Modular metabolite assembly in *C. elegans* lysosome-related organelles

Henry H. Le[1‡], Chester J. J. Wrobel[1‡], Sarah M. Cohen[2], Jingfang Yu[1], Heenam Park[2], Maximilian J. Helf[1], Brian J. Curtis[1], Pedro R. Rodrigues[1], Paul W. Sternberg[2,*], and Frank C. Schroeder[1,*]

[1]Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, United States.

[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, United States.

[‡]These authors contributed equally.

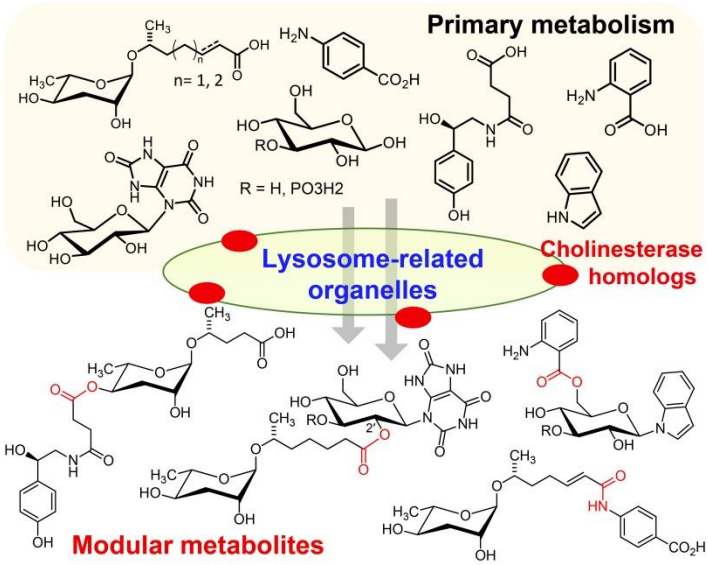
Corresponding authors:

Paul W. Sternberg, pws@caltech.edu

Frank C. Schroeder, fs31@cornell.edu

Abstract

Signaling molecules derived from attachment of diverse primary metabolic building blocks to ascarosides play a central role in the life history of *C. elegans* and other nematodes; however, many aspects of their biogenesis remain unclear. Using comparative metabolomics, we show that lysosome-related organelles (LROs) are required for biosynthesis of most modular ascarosides as well as previously undescribed modular glucosides. Both modular glucosides and ascarosides are derived from highly selective assembly of moieties from nucleoside, amino acid, neurotransmitter, and lipid metabolism. We further show that cholinesterase (*cest*) homologs that localize to the LROs are required for assembly of both modular ascarosides and glucosides, mediating formation of ester and amide linkages between subsets of building blocks. Their specific biosynthesis suggests that modular glucosides, like ascarosides, serve dedicated signaling functions. Further exploration of LRO function and *cest* homologs in *C. elegans* and other animals may reveal additional new compound families and signaling paradigms. (150 Words)



Introduction

Recent studies indicate that the metabolomes of animals, from model systems such as *Caenorhabditis elegans* and Drosophila to humans, may include >100,000 of compounds[1,2]. The structures and functions of most of these small molecules have not been identified, representing a largely untapped reservoir of chemical diversity and bioactivities. In *C. elegans*[3] a large modular library of small-molecule signals, the ascarosides, have been shown to be involved in almost every aspect of its life history, including aging, development, and behavior[4-7]. The ascarosides represent a structurally diverse chemical language, derived from glycosides of the dideoxysugar ascarylose and hydroxylated short-chain fatty acid (Fig. 1a)[8]. Structural and functional specificity arises from optional attachment of additional moieties to the sugar, for example indole-3-carboxylic acid (e.g. icas#3 (**1**)), or carboxy-terminal additions to the fatty acid chain, such as *p*-aminobenzoic acid (PABA, as in ascr#8 (**2**)) or *O*-glucosyl uric acid (e.g. uglas#11 (**3**), Fig. 1b)[2,9-12]. Given that even small changes in the chemical structures of the ascarosides often result in starkly altered biological function, ascaroside biosynthesis appears to correspond to a carefully regulated encoding process in which biological state is translated into chemical structures[13]. Thus, the biosynthesis of ascarosides and other *C. elegans* signaling molecules (e.g. nacq#1)[14] represents a fascinating model system for the endogenous regulation of inter-organismal small-molecule signaling in metazoans. However, for most of the >200 recently identified *C. elegans* metabolites[2,8,9], biosynthetic knowledge is sparse. Previous studies have demonstrated that conserved primary metabolic pathways, e.g. peroxisomal *β*-oxidation[9,10] and amino acid catabolism[8,15] (Fig. 1a), contribute to ascaroside biosynthesis; however, many aspects of the mechanisms underlying assembly of multi-modular metabolites remains unclear.

Recently, metabolomic analysis of null mutants of the Rab-GTPase *glo-1*, which lack a specific type of lysosome-related organelles (LROs, also referred to as gut granules), revealed complete loss of 4′-modified ascarosides in this mutant[13]. The *glo-1*-dependent LROs are acidic, pigmented compartments that are related to mammalian melanosomes and drosophila eye pigment organelles[16,17]. LROs form when lysosomes fuse with other cellular compartments, e.g. peroxisomes, and appear to play an important role for recycling proteins and metabolites[16]. Additionally, it has been suggested that LROs may be involved in the production and secretion of diverse signaling molecules[18,19], and the observation that *glo-1* mutant worms are deficient in 4′-modified ascarosides suggested that the LROs may serve as hubs for their assembly (Fig. 1a)[13].

Parallel studies of other *Caenorhabditis* species[20-22] and *Pristionchus pacificus*[23], a nematode species being developed as a satellite model system to *C. elegans*[24], revealed that production of modular ascarosides is widely conserved among nematodes. Leveraging the high genomic diversity of sequenced *P. pacificus* isolates, genome-wide association studies coupled to metabolomic analysis revealed that the serine hydrolase *uar-1*, a homolog of mammalian cholinesterases, is required for 4′-attachment of an ureidoisobutyryl moiety to a subset of ascarosides, e.g. ubas#3 (**4**, Fig. 1c)[23]. Homology searches revealed a large expansion of cholinesterase (*cest*) homologs in *P. pacificus* as well as *C. elegans* (Fig. S1), and recently it was shown that in *C. elegans*, the *uar-1* homologs *cest-3*, *cest-8*, and *cest-9.2* are involved in the 4′-attachment of other acyl groups in modular ascarosides[25,26]. Based on these findings, we posited that *cest* homologs localize to the LROs where they control assembly of modular ascarosides, and perhaps other modular metabolites. In this work, we present a comprehensive assessment of the role of the LROs in *C. elegans* small molecule biosynthesis and uncover the central role of LRO-localized *cest* homologs in the biosynthesis of diverse modular metabolites derived from highly selective assembly of primary metabolic building blocks.

Results

Novel classes of LRO-dependent metabolites. To gain a comprehensive overview of the LROs in *C. elegans* metabolism, we employed a fully untargeted comparison of the metabolomes of LRO-deficient *glo-1(zu437)* mutant and wildtype worms (Fig. 1d). HPLC-high resolution mass spectrometry (HPLC-HRMS) data for the *exo*- and *endo*- metabolomes of the two strains were analyzed using the Metaboseek comparative metabolomics platform, which integrates the *xcms* package[27]. These comparative analyses revealed that *glo-1* deletion has a dramatic impact on *C. elegans* metabolism. For example, in negative ionization mode we detected >1000 molecular features that were at least 10-fold less abundant in the *glo-1 exo*- and *endo*-metabolomes, as well as >3000 molecular features that are 10-fold upregulated in *glo-1* mutants. For further characterization of differential features, we employed tandem mass spectrometry (MS^2) based molecular networking, a method which groups metabolites based on shared fragmentation patterns (Fig. 1d, S2-5)[28]. The resulting four MS^2 networks – for data obtained in positive and negative ionization mode for the *exo*- and *endo*-metabolomes – revealed several large clusters of features whose abundance was largely abolished or greatly increased in *glo-1* worms. Notably, although some differential MS^2 clusters represented known compounds, e.g. ascarosides, the majority of clusters were found to represent previously undescribed metabolite families.

