Article

Biosynthesis of selenium-containing small molecules in diverse microorganisms

https://doi.org/10.1038/s41586-022-05174-2

Received: 18 February 2022

Accepted: 2 August 2022

Published online: 7 September 2022

Check for updates

Chase M. Kayrouz[1], Jonathan Huang[1], Nicole Hauser[1] & Mohammad R. Seyedsayamdost[1,2]✉

Selenium is an essential micronutrient in diverse organisms. Two routes are known for its insertion into proteins and nucleic acids, via selenocysteine and 2-selenouridine, respectively[1]. However, despite its importance, pathways for specific incorporation of selenium into small molecules have remained elusive. Here we use a genome-mining strategy in various microorganisms to uncover a widespread three-gene cluster that encodes a dedicated pathway for producing selenoneine, the selenium analogue of the multifunctional molecule ergothioneine[2,3]. We elucidate the reactions of all three proteins and uncover two novel selenium–carbon bond-forming enzymes and the biosynthetic pathway for production of a selenosugar, which is an unexpected intermediate en route to the final product. Our findings expand the scope of biological selenium utilization, suggest that the selenometabolome is more diverse than previously thought, and set the stage for the discovery of other selenium-containing natural products.

Selenium (Se) is an essential trace element across all kingdoms of life and is found mainly in selenoprotein and selenonucleic acid biopolymers[1]. Selenoproteins carry Se in the form of selenocysteine (Sec), which often constitutes the redox active centre in the catalytic reduction of peroxides, disulfides, sulfoxides and aryl iodides[4–8]. Selenonucleic acids carry Se in 5-methylaminomethyl-2-selenouridine, which is incorporated into the wobble position of several tRNAs by various microorganisms[9]. Aside from its presence in important biopolymers, Se has also been detected in place of sulfur in many common sulfur-containing metabolites, such as methionine, cystathionine and dimethyl sulfide, when organisms are grown under Se-enriched conditions. However, sulfur and Se are poorly discriminated by biosynthetic enzymes, and thus the observed selenometabolites are usually nonspecific products of sulfur-utilizing enzymes[10]. Se-specific biosynthetic pathways have so far been limited exclusively to Sec and selenouridine; they encompass a grand total of three Se-specific enzymes. Common to both pathways is selenophosphate synthetase (SelD), which phosphorylates hydrogen selenide to generate selenophosphate (SeP), the substrate for selenocysteine synthase (SelA) and selenouridine synthase (SelU; Fig. 1a). Using SeP, SelA converts seryl-tRNA$_{Sec}$ into selenocysteinyl-tRNA$_{Sec}$ in a pyridoxal-5-phosphate (PLP)-dependent manner, whereas the rhodanese enzyme SelU catalyses a S-to-Se substitution on 2-thiouridine-containing tRNAs[11–13] (Fig. 1a).

Given the taxonomic ubiquity of Se utilization, it is hard to imagine that Se would be limited to these two pathways. Furthermore, both SelA and SelU, the only known C–Se bond-forming enzymes, act on macromolecular substrates, thus leaving no known mechanism for specific incorporation of Se into organic small molecules, notably natural products, which can harbour unusual atoms and complex architectures. We therefore speculated that nature may have indeed evolved additional biosynthetic strategies for endowing small molecules with Se and its unique properties. Guided by a bioinformatic search, we herein report a widespread pathway for incorporation of Se into microbial metabolites, involving two novel and unusual Se–C bond-forming enzymes.

Genome mining for selenometabolite pathways

To identify new selenometabolite pathways, we first envisioned potential genetic characteristics of organisms that may encode them. Given that SeP is the common Se donor to both selenoproteins and selenonucleic acids, we hypothesized that new pathways may follow a similar biosynthetic trajectory. Conveniently, genes involved in the production of natural products are often grouped into biosynthetic gene clusters (BGCs) in microbial genomes, thereby allowing us to formulate a basic criterion for genome mining: the presence of *selD* within a BGC[14].

