and GST-*Ol*SUMO-4gg (*lane 6*). Following incubation, a GST-pulldown assay was carried out and the proteins associated with the beads were subjected to immunoblot analysis using anti-$(His)_6$ antibody. For positive controls, GST-*Hs*SUMO-1 (*lane 1*) and GST-*Hs*SUMO-3 (*lane 2*) were used. Amounts of GST-SUMOs used in the pull-down assay were visualized by immunoblot analysis with anti-GST antibody (*bottom*).

Figure 4. Expression of *Ol*SUMO paralogs embryos at different developmental stages.

Quantitative detection of the transcripts of *Ol*SUMO paralogs in 0-, 3-, 6- and 9-dpf embryos. Quantitative RT-PCR was used to compare the amounts of *Ol*SUMO-1 (*black*), *Ol*SUMO-2 (*gray*), *Ol*SUMO-3 (*white*) and *Ol*SUMO-4 (*dark gray*) transcripts using total RNA prepared from at least three individual embryos at the different developmental stages. In this experiment we used β-actin for the internal control; it showed a 2-fold reduction during medaka embryogenesis.

553

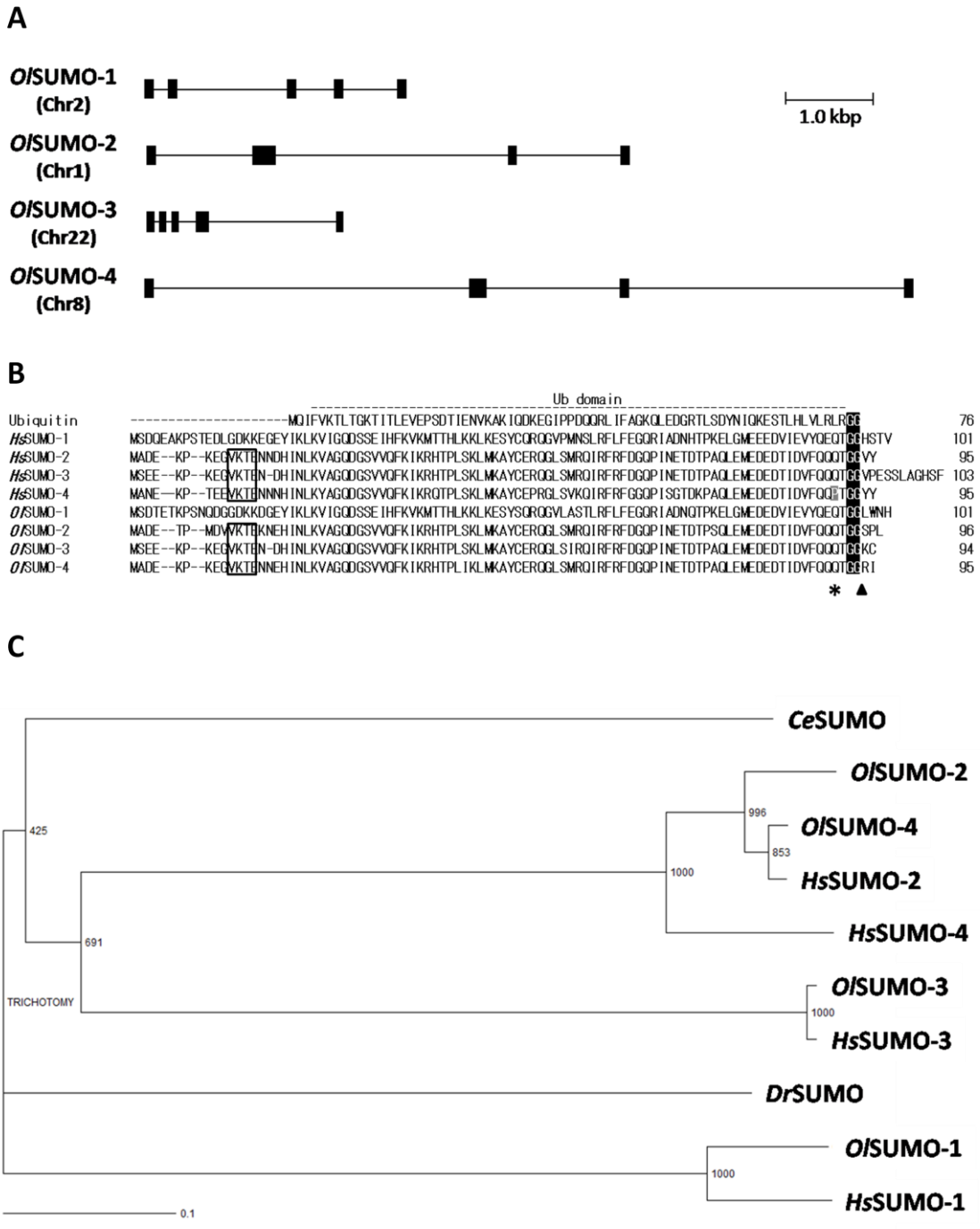

Fig. 1

554

555

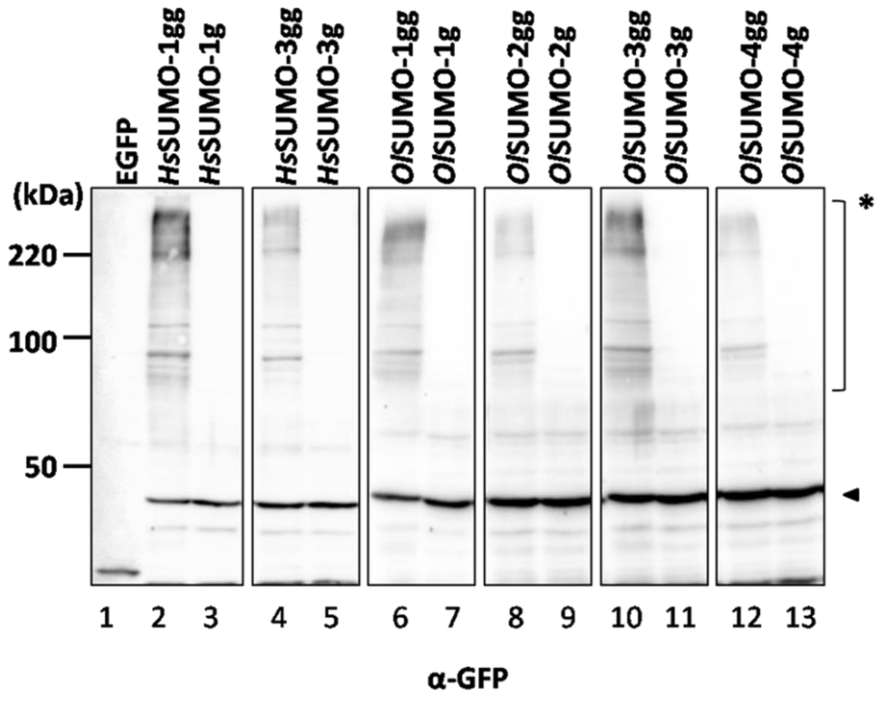
EGFP
HsSUMO-1gg
HsSUMO-1g
HsSUMO-3gg
HsSUMO-3g
OlSUMO-1gg
OlSUMO-1g
OlSUMO-2gg
OlSUMO-2g
OlSUMO-3gg
OlSUMO-3g
OlSUMO-4gg
OlSUMO-4g
(kDa)
220
100
50
*
1 2 3 4 5 6 7 8 9 10 11 12 13
α-GFP