In agreement with previous studies[13], biosynthesis of most modular ascarosides was abolished or substantially reduced in *glo-1* mutants, including all 4′-modified ascarosides, e.g. icas#3 (**1**) (Figs. 1b, and S6a). Similarly, production of ascarosides modified at the carboxy terminus, e.g. uglas#11 (**3**) derived from ester formation between ascr#1 (**5**) and uric acid glucoside[12] (**6**), and ascr#8 (**2**), derived from formation of an amide bond between ascr#7 (**7**) and of *p*-amino benzoic acid (**8**), was largely abolished in *glo-1* mutants (Figs. 1a, 1b, and S6a). Metabolites plausibly representing building blocks of these modular ascarosides were not strongly perturbed in *glo-1* mutants (Fig. S7). For example, abundances of unmodified ascarosides, e.g. ascr#3 (**9**) and ascr#10 (**10**), or metabolites representing 4′-modifications, e.g. indole-3-carboxylic acid (**11**) and octopamine succinate (**12**), were not significantly perturbed in the mutant (Figs. 1a, S6a and S7). In contrast, a subset of modular ascaroside glucose esters (e.g. iglas#1 (**13**) and glas#10 (**14**), Fig. 1e), was strongly increased in *glo-1* mutants (Fig. S6b). These results confirm that the LROs function as a central hub for the biosynthesis of most modular ascarosides, with the exception of a subset of ascarosylated glucosides, whose increased production in *glo-1* mutants may be indicative of a shunt pathway for ascarosyl-CoA derivatives[29-31], which represent plausible precursors for modular ascarosides modified at the carboxy terminus.

Next, we analyzed the most prominent MS^2 clusters representing previously uncharacterized metabolites whose production is abolished or strongly reduced in *glo-1* mutants (Fig. 2). Detailed analysis of their MS^2 spectra indicated that they may represent a large family of modular hexose derivatives incorporating moieties from diverse primary metabolic pathways. For example, MS^2 spectra from clusters **I**, **II**, and **III** of the positive-ionization network suggested phosphorylated hexose glycosides of indole, anthranilic acid, tyramine, or octopamine, which are further decorated with a wide variety of fatty acyl moieties derived from fatty acid or amino acid metabolism, for example nicotinic acid, pyrrolic acid, or tiglic acid (Fig. 2, Table S1)[16,32]. Given the previous identification of the glucosides iglu#1/2 (**15/16**, Fig. 2e) and angl#1/2 (**17/18**), we hypothesized that clusters **I**, **II**, and **III** represent a modular library of glucosides, in which *N*-glucosylated indole, anthranilic acid, tyramine, or octopamine[33] serve as scaffolds for attachment of diverse building blocks. To further support these structural assignments, a series of modular metabolites based on *N*-glucosylated indole ("iglu") were selected for total synthesis. Synthetic standards for the non-phosphorylated parent compounds of iglu#4 (**19**), iglu#6 (**20**), iglu#8 (**21**), and iglu#10 (**22**) matched HPLC retention times and MS^2 spectra of the corresponding natural compounds (Fig. S8), confirming their structures and enabling tentative structural assignments for a large number of additional modular glucosides, including their

phosphorylated derivatives, e.g. iglu#12 (**23**), iglu#41 (**24**), angl#4 (cluster **II**, **25**), and tyglu#4 (cluster **III**, **26**) (Fig. 2). The proposed structures include several glucosides of the neurotransmitters tyramine and octopamine, whose incorporation could be verified by comparison with data from a recently described feeding experiment with stable isotope labeled tyrosine[33]. Similar to ascaroside biosynthesis, the production of modular glucosides is life stage dependent; for example, production of specific tyramine glucosides peaks at the L3 larval stage, whereas production of angl#4 increases until the adult stage (Figs. S9 and S10). Notably, modular glucosides were detected primarily as their phosphorylated derivatives, as respective non-phosphorylated species were generally less abundant. In contrast to most ascarosides, the phosphorylated glucosides are more abundant in the *endo*-metabolome (metabolites inside the worm) than the *exo*-metabolome (excreted metabolites), suggesting that phosphorylated glucosides may be specifically retained in the body (Fig. S9).

As in the case of modular ascarosides, the abundances of putative building blocks of the newly identified modular glucosides were not strongly perturbed in *glo-1* mutants. For example, abundances of anthranilic acid, indole, octopamine, and tyramine were not significantly affected in *glo-1* null animals (Fig. S11). Notably, abundances of the glucosides scaffold, e.g. iglu#1 and angl#1, were also largely unaltered or even slightly increased in *glo-1* mutants (Fig. S11). In addition, production of some of the identified modular glucosides, e.g. iglu#5, is reduced but not fully abolished in *glo-1* worms (Fig. S8).

To confirm our results, we additionally compared the *glo-1* metabolome with that of *glo-4* mutants. *glo-4* encodes a predicted guanyl-nucleotide exchange factor acting upstream of *glo-1*, and like *glo-1* mutants, *glo-4* worms do not form LROs[17]. We found that the *glo-4* metabolome closely resembles that of *glo-1* worms, lacking most of the modular ascarosides and ascarosides detected in wildtype worms (Fig. S6c). Correspondingly, similar sets of compounds are upregulated in *glo-1* and *glo-4* mutants relative to wildtype, including ascarosyl glucosides and ascaroside phosphates. Compounds accumulating in *glo-1* and *glo-4* mutant worms further include a diverse array of small peptides (primarily three to six amino acids), consistent with the proposed role of LROs in the breakdown of peptides derived from proteolysis (Fig. S12)[34]. Taken together, our results indicate that, in addition to their roles in the degradation of metabolic waste, the LROs serve as hotspots of biosynthetic activity, where building blocks from diverse primary metabolic pathways are attached to glucoside and ascaroside scaffolds (Fig. 1a).

AChE homologs are required for modular assembly. Comparing the relative abundances of different members of the identified families of modular glucosides and ascarosides, it appears

that combinations of different building blocks and scaffolds are highly specific, suggesting the presence of dedicated biosynthetic pathways. For example, uric acid glucoside, gluric#1 (**6**), is preferentially combined with an the ascaroside bearing a 7-carbon side chain (to form uglas#11, **3**), whereas ascarosides bearing a 9-carbon side chain are preferentially attached to the anomeric position of free glucose, as in glas#10 (**14**)[2,8]. Similarly, tiglic acid is preferentially attached to indole and tyramine glucosides but not to anthranilic acid glucosides (Table S1). Given that 4′-modification of ascarosides in *P. pacificus* and *C. elegans* have been shown to require *cest* homologs, we hypothesized that the biosynthesis of other modular ascarosides as well as the newly identified glucosides may be under the control of *cest* family enzymes[23,26]. From a list of 44 *uar-1* homologs from BLAST analysis (Table S2), we selected seven for further study (Fig. 3a, S2). The selected homologs are predicted to have intestinal expression, one primary site of small molecule biosynthesis in *C. elegans*, and are closely related to the UAR-1 gene, while representing different sub-branches of the phylogenetic tree. Utilizing a recently optimized CRISPR/Cas9 method, we obtained two null mutant strains for each of the five selected genes[35]. Mutants for the remaining two homologs, *ges-1* and *cest-6*, had been previously obtained (Table S3). We then analyzed the *exo*- and *endo*-metabolomes of this set of mutant strains by HPLC-HRMS to identify features that are absent or strongly downregulated in null mutants of a specific candidate gene compared to wild type worms and all other mutants in this study. We found that two of the seven tested homologs (*cest-1.1*, *cest-2.2*) are defective in the production of two different families of modular ascarosides, whereas *cest-4* mutants were defective in the biosynthesis of a specific subset of modular indole glucosides (Fig. 3). The metabolomes of mutants for the remaining four *cest* homologs did not exhibit any significant differences compared to wildtype under the tested conditions.