We retrieved all *selD* genes from the National Center for Biotechnology Information (NCBI) database and, after a strict dereplication process, extracted the genetic contexts of each *selD* gene to search for the presence of common natural product biosynthetic genes. Unfortunately, we were unable to identify obvious instances of *selD* within a natural product BGC. We therefore inverted the search to interrogate *selD* genetic contexts in a hypothesis-free manner and identify genes commonly colocalized with *selD*. Specifically, we quantified the abundance of genes that overlapped with a *selD* open reading frame by one or more base pairs, reasoning that this additional constraint would select for genes whose products are transcriptionally linked to SelD (Fig. 1b). Among the top five *selD*-overlapping genes are *selA* and *selU*, validating the ability of the strategy to identify functionally related genes, along with two genes, *yedF* and *duf3343*, that are thought to serve an as-yet-unidentified role in Se reduction and/or trafficking[15].

The second most commonly overlapping gene, annotated as a putative glycosyltransferase from the TIGR04348 family, caught our attention (Fig. 1b and Supplementary Table 1). Although accounting for nearly 5% of all *selD*-overlapping genes, to the best of our knowledge, TIGR04348 has never been implicated in Se metabolism, nor has any

[1]Department of Chemistry, Princeton University, Princeton, NJ, USA. [2]Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ✉e-mail: mrseyed@princeton.edu

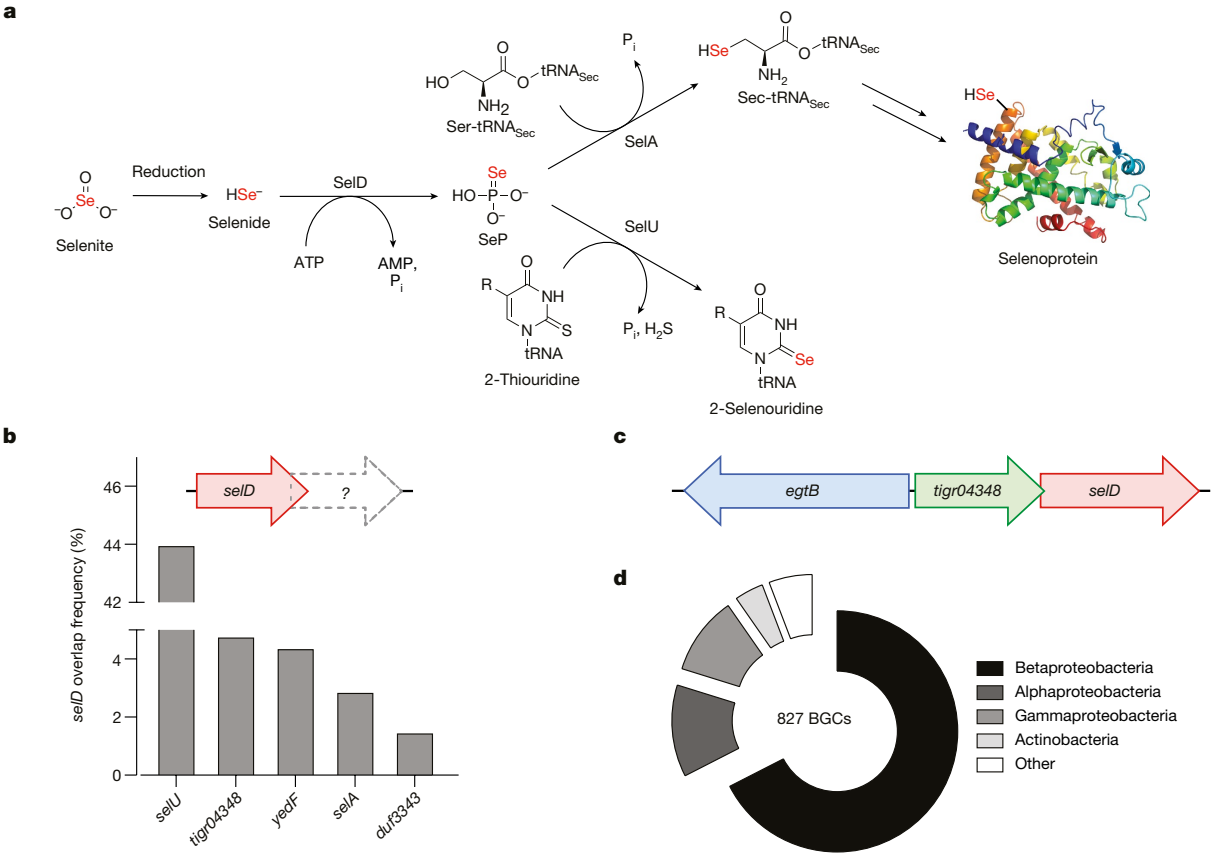

Fig. 1 | Known and new biological pathways for Se incorporation. a, Pathway for insertion of Se into protein (top) and nucleic acid biopolymers (bottom). **b**, Distribution of genes that overlap with all known *selD* sequences in the available genome database. TIGR04348, a previously uncharacterized protein family, is the second most common. **c**, A new potential Se-insertion BGC, consisting of an *egtB* homologue, *tigr04348* and *selD*. **d**, Prevalence of the three-gene cassette in **c**.