Fig. 2

556

557

A

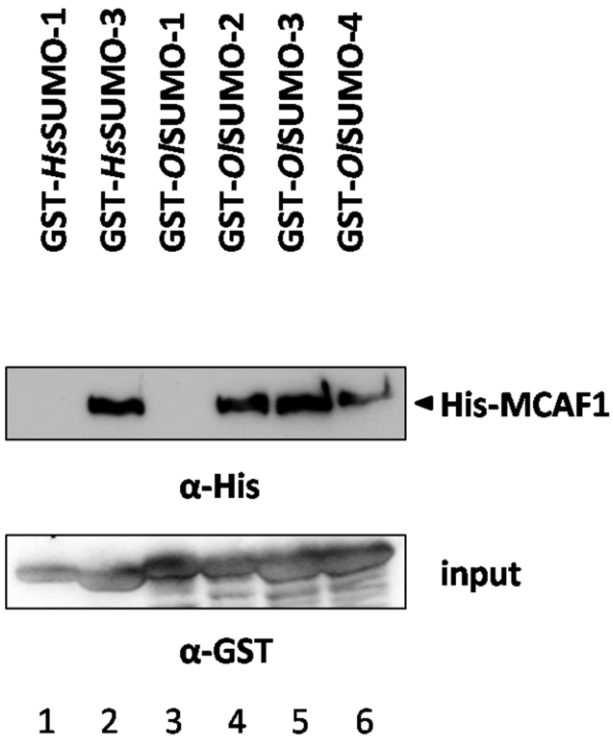

Fig. 3

558

559

B

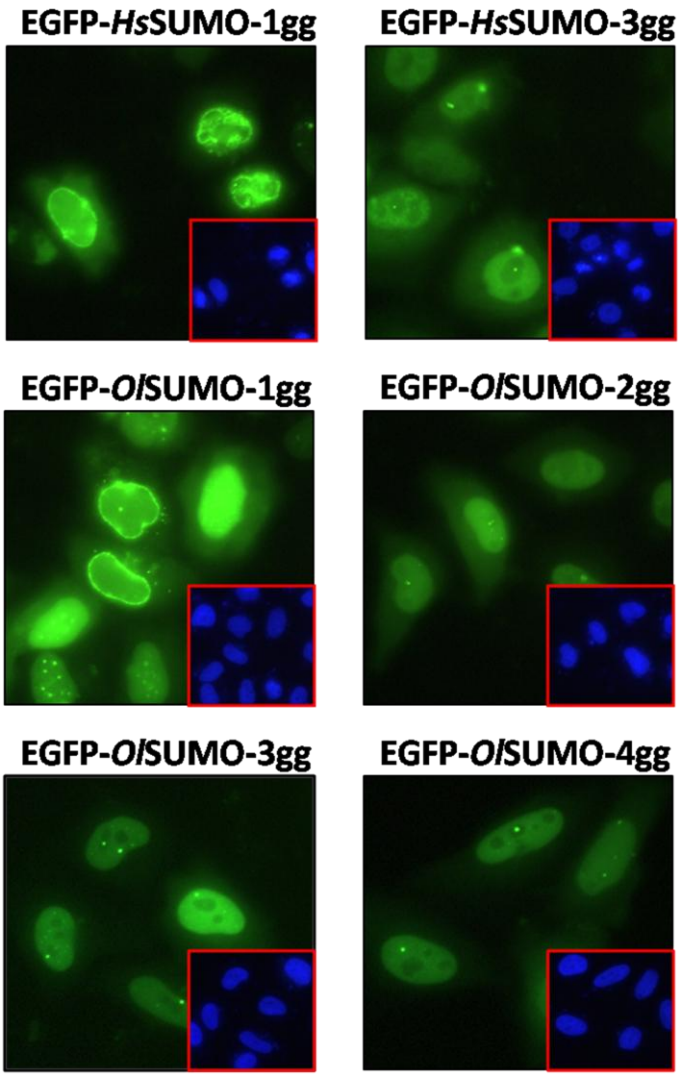

Fig. 3

560

561

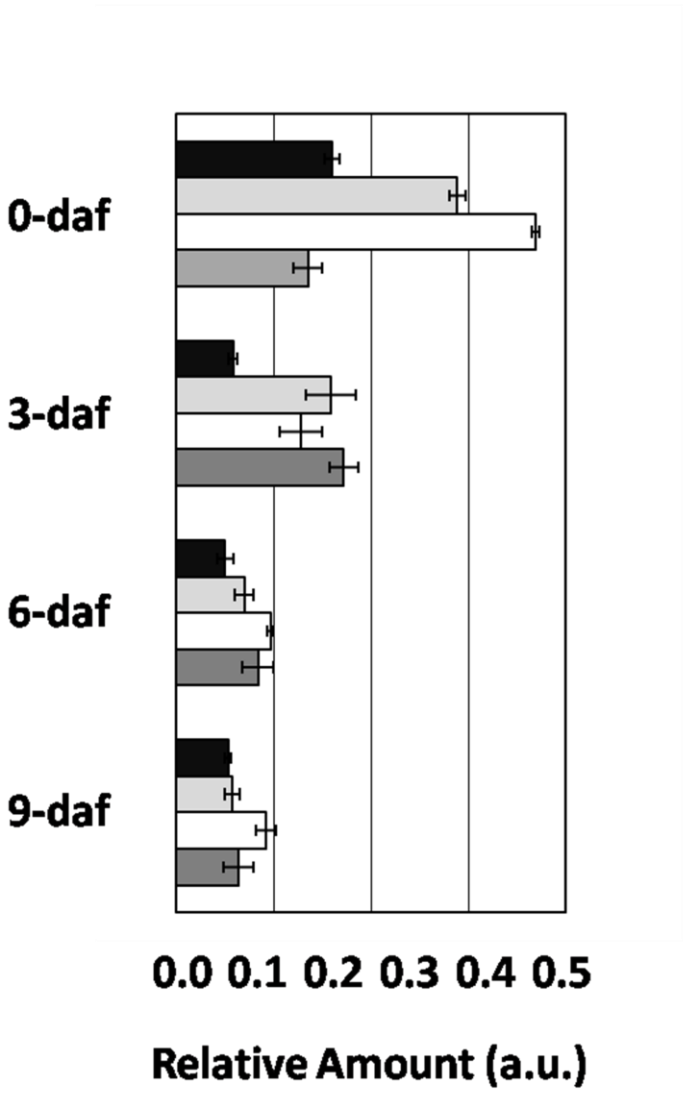

Fig. 4

562