Analysis of the metabolomes of the two *cest-2.2* null mutants revealed loss of dauer pheromone component and male attractant ascr#8 (**2**) as well as of the closely related ascr#81 (**27**) and ascr#82 (**28**) (Fig. 3b, S13a). Biosynthetically, the ascr#8 family of ascarosides are derived from amide formation between ascr#7 (ΔC7, **7**) and folate-derived *p*-aminobenzoic acid (PABA, **8**), PABA-glutamate (**29**), or PABA-diglutamate, respectively. We did not detect any significant reduction in the production of plausible ascr#8 precursors, including PABA and PABA-glutamate, or ascr#7 (Fig. 3c, S14b). Biosynthesis of ascr#8, ascr#81, and ascr#82 was recovered in *cest-2.2* mutant worms in which the *cest-2.2* sequence had been restored to wild type using CRISPR/Cas9 (Fig. 3c, S15b). These results indicate that CEST-2.2 is required specifically for biosynthesis of the amide linkage between the carboxy terminus of ascr#7 and PABA derivatives, in contrast to the implied functions of UAR-1, CEST-8, CEST-3, and CEST-

9.2, which are involved in the formation of ester bonds between various head groups and the 4′-hydroxy group of ascarylose[23,26].

In the two *cest-1.1* null mutants, biosynthesis of the nucleoside-like ascaroside uglas#1 (**30**) and its phosphorylated derivative uglas#11 (**3**) was abolished (Fig. 3d, S13c). uglas#1 and uglas#11 are derived from the attachment of ascr#1, bearing a seven carbon (C7) side chain, to the uric acid gluconucleoside gluric#1 (**6**). Production of ascr#1 (**5**) and gluric#1 (**6**), representing plausible building blocks of uglas#1 (**30**), was not reduced (Fig. S14a). Furthermore, production of uglas#14 (**31**) and uglas#15 (**32**), isomers of uglas#1 and uglas#11 bearing the ascarosyl moiety at the 6′ position instead of the 2′ position, was not abolished but rather slightly increased in the *cest-1.1* mutants (Fig. 3d-e). These results indicate that CEST-1.1 is required for the formation of the ester bond specifically between ascr#1 (**5**) and the 2′-hydroxyl group in gluric#1. As in the case of *cest-2.2*, biosynthesis of uglas#1 and uglas#11 was fully recovered in *cest-1.1* mutant worms in which the *cest-1.1* sequence had been restored to wild type using CRISPR/Cas9 (Fig. 3f, S15a).

Previous work implicated *cest-1.1* with longevity phenotypes associated with argonaute-like gene 2 (*alg-2*)[36]. *alg-2* mutant worms are long lived compared to wild type and their long lifespan was further shown to require the *daf-16*, the sole ortholog of the FOXO family of transcription factors in *C. elegans*, as well as *cest-1.1*. Moreover, uglas#11 biosynthesis is significantly increased in mutants of the insulin receptor homolog *daf-2*, a central regulator of lifespan in *C. elegans* upstream of *daf-16*.[12] These findings suggest the possibility that the production of uglas ascarosides underlies the *cest-1.1*-dependent extension of adult lifespan in *C. elegans*.

In contrast, comparative metabolomic analysis of the *cest-4* mutant strains did not reveal any defects in the biosynthesis of known ascarosides. Instead, we found that the levels of a specific subset of modular anthranilic acid (**33**) bearing indole glucosides, including iglu#3 (**34**) and its phosphorylated derivative iglu#4 (**35**) were abolished in the *cest-4* mutant worms (Fig. 3g, S13b). Abundances of the putative precursor glucosides, iglu#1 (**15**) and iglu#2 (**16**), were not significantly changed in *cest-4* (Fig. 3h, S14c). Notably, production of other indole glucosides, e.g. iglu#6 (**36**) and iglu#8 (**37**), was not significantly reduced in *cest-4* worms (Fig. 3i, S16). Biosynthesis of iglu#3 and iglu#4 was restored to wild type levels in genetic revertant strains for *cest-4* (Fig. 3h, S15c). Therefore, it appears that *cest-4* is specifically required for attachment of anthranilic acid to the 6′ position of glucosyl indole precursors, whereas attachment of tiglic acid, nicotinic acid, and other moieties is *cest-4*-independent (Fig. 3i, S16). The role of *cest-4* in the biosynthesis of the iglu family of modular glucosides thus parallels that

of *cest-1.1* in the biosynthesis of the uglas ascarosides: whereas *cest-4* appears to be required for the attachment of anthranilic acid (**33**) to the 6' position of a range of indole glucosides, *cest-1.1* is required for attaching the ascr#1 side chain to the 2′ position in uric acid glucosides.

AChE homologs localize to the LROs. All *cest* homologs selected for this study exhibit domain architectures typical of the α/ß-hydrolase superfamily of proteins, including a conserved catalytic triad, and further contain a predicted disulfide bridge, as in mammalian AChE[37] (Fig. S17). The *cest* genes also share homology with neuroligin, a membrane bound member of the α/ß-hydrolase fold family, that mediates the formation and maintenance of synapses between neurons[38]. Sequence analysis suggests that five of the seven CEST homologs studied here are membrane anchored (Fig. S18), given the presence of a predicted *C*-terminal transmembrane domain[39] (consisting of ~20 residues), with the *N* terminus on the luminal side of a vesicle or organelle (Fig. S18). Since the production of all so far identified *cest*-dependent metabolites is abolished in *glo-1* mutants and thus appears to require the LROs, it seemed likely that the CEST proteins localize to the LROs in the *C. elegans* intestine. To test this idea, we created mutant strains that express *cest-2.2* either *N*- or *C*-terminally tagged with mCherry at the native genomic locus. The red fluorescent mCherry was chosen because of the strong green autofluorescence of the LROs[16]. We confirmed that the mutant strains are still able to produce ascr#8 (**2**), #81 (**27**), and #82 (**28**) (Fig. 4a), indicating that CEST-2.2 remained functional in the tagged strains. Using fluorescence microscopy, we found that the mCherry signal co-localized with the autofluorescence of the LROs, for both *N*- and *C*-terminally tagged *cest-2.2* (Fig. 4b). Given the highly conserved sequences of the *cest* genes and the requirement of LROs for production of all metabolites so far shown to be *cest*-dependent, our results suggest these *cest* homologs function in the LROs to mediate the biosynthesis of specific sets of modular metabolites.

***Glo-1*-dependent metabolites in *C. briggsae*.** In addition to *C. elegans* and *P. pacificus*, modular ascarosides have been reported from several other *Caenorhabditis* species[40,41], including *C. briggsae*[20,42]. To assess whether the role of LROs in the biosynthesis of modular metabolites is conserved across species, we created two *Cbr-glo-1* (CBG01912.1) knock-out strains using CRISPR/Cas9. As in *C. elegans*, *Cbr-glo-1* mutant worms lacked autofluorescent LROs, which are prominently visible in wildtype *C. briggsae* (Fig. S19). Comparative metabolomic analysis of the *endo*- and *exo*-metabolomes of wildtype *C. briggsae* and the *Cbr-glo-1* mutant strains revealed that biosynthesis of all known modular ascarosides is abolished in

Cbr-glo-1 worms, including the indole carboxy derivatives icas#2 (**38**) and icas#6 (**39**), which are highly abundant in wildtype *C. briggsae* (Fig. 5a).[20] In addition, the *C. briggsae* MS^2 networks included several large *Cbr-glo-1*-dependent clusters representing modular glucosides, including many of the compounds also detected in *C. elegans*, e.g. iglu#4 and angl#4. As in *C. elegans*, production of unmodified glucoside scaffolds, e.g. iglu#1 (**15**) and angl#1 (**17**), was not reduced or increased in *Cbr-glo-1* mutants, whereas biosynthesis of most modular glucosides derived from attachment of additional moieties to these scaffolds was abolished (Fig. 5b). Taken together, these results indicate that the role of LROs as a central hub for the assembly of diverse small molecule architectures, including modular glucosides and ascarosides, may be widely conserved among nematodes (Fig. 5c).

Discussion

Taken together, our results demonstrate that in *C. elegans* intestinal LROs play a central role in the biosynthesis of several large compound families that are derived from combinatorial assembly of primary metabolism-derived building blocks via cholinesterases. The *glo-1*-dependent LROs co-exist with conventional lysosomes and are perhaps most closely related to mammalian melanosomes, whose maturation requires two *glo-1* orthologs, the GTPases RAB32 and RAB38[43]. Lysosomes and LROs are generally presumed to function primarily in autophagy, phagocytosis, and the hydrolytic degradation of proteins, and Rab32 family GTPases have been shown to be required for these processes in diverse organisms[44].