member of its protein family been characterized. Furthermore, a closer examination of *selD–tigr04348* genetic contexts revealed a third commonly colocalized gene: a homologue of *egtB*, which encodes the C–S bond-forming enzyme in the biosynthesis of ergothioneine[16] (Fig. 1c), a ubiquitous metabolite found in microorganisms and mammals to which several functions have been ascribed, including antioxidant and cytoprotectant[3]. Over 800 examples of this three-gene cassette were found among diverse bacterial genomes, suggesting that it may encode a widespread pathway dedicated to the biosynthesis of a new selenometabolite (Fig. 1d). This hypothesis is further endorsed by the presence of an additional gene encoding a putative selenate transporter in close proximity to many of the retrieved clusters (Supplementary Fig. 1). We therefore set out to characterize this pathway using metabolomic and biochemical approaches.

Microbial metabolite profiling

We began by performing a metabolomic analysis of two species that have the *selD–egtB–tigr04348* cluster. The actinomycete *Amycolatopsis palatopharyngis* DSM 444832 and the β-proteobacterium *Variovorax paradoxus* DSM 30034 were grown in the presence of sodium selenite (Supplementary Fig. 1). The cells were then pelleted and analysed by high-performance liquid chromatography-coupled mass spectrometry (HPLC–MS), revealing the production of ergothioneine and its Se analogue, selenoneine, by both bacteria (Fig. 2a,b and Supplementary Table 2). Selenoneine has been observed on several occasions in biological samples, with particularly high abundance in the blood and tissues of marine animals[2,17]. It is more reducing than ergothioneine ($E^{o\prime}$ = −0.49 V versus −0.06 V normal hydrogen electrode), making it an effective antioxidant[3,18,19]. The natural origin of selenoneine has thus far remained enigmatic, although researchers have speculated that it may arise from nonspecific incorporation by the ergothioneine pathway[20]. Indeed, upon further genomic examination, both strains were found to have a canonical ergothioneine BGC (*egt*) in addition to the putative selenometabolite BGC. To examine whether selenoneine may be the product of nonspecific Se incorporation by the *egt* cluster, two close relatives of the producing strains, the actinomycete *Streptomyces rimosus* ATCC 10970 and the β-proteobacterium *Burkholderia thailandensis* E264, which encode a canonical *egt* cluster but lack the putative selenometabolite cassette, were analysed in the same manner, revealing exclusive production of ergothioneine, and not selenoneine (Fig. 2a). These observations are consistent with previous reports of the inability of the bacterial ergothioneine biosynthetic machinery to generate selenoneine in vitro[21]. Our results suggest that selenoneine may in fact be the product of the new cluster, which we have termed *sen*, with *senA*, *senB* and *senC* encoding an *egtB* homologue, a putative glycosyltransferase and a *selD* homologue, respectively (Fig. 2c).

If correct, our hypothesis would indicate two distinct biosynthetic routes to ergothioneine and selenoneine. In many bacteria, ergothioneine is produced by *egtABCDE*. Early steps involve an EgtA-catalysed isopeptide bond formation between cysteine and glutamic acid, and *S*-adenosylmethionine (SAM)-dependent trimethylation of histidine by EgtD (Fig. 2d). The products, γ-glutamyl-cysteine (GGC) and hercynine, are then oxidatively linked via a sulfoxide by the atypical nonhaem iron enzyme EgtB. Finally, following EgtC-catalysed hydrolysis of the glutamic acid residue, EgtE performs PLP-dependent C–S bond cleavage to furnish ergothioneine[16]. By contrast, the *sen* cluster contains only three genes. We speculated that SenB utilizes SeP (generated by SenC)