In contrast, our findings indicate that, in *C. elegans*, the metabolic roles of LROs extend beyond catabolism. We show that the LROs function as an assembly hub for the biosynthesis of complex molecular architectures that combine diverse building blocks from amino acid, nucleoside, carbohydrate, and lipid metabolism via ester and amide bonds. Consistent with the notion that lysosomes and LROs are degradation hotspots, many of the building blocks of the identified modular ascarosides and glucosides are derived from catabolic pathways, for example, anthranilic acid is derived from tryptophan catabolism, uric acid stems from purine metabolism, and the short chain ascarosides are the end products of peroxisomal *β*-oxidation of very long-chain precursors.

Notably, our results demonstrate that the modular assembly paradigm extends beyond ascarosides. The modular glucosides represent a previously unknown family of nematode metabolites. In contrast to the well-established role of modular ascarosides as pheromones, it is unknown whether modular glycosides serve specific biological functions, e.g., as signaling molecules; however, their specific biosynthesis via *cest-4* as well as their life stage-dependent

production strongly supports this hypothesis (Fig. S10). Like the ascaroside pheromones, some modular glucosides are excreted into the media, suggesting that they could be involved in inter-organismal communication. Identifying developmental and environmental conditions that affect modular glucoside production, as well as a more comprehensive understanding of their biosyntheses, may help uncover potential signaling and other biological roles. In particular, the apparent peroxisomal origin of the ascaroside scaffolds suggests a link between peroxisome and LRO activity, perhaps via pexophagy[45], and characterization of the role of autophagy for LRO-dependent metabolism may contribute to uncovering the functions of modular glucoside and ascarosides.

The high degree of selectivity in which different building blocks are combined in the modular ascarosides and glucosides strongly suggests that these compounds, despite their numbers and diversity, represent products of dedicated enzymatic pathways, as has recently been established for 4′-acylated ascarosides. Our results revealed a wider range of biosynthetic functions associated with *cest* homologs, including esterification and amide formation at the carboxy terminus of ascarosides and acylation of glucosides (Figure 5c). Notably, all *cest* homologs characterized so far appear to have a narrow substrate scope, further supporting the view that the resulting selectively assembled molecular architectures serve dedicated functions.

All CEST proteins that so far have been associated with modular metabolite assembly contain membrane-anchors and exhibit domain architectures typical of serine hydrolases of the AChE family, including an α/β-hydrolase fold, a conserved catalytic serine-histidine-glutamate triad, and bridging disulfide cysteines (Fig. S17)[37]. Based on our localization experiments showing delivery of CEST-2.2 to the LROs, we envision an enzyme anchored along the luminal face where catalysis could be activated at low pH. While our efforts at heterologous expression of CEST proteins were unsuccessful and the exact biosynthetic mechanisms remain to be elucidated, we hypothesize that CEST proteins, after translating from the endomembrane system to the acidic LROs, partake in the assembly of diverse ascaroside or glucoside-based architectures via acyltransfer from corresponding activated intermediates, e.g. CoA or phosphate esters[37,46]. α/β-hydrolase fold enzymes are functionally highly diverse[47] and include esterases, peptidases, as well as oxidoreductases and lyases, serving varied biosynthetic roles in animals, plants,[48] and bacteria[49]. While acyltransferase activity is often observed as a side reaction for esterases and lipases, α/β-hydrolase fold enzymes have been shown to function as dedicated acyltransferases, e.g. in microbial natural product biosyntheses[47,50].

Finally, although the LROs appear to act as central metabolic hubs for the biosynthesis of most modular metabolites we have detected so far, it is notable that some modular

ascarosides, e.g. iglas#1 (**13**), and modular glucosides, e.g. iglu#6 (**36**) and iglu#8 (**37**), do not appear to be *glo-1*-dependant (Fig. S8). These findings suggest that cell compartments other than the LROs contribute to modular metabolite biosynthesis and may also indicate that not all CEST proteins are delivered to the LROs. Similarly, *glo-1* mutants continue to generate the simple glucosides and ascarosides that serve as scaffolds for further elaboration in the LROs, which may be derived from UDP-glycosyltransferases[51].

Reminiscent of the role of AChE for neuronal signal transduction in animals, it appears that, in *C. elegans*, AChE homologs have been co-opted to establish additional signal transduction pathways that are based on a modular chemical language, for inter-organismal communication, and perhaps also intra-organismal signaling. The biosynthetic functions of most of the 200 serine hydrolases in *C. elegans*, including more than 30 additional *cest* homologs, remain to be assessed, and it seems likely that this enzyme family contributes to the biosynthesis of a large number of additional, yet unidentified compounds. Similarly, the exact enzymatic roles of many families of mammalian serine hydrolases have not been investigated using HRMS-based untargeted metabolomics. Our results may motivate a systematic characterization of metazoan serine hydrolases, with regard to their roles in metabolism and small molecule signaling, associated enzymatic mechanisms, and cellular localization.

Figures

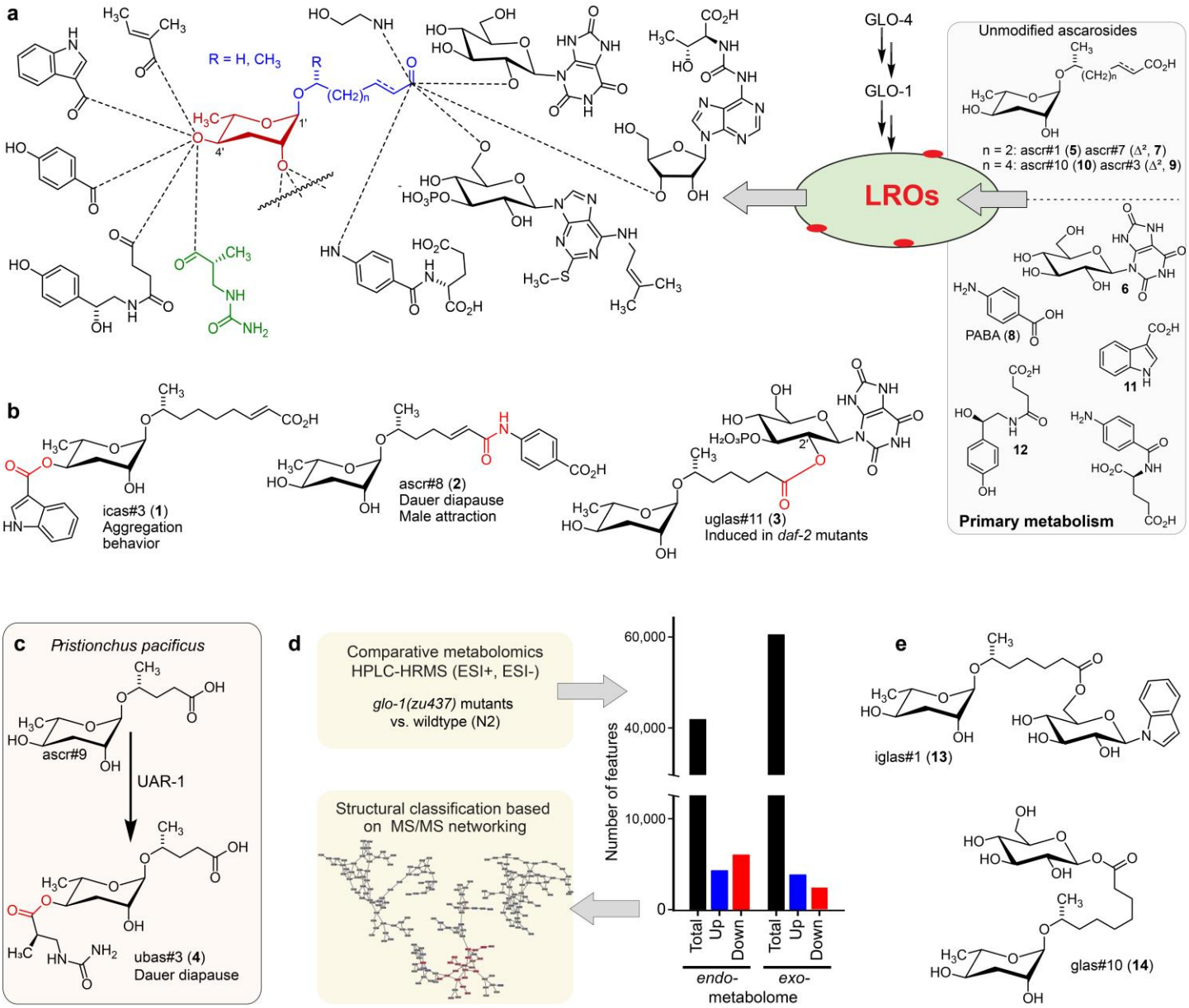


Figure 1: (a) Modular ascarosides are assembled from simple ascarosides, e.g. ascr#1 (**5**) or ascr#3 (**9**), and other primary metabolism-derived building blocks, e.g. glucosyl uric acid (**6**), *p*-aminobenzoic acid (PABA, **8**) indole-3-carboxylic acid (**11**), or succinyl octopamine (**12**). We hypothesize that the *glo-1*-dependent LROs play a central role in their biosynthesis. (b) Examples for modular ascarosides and their biological context. (c) UAR-1 in *P. pacificus* converts simple ascarosides into the 4′-ureidoisobutyric acid-bearing ascarosides, e.g. ubas#3 (**4**). (d) Strategy for comparative metabolomic analysis of LRO-deficient *glo-1* mutants. (e) Example for modular ascarosides whose production is increased in *glo-1* mutants.

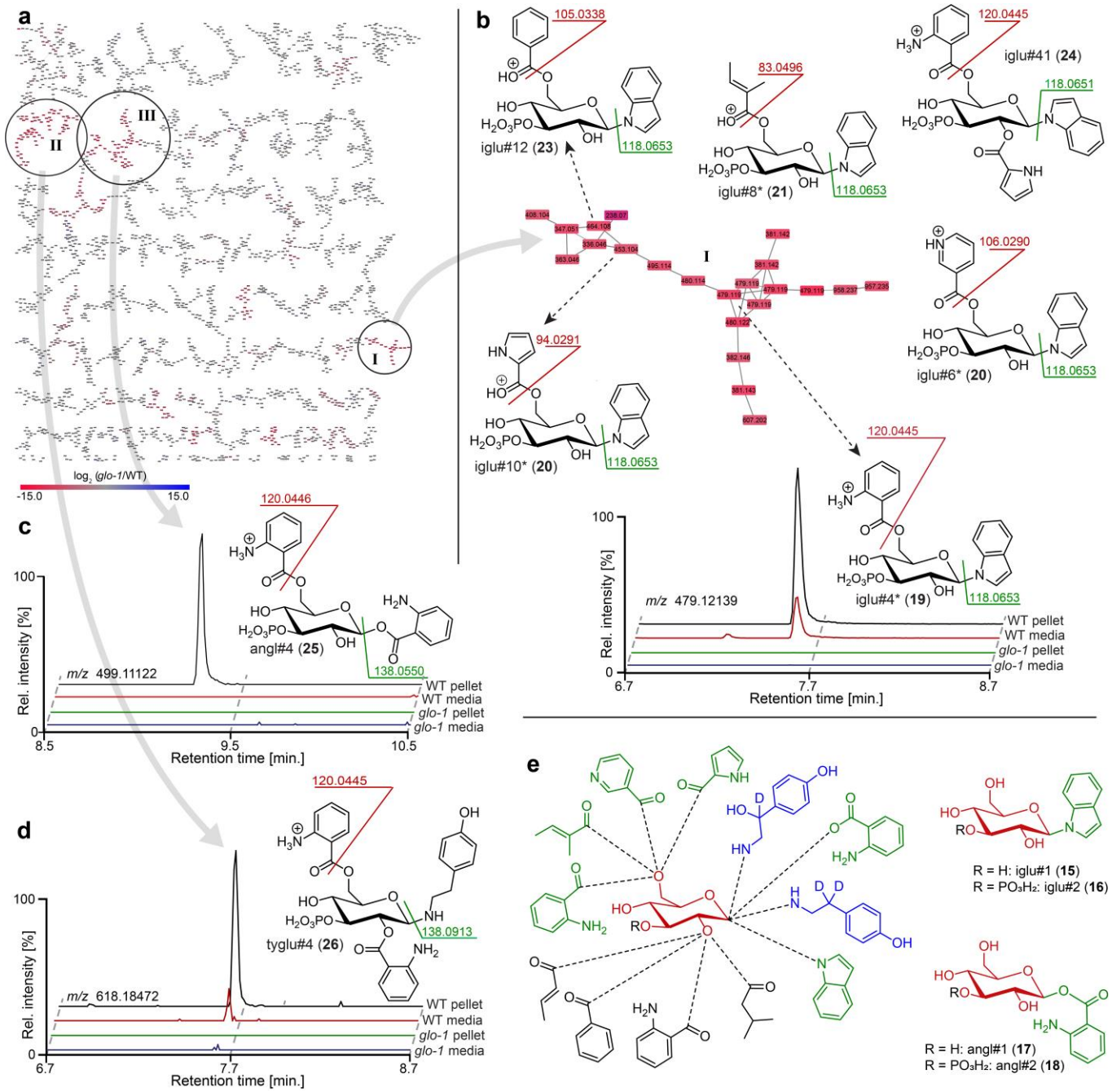


Figure 2: (a) Partial MS^2 network (positive ion mode) for *C. elegans endo*-metabolome highlighting three clusters of modular glucosides that are down regulated in the *glo-1* mutants (also see Fig. S1-4). Red represents downregulated and blue upregulated features compared to wildtype *C. elegans*. (b) Cluster **I** features several modular indole glucoside derivatives. Shown structures were based on MS^2 fragmentation patterns, also see Table S1. Compounds whose non-phosphorylated analogs were synthesized are marked (*).Shown ion chromatograms demonstrate loss of iglu#4 in *glo-1* mutants. (c,d) Examples for modular glucosides detected as

part of clusters **II** and **III**. Ion chromatograms show abolishment of angl#4 (**25**) (c) and tyglu#4 (**26**) (d) production in *glo-1* mutants. (e) Modular glucosides are derived from combinatorial assembly of a wide range of building blocks. Incorporation of moieties was confirmed via total synthesis of example compounds (green) or stable isotope labeling (blue). For all compounds, 3-phosphorylation was assigned based on the established structures of iglu#2 (**16**), angl#2 (**18**), and uglas#11 (**3**).