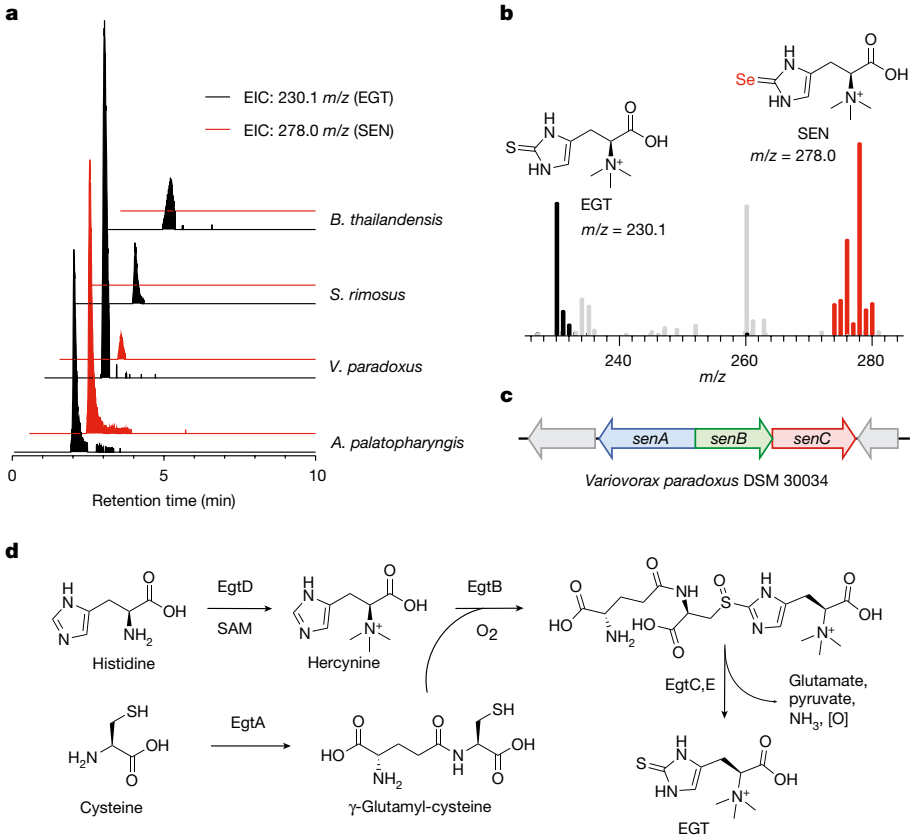

Fig. 2 | Production of selenoneine by bacteria that carry the three-gene *sen* cluster. a, Production of selenoneine (SEN) and ergothioneine (EGT), determined by HPLC–MS analysis, by the strains indicated. Extracted ion chromatograms (EICs) are shown. Only those that have the *sen* cluster synthesize SEN. Results are representative of at least three biological replicates. **b**, HR-MS analysis of *A. palatopharyngis* culture extracts, revealing production of both EGT and SEN. **c**, The three-gene *sen* cluster, consisting of the *egtB* homologue *senA*, a member of an uncharacterized superfamily *senB* and a *selD* homologue *senC*. **d**, EGT biosynthetic pathway.

to generate a new selenometabolite. SenA could then catalyse C–Se bond formation between this new species and hercynine, followed by C–Se bond cleavage to give selenoneine. Hercynine would probably be siphoned from the canonical ergothioneine pathway, although a small subset of *sen* clusters feature a colocalized *egtD* gene (Supplementary Fig. 1).

Reconstitution of selenoneine biosynthesis

To evaluate this proposal and conclusively link the production of selenoneine to the *sen* cluster, we reconstituted the entire biosynthetic pathway in vitro. SenA, SenB, SenC and EgtD were cloned from *V. paradoxus* and isolated as soluble, His-tagged proteins through heterologous expression in *Escherichia coli* (Supplementary Tables 3 and 4). As expected, SenC was found to catalyse the ATP-dependent phosphorylation of sodium selenide to yield SeP (Fig. 3a and Supplementary Fig. 2). Next, we evaluated the function of SenB, the C terminus of which shows weak homology to NDP-sugar-binding domains of glycosyltransferases. When SenB was added to the SenC reaction along with the common sugar donor UDP-glucose, production of a new species was observed. It contained two Se atoms, as judged by the characteristic Se isotope pattern obtained by high-resolution MS (HR-MS), implying a potential diselenide dimer (Fig. 3b). Derivatization of the species with the thiol-labelling reagent monobromobimane (mBBr) resulted in an mBBr derivative with a single Se atom and a mass consistent with the monomeric form of the underivatized species (Fig. 3c). Purification of the mBBr derivative and structural elucidation by multidimensional nuclear magnetic resonance (NMR) spectroscopy revealed 1-seleno-β-D-glucose (SeGlc), a rare example of a naturally occurring selenosugar, as the SenB reaction product (Fig. 3d, Supplementary Table 5 and Supplementary Fig. 3).