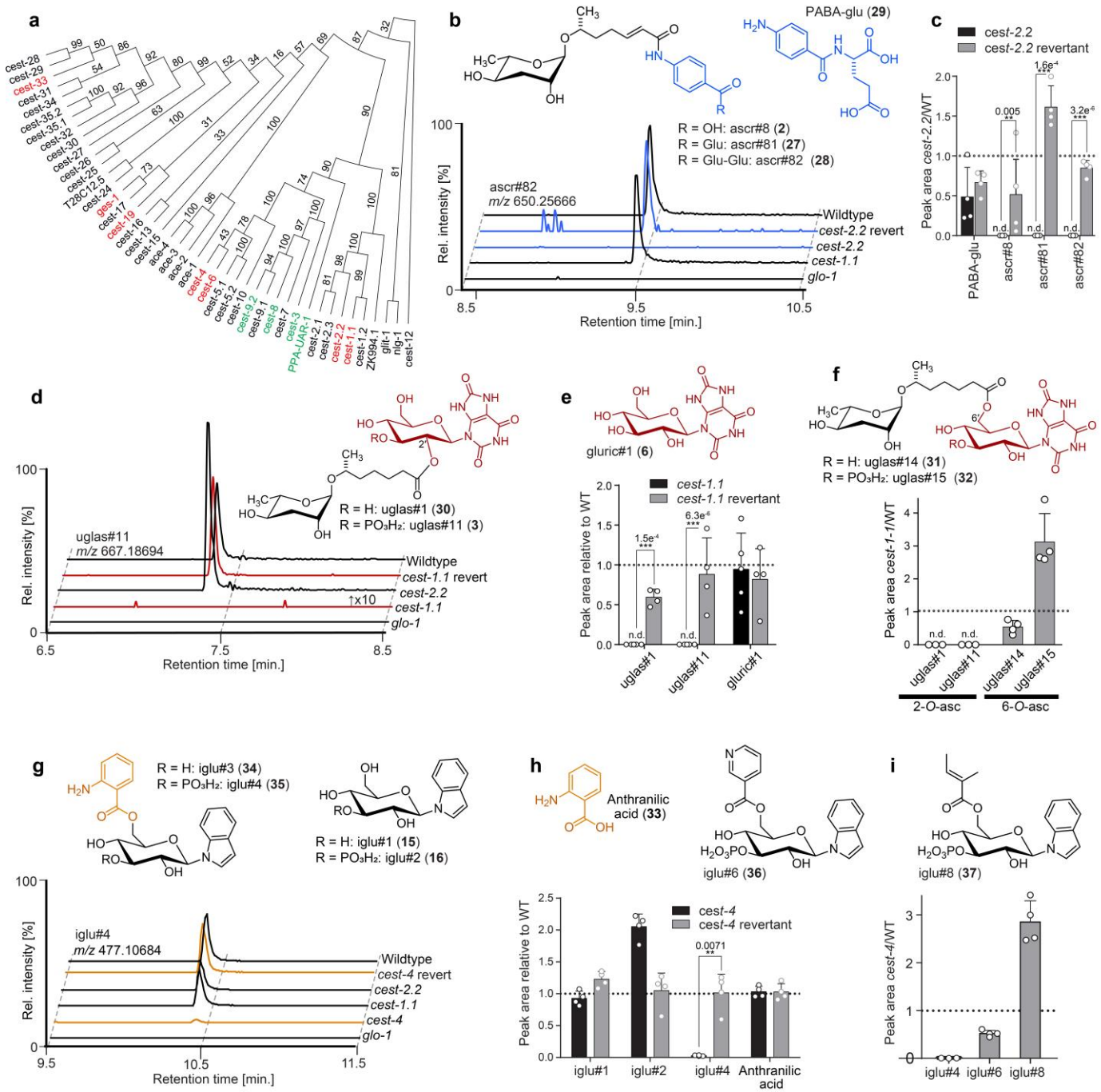


Figure 3: (a) Phylogenetic tree relating *P. pacificus uar-1* to homologous predicted genes in *C. elegans*. *Ppa-uar-1*, *cest-3*, *cest-8*, *cest-9.2* (green) have been shown to mediate ester formation at the 4′-position of ascarosides in *P. pacificus* and *C. elegans*. Genes shown in red color were selected for the current study. (b,c) Production of ascr#8 (**2**), ascr#81 (**27**), and ascr#82 (**28**) is abolished in *cest-2.2* mutants Isogenic revertant strains of the *cest-2.2* null mutants in which the STOP-IN cassette was precisely excised, demonstrate wildtype-like

recovery of the associated metabolite. (d,e) Production of uglas#1 and uglas#11 is abolished in *cest-1.1* mutants and recovered in genetic revertants. (f) Biosynthesis of positional isomers uglas#14 (**31**) and uglas#15 (**32**) is unaltered or increased in *cest-1.1* mutants (f). (g,h) Production of the anthranilic acid-modified glucoside iglu#4 is largely abolished *cest-4* mutants and fully recovered in genetic revertants. (i) Production of iglu#6 (**36**) and iglu#8 (**37**), whose structures are closely related to that of iglu#4, is not abolished in *cest-4* mutants. Ion chromatograms in panels b, d, and g further demonstrate abolishment in *glo-1* mutants. n.d., not detected. Error bars are standard deviation of the mean, and p-values are depicted in the Figure.

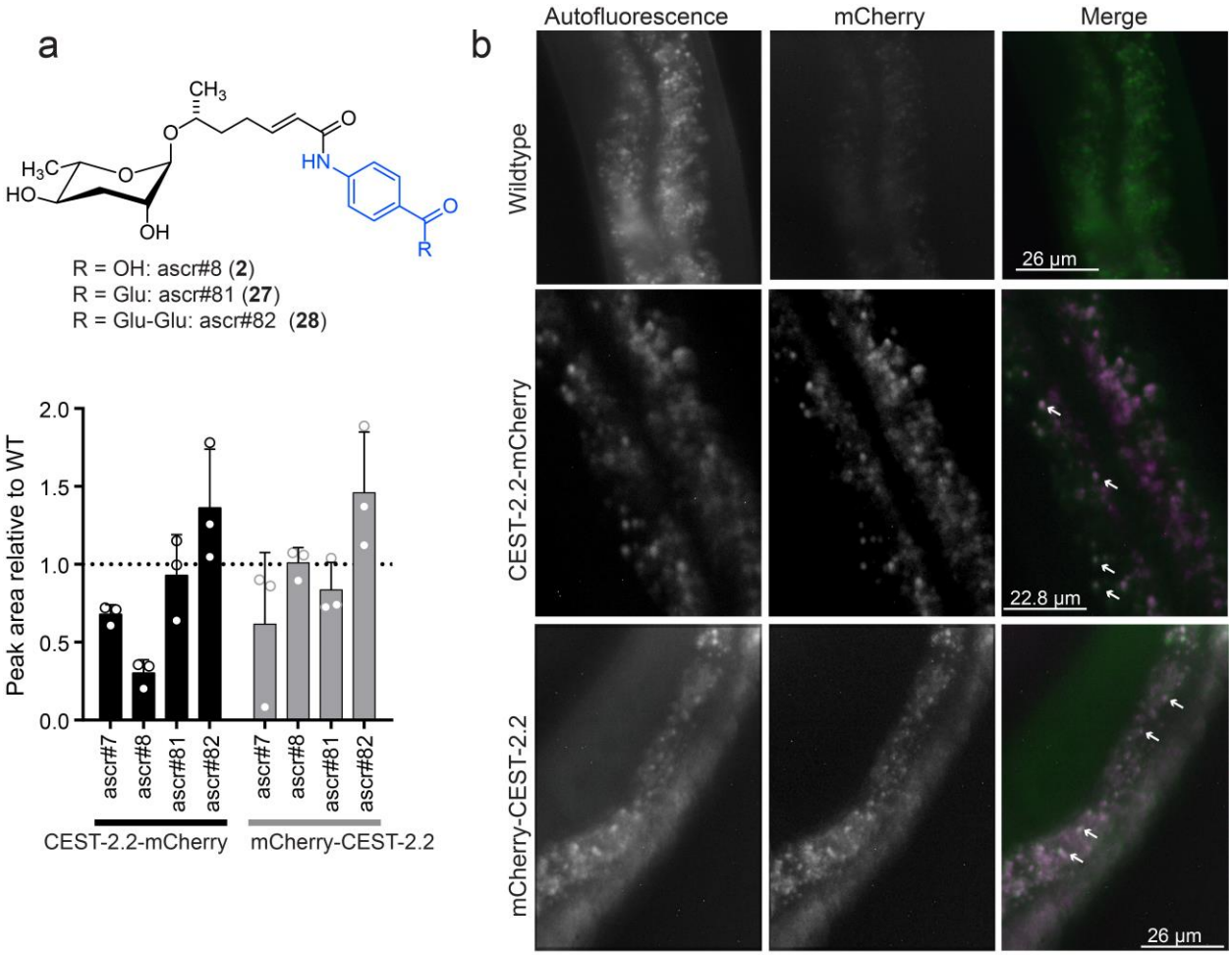


Figure 4: (a) Relative amounts of *cest-2.2* dependent metabolites in *N*- and *C*-terminally mCherry-tagged CEST-2.2. (b) Localization of CEST-2.2 to acidic gut granules in *C. elegans*. Top, wildtype (N2) control; middle, *C*-terminally tagged CEST-2.2; bottom, *N*-terminally tagged CEST-2.2. White arrows depict co-localization of mCherry and autofluorescent signals.

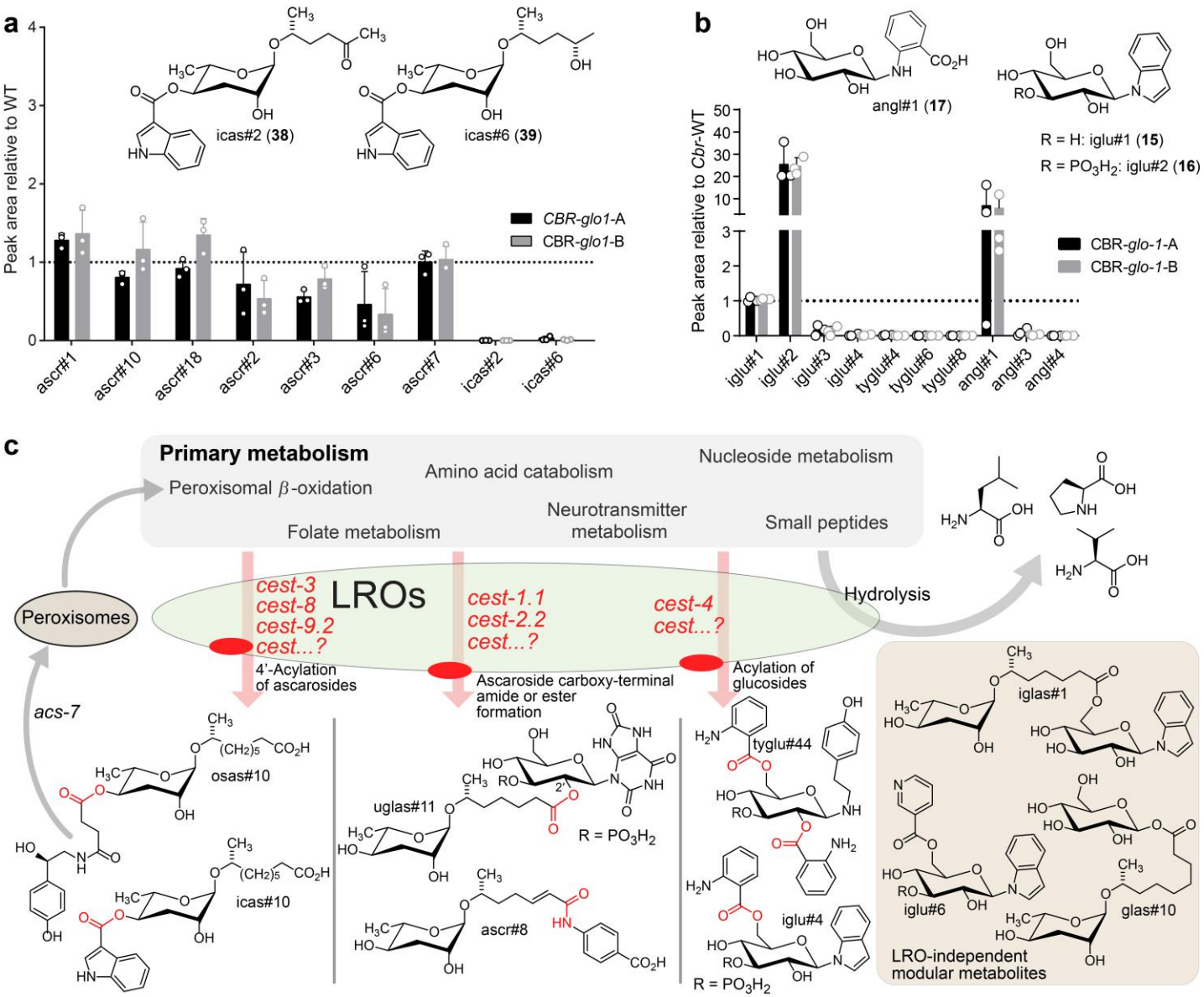


Figure 5: Relative abundance of (a) simple and modular ascarosides and (b) simple and modular glucosides in the *endo*-metabolome of *Cbr-glo-1* mutants relative to wildtype *C. briggsae*. n.d., not detected. (c) Model for modular metabolite assembly. CEST proteins (membrane-bound in the LROs, red) mediate attachment of diverse primary metabolism-derived building blocks to glucose scaffolds and peroxisomal *β*-oxidation-derived ascarosides via ester and amide bonds. Some of the resulting modular ascarosides may undergo additional peroxisomal *β*-oxidation following activation by *acs-7* [25].

Author Contributions

The manuscript was written through contributions of all authors and all authors have given approval to the final version of the manuscript.

[‡]These authors contributed equally.

Acknowledgements

This research was funded by an NIH Chemical Biology Interface (CBI) Training Grant 5T32GM008500 (to B.C.), National Institutes of Health grants R35 GM131877 (to F.C.S.), and R24OD023041 (to P.W.S.). F.C.S. is a Faculty Scholar of the Howard Hughes Medical Institute. We thank WormBase for sequences, Tsui-Fen Chou for Cas9 protein, Ying (Kitty) Zhang for assistance with NMR spectroscopy, and Navid Movahed for assistance with mass spectrometry.

Competing Interests.

The authors declare no competing interests.

References

1 da Silva, R. R., Dorrestein, P. C. & Quinn, R. A. Illuminating the dark matter in metabolomics. *Proc Natl Acad Sci U S A* **112**, 12549-12550, doi:10.1073/pnas.1516878112 (2015).

2 Artyukhin, A. B. *et al.* Metabolomic "Dark Matter" Dependent on Peroxisomal beta-Oxidation in Caenorhabditis elegans. *J Am Chem Soc* **140**, 2841-2852, doi:10.1021/jacs.7b11811 (2018).

3 Girard, L. R. *et al.* WormBook: the online review of Caenorhabditis elegans biology. *Nucleic Acids Res* **35**, D472-475, doi:10.1093/nar/gkl894 (2007).

4 Schroeder, F. C. Modular assembly of primary metabolic building blocks: a chemical language in C. elegans. *Chem Biol* **22**, 7-16, doi:10.1016/j.chembiol.2014.10.012 (2015).

5 Butcher, R. A. Decoding chemical communication in nematodes. *Nat Prod Rep* **34**, 472-477, doi:10.1039/c7np00007c (2017).

6 Butcher, R. A., Fujita, M., Schroeder, F. C. & Clardy, J. Small-molecule pheromones that control dauer development in Caenorhabditis elegans. *Nat Chem Biol* **3**, 420-422, doi:10.1038/nchembio.2007.3 (2007).

7 Jeong, P. Y. *et al.* Chemical structure and biological activity of the Caenorhabditis elegans dauer-inducing pheromone. *Nature* **433**, 541-545, doi:10.1038/nature03201 (2005).

8 von Reuss, S. H. *et al.* Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small-molecule signals in C. elegans. *J Am Chem Soc* **134**, 1817-1824, doi:10.1021/ja210202y (2012).

9 Artyukhin, A. B. *et al.* Succinylated octopamine ascarosides and a new pathway of biogenic amine metabolism in Caenorhabditis elegans. *J Biol Chem* **288**, 18778-18783, doi:10.1074/jbc.C113.477000 (2013).

10 Bose, N. *et al.* Natural variation in dauer pheromone production and sensing supports intraspecific competition in nematodes. *Curr Biol* **24**, 1536-1541, doi:10.1016/j.cub.2014.05.045 (2014).

11 Aprison, E. Z. & Ruvinsky, I. Counteracting Ascarosides Act through Distinct Neurons to Determine the Sexual Identity of C. elegans Pheromones. *Curr Biol* **27**, 2589-2599.e2583, doi:10.1016/j.cub.2017.07.034 (2017).

12 Curtis, B. J. *et al.* Identification of Uric Acid Gluconucleoside–Ascaroside Conjugates in Caenorhabditis elegans by Combining Synthesis and MicroED. *Organic Letters*, doi:10.1021/acs.orglett.0c02038 (2020).

13 Panda, O. *et al.* Biosynthesis of Modular Ascarosides in C. elegans. *Angew Chem Int Ed Engl* **56**, 4729-4733, doi:10.1002/anie.201700103 (2017).

14 Ludewig, A. H. *et al.* An excreted small molecule promotes C. elegans reproductive development and aging. *Nat Chem Biol* **15**, 838-845, doi:10.1038/s41589-019-0321-7 (2019).

15 Srinivasan, J. *et al.* A modular library of small molecule signals regulates social behaviors in Caenorhabditis elegans. *PLoS Biol* **10**, e1001237, doi:10.1371/journal.pbio.1001237 (2012).

16 Coburn, C. & Gems, D. The mysterious case of the C. elegans gut granule: death fluorescence, anthranilic acid and the kynurenine pathway. *Front Genet* **4**, 151, doi:10.3389/fgene.2013.00151 (2013).