Control assays further confirmed the requirement of each reaction component (Fig. 3e). In the absence of SenB, no SeGlc was formed. Controls lacking either SenC or ATP also abolished production of SeGlc, verifying SeP as the source of Se for SenB. Moreover, no product was observed with sodium sulfide, suggesting that SenC discriminates between Se and S, thus synthesizing only SeP for the downstream SenB-catalysed reaction (Supplementary Fig. 2). When assayed against various nucleotide-sugar substrates, SenB was found to efficiently utilize UDP-*N*-acetylglucosamine (UDP-GlcNAc) and UDP-*N*-acetylgalactosamine (UDP-GalNAc; Supplementary Table 6 and Supplementary Figs. 4 and 5), but not GDP-mannose or GDP-glucose, suggesting that SenB is specific for UDP-sugars (Fig. 3e). In competition experiments using equimolar concentrations of the three UDP-sugars, the enzyme exclusively accepted UDP-GlcNAc as a substrate forming *N*-acetyl-1-seleno-β-D-glucosamine (SeGlcNAc; Supplementary Fig. 6). Together, these results allow us to designate SenB as a novel selenosugar synthase, which joins SelA and SelU as only the third bona fide Se–C bond-forming enzyme characterized to date. Like SelA and SelU, the reaction of SenB involves Se–P bond cleavage, and the underlying reaction mechanism remains to be determined.

The remaining enzyme encoded in the *sen* cluster, SenA, is a distant homologue of the C–S bond-forming sulfoxide synthase EgtB[16]. Previous work by the groups of Seebeck and Liu has provided extensive characterization of this family of nonhaem iron enzymes, including crystal structures that have pinpointed residues involved in iron, hercynine

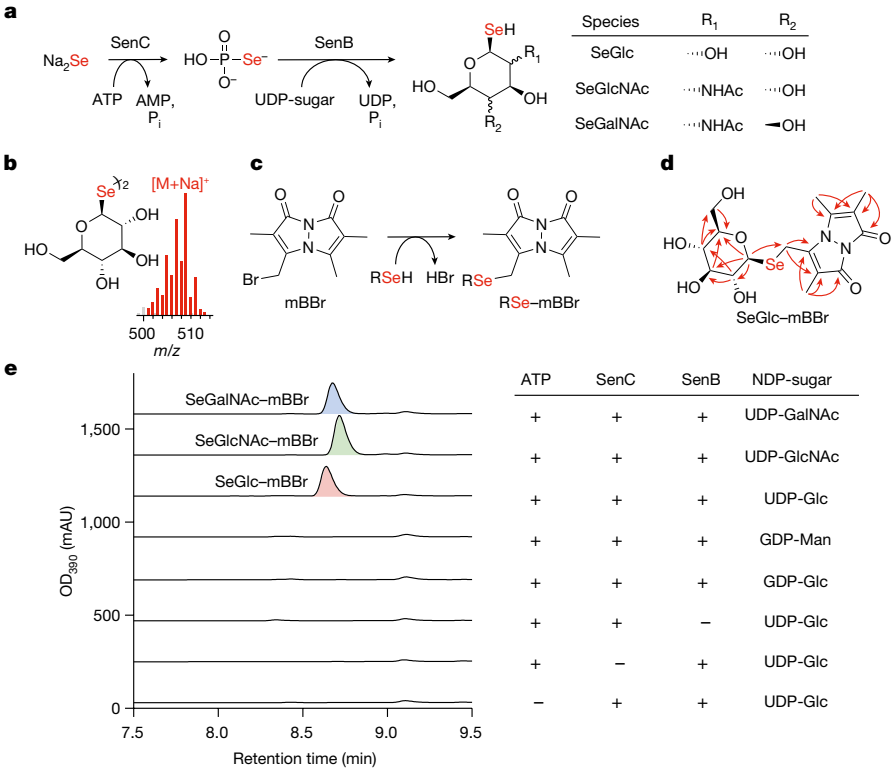

Fig. 3 | SenB, a novel selenosugar synthase. a, Reactions shown to be carried out by SenC (SelD homologue) and SenB, the first known selenosugar synthase, to our knowledge. SenB accepts three different UDP-sugars indicated to generate SeGlc, SeGlcNAc or SeGalNAc. **b**, Diselenide product of the SenB reaction. **c**, The mBBr derivatization reaction. **d**, Structure of the mBBr-derivatized SeGlc; relevant ^{1}H–^{13}C heteronuclear multiple bond correlations (red arrows) used to solve the structure are shown. **e**, Substrate tolerance and control reactions for SenB. The reaction content for each trace is shown in the table on the right of the HPLC–UV chromatogram. The product peaks in the chromatogram are labelled. Results of each reaction are representative of at least three independent assays. Man, mannose; OD_{390}, optical density at 390 nm.

and thiol binding[22–24]. SenA proteins bear 30–50% similarity to members of the EgtB family and share its conserved iron-binding three-His facial triad, catalytic Tyr residue, and motifs involved in hercynine binding (Supplementary Fig. 7). However, many conserved amino acid differences are observed between the two groups, including those implicated in thiol binding within the EgtB active site. These substitutions suggested to us that SenA catalyses Se–C bond formation between hercynine and a different substrate, presumably a selenosugar, en route to selenoneine. We tested this hypothesis by first recapitulating the activity of *V. paradoxus* EgtD in vitro, allowing for rapid formation of hercynine from histidine and SAM (Fig. 4a and Supplementary Fig. 8). Upon incubation of SenA with hercynine (generated in situ with EgtD, His and SAM), various selenosugars (from lyophilized SenBC reaction mixtures), dithiothreitol (DTT) and Fe(II), followed by derivatization with mBBr, we observed a new species containing a single Se atom and a high-resolution mass consistent with that of selenoneine–mBBr (Fig. 4b). The structure was confirmed by NMR spectroscopy upon purification of the product from large-scale reactions with SeGlcNAc (Fig. 4c, Supplementary Table 7 and Supplementary Fig. 9), the substrate for which SenA showed the greatest preference. More importantly, in the absence of mBBr, selenoneine itself was directly observed in the product mixture as evidenced by HR-MS and HR-MS/MS analysis (Supplementary Table 8 and Supplementary Fig. 10), thus completing reconstitution of the entire selenoneine biosynthetic pathway in vitro and confirming that the *sen* cluster is responsible for building this unusual natural product (Fig. 4d).

In ergothioneine biosynthesis, the EgtB reaction yields hercynyl-GGC sulfoxide, which is further processed by EgtC and EgtE to produce ergothioneine[25,26] (Fig. 2d). The analogous product by SenA, hercynyl-SeGlcNAc selenoxide, was not detected in our reaction. However, further examination of the mBBr-derivatized reaction revealed the reduced form of this product, hercynyl-SeGlcNAc selenoether; its structure was corroborated by HR-MS and HR-MS/MS analysis (Fig. 4d, Supplementary Table 8 and Supplementary Fig. 10). We suspect that this product forms as a result of reduction of the selenoxide intermediate by excess DTT in the reaction, implying that the selenoxide intermediate is indeed formed. Spontaneous selenoxide eliminations are well documented, and this intermediate could undergo, aside from reduction by DTT, rapid internal elimination to form the selenenic acid (Fig. 4e), which would convert to selenoneine through either reduction or disproportionation[27]. Selenoxide eliminations occur upwards of 10^5 times more rapidly than the related sulfoxide eliminations[1]. Thus, the tendency of selenoxides to undergo internal elimination obviates the need for an EgtE-like reaction in selenoneine biosynthesis. In support of this proposal, replacement of SeGlcNAc with commercially available thioglucose in the SenA reaction mixture resulted in trace production of hercynyl-thioglucose sulfoxide, with no observed formation of ergothioneine or reduction to the corresponding thioether (Fig. 4f, Supplementary Table 8 and Supplementary Fig. 10).

Further assessment of SenA

SenA catalyses a remarkable Se–C bond-forming reaction and we therefore assessed its features further. To do so, we first prepared SeGlcNAc synthetically (Supplementary Schemes 1–4 and Supplementary Figs. 11 and 12). Assays with SenA, synthetic SeGlcNAc and commercially available hercynine allowed us to recapitulate the results from the multi-enzyme reaction above. Moreover, we were able to locate the eliminated enol ether sugar product of the reaction (2-acetamidoglucal (2AG)), the identity of which we verified by comparison with a synthetic