17 Hermann, G. J. *et al.* Genetic analysis of lysosomal trafficking in Caenorhabditis elegans. *Mol Biol Cell* **16**, 3273-3288, doi:10.1091/mbc.e05-01-0060 (2005).

18 Dell'Angelica, E. C., Mullins, C., Caplan, S. & Bonifacino, J. S. Lysosome-related organelles. *FASEB J* **14**, 1265-1278, doi:10.1096/fj.14.10.1265 (2000).

19 Luzio, J. P., Hackmann, Y., Dieckmann, N. M. & Griffiths, G. M. The biogenesis of lysosomes and lysosome-related organelles. *Cold Spring Harb Perspect Biol* **6**, a016840, doi:10.1101/cshperspect.a016840 (2014).

20 Dong, C., Dolke, F. & von Reuss, S. H. Selective MS screening reveals a sex pheromone in Caenorhabditis briggsae and species-specificity in indole ascaroside signalling. *Org Biomol Chem* **14**, 7217-7225, doi:10.1039/c6ob01230b (2016).

21 Bergame, C. P., Dong, C., Sutour, S. & von Reuss, S. H. Epimerization of an Ascaroside-Type Glycolipid Downstream of the Canonical beta-Oxidation Cycle in the Nematode Caenorhabditis nigoni. *Org Lett* **21**, 9889-9892, doi:10.1021/acs.orglett.9b03808 (2019).

22 Dolke, F. *et al.* Ascaroside Signaling in the Bacterivorous Nematode Caenorhabditis remanei Encodes the Growth Phase of Its Bacterial Food Source. *Org Lett* **21**, 5832-5837, doi:10.1021/acs.orglett.9b01914 (2019).

23 Falcke, J. M. *et al.* Linking Genomic and Metabolomic Natural Variation Uncovers Nematode Pheromone Biosynthesis. *Cell Chem Biol* **25**, 787-796.e712, doi:10.1016/j.chembiol.2018.04.004 (2018).

24 Rae, R., Schlager, B. & Sommer, R. J. Pristionchus pacificus: A Genetic Model System for the Study of Evolutionary Developmental Biology and the Evolution of Complex Life-History Traits. *CSH Protoc* **2008**, pdb emo102, doi:10.1101/pdb.emo102 (2008).

25 Chen, A. L. *et al.* Pharmacological convergence reveals a lipid pathway that regulates C. elegans lifespan. *Nat Chem Biol* **15**, 453-462, doi:10.1038/s41589-019-0243-4 (2019).

26 Faghih, N. *et al.* A large family of enzymes responsible for the modular architecture of nematode pheromones. *J Am Chem Soc*, doi:10.1021/jacs.0c04223 (2020).

27 Tautenhahn, R., Bottcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 504, doi:10.1186/1471-2105-9-504 (2008).

28 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**, 828-837, doi:10.1038/nbt.3597 (2016).

29 Zhang, X. *et al.* Acyl-CoA oxidase complexes control the chemical message produced by Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **112**, 3955-3960, doi:10.1073/pnas.1423951112 (2015).

30 Zhang, X., Li, K., Jones, R. A., Bruner, S. D. & Butcher, R. A. Structural characterization of acyl-CoA oxidases reveals a direct link between pheromone biosynthesis and metabolic state in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **113**, 10055-10060, doi:10.1073/pnas.1608262113 (2016).

31 Zhang, X., Wang, Y., Perez, D. H., Jones Lipinski, R. A. & Butcher, R. A. Acyl-CoA Oxidases Fine-Tune the Production of Ascaroside Pheromones with Specific Side Chain Lengths. *ACS Chem Biol* **13**, 1048-1056, doi:10.1021/acschembio.7b01021 (2018).

32 Stupp, G. S. *et al.* Chemical detoxification of small molecules by Caenorhabditis elegans. *ACS Chem Biol* **8**, 309-313, doi:10.1021/cb300520u (2013).

33 O'Donnell, M. P., Fox, B. W., Chao, P. H., Schroeder, F. C. & Sengupta, P. A neurotransmitter produced by gut bacteria modulates host sensory behaviour. *Nature*, doi:10.1038/s41586-020-2395-5 (2020).

34 Bird, P. I., Trapani, J. A. & Villadangos, J. A. Endolysosomal proteases and their inhibitors in immunity. *Nat Rev Immunol* **9**, 871-882, doi:10.1038/nri2671 (2009).

35 Wang, H., Park, H., Liu, J. & Sternberg, P. W. An Efficient Genome Editing Strategy To Generate Putative Null Mutants in Caenorhabditis elegans Using CRISPR/Cas9. *G3 (Bethesda)* **8**, 3607-3616, doi:10.1534/g3.118.200662 (2018).

36 Aalto, A. P. *et al.* Opposing roles of microRNA Argonautes during Caenorhabditis elegans aging. *PLoS Genet* **14**, e1007379, doi:10.1371/journal.pgen.1007379 (2018).

37 Soreq, H. & Seidman, S. Acetylcholinesterase--new roles for an old actor. *Nat Rev Neurosci* **2**, 294-302, doi:10.1038/35067589 (2001).

38 Bemben, M. A., Shipman, S. L., Nicoll, R. A. & Roche, K. W. The cellular and molecular landscape of neuroligins. *Trends Neurosci* **38**, 496-505, doi:10.1016/j.tins.2015.06.004 (2015).

39 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580, doi:10.1006/jmbi.2000.4315 (2001).

40 Dong, C., Dolke, F., Bandi, S., Paetz, C. & von Reuss, S. H. Dimerization of conserved ascaroside building blocks generates species-specific male attractants in Caenorhabditis nematodes. *Org Biomol Chem*, doi:10.1039/d0ob00799d (2020).

41 Kanzaki, N. *et al.* Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. *Nat Commun* **9**, 3216, doi:10.1038/s41467-018-05712-5 (2018).

42 von Reuss, S. H. Exploring Modular Glycolipids Involved in Nematode Chemical Communication. *Chimia (Aarau)* **72**, 297-303, doi:10.2533/chimia.2018.297 (2018).

43 Marks, M. S., Heijnen, H. F. & Raposo, G. Lysosome-related organelles: unusual compartments become mainstream. *Curr Opin Cell Biol* **25**, 495-505, doi:10.1016/j.ceb.2013.04.008 (2013).

44 Morris, C. *et al.* Function and regulation of the Caenorhabditis elegans Rab32 family member GLO-1 in lysosome-related organelle biogenesis. *PLoS Genet* **14**, e1007772, doi:10.1371/journal.pgen.1007772 (2018).

45 Sakai, Y., Oku, M., van der Klei, I. J. & Kiel, J. A. Pexophagy: autophagic degradation of peroxisomes. *Biochim Biophys Acta* **1763**, 1767-1775, doi:10.1016/j.bbamcr.2006.08.023 (2006).

46 Vaz, F. M. & Wanders, R. J. A. Carnitine biosynthesis in mammals. *Biochemical Journal* **361**, 417-429, doi:10.1042/bj3610417 (2002).

47 Rauwerdink, A. & Kazlauskas, R. J. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS Catal* **5**, 6153-6176, doi:10.1021/acscatal.5b01539 (2015).

48 Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D. & Noel, J. P. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Current opinion in structural biology* **41**, 233-246, doi:10.1016/j.sbi.2016.08.005 (2016).

49 Zheng, Q. *et al.* An alpha/beta-hydrolase fold protein in the biosynthesis of thiostrepton exhibits a dual activity for endopeptidyl hydrolysis and epoxide ring opening/macrocyclization. *Proc Natl Acad Sci U S A* **113**, 14318-14323, doi:10.1073/pnas.1612607113 (2016).

50 Lejon, S., Ellis, J. & Valegård, K. The last step in cephalosporin C formation revealed: crystal structures of deacetylcephalosporin C acetyltransferase from Acremonium chrysogenum in complexes with reaction intermediates. *J Mol Biol* **377**, 935-944, doi:10.1016/j.jmb.2008.01.047 (2008).

51 Mackenzie, P. I. *et al.* Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics* **15**, 677-685, doi:10.1097/01.fpc.0000173483.13689.56 (2005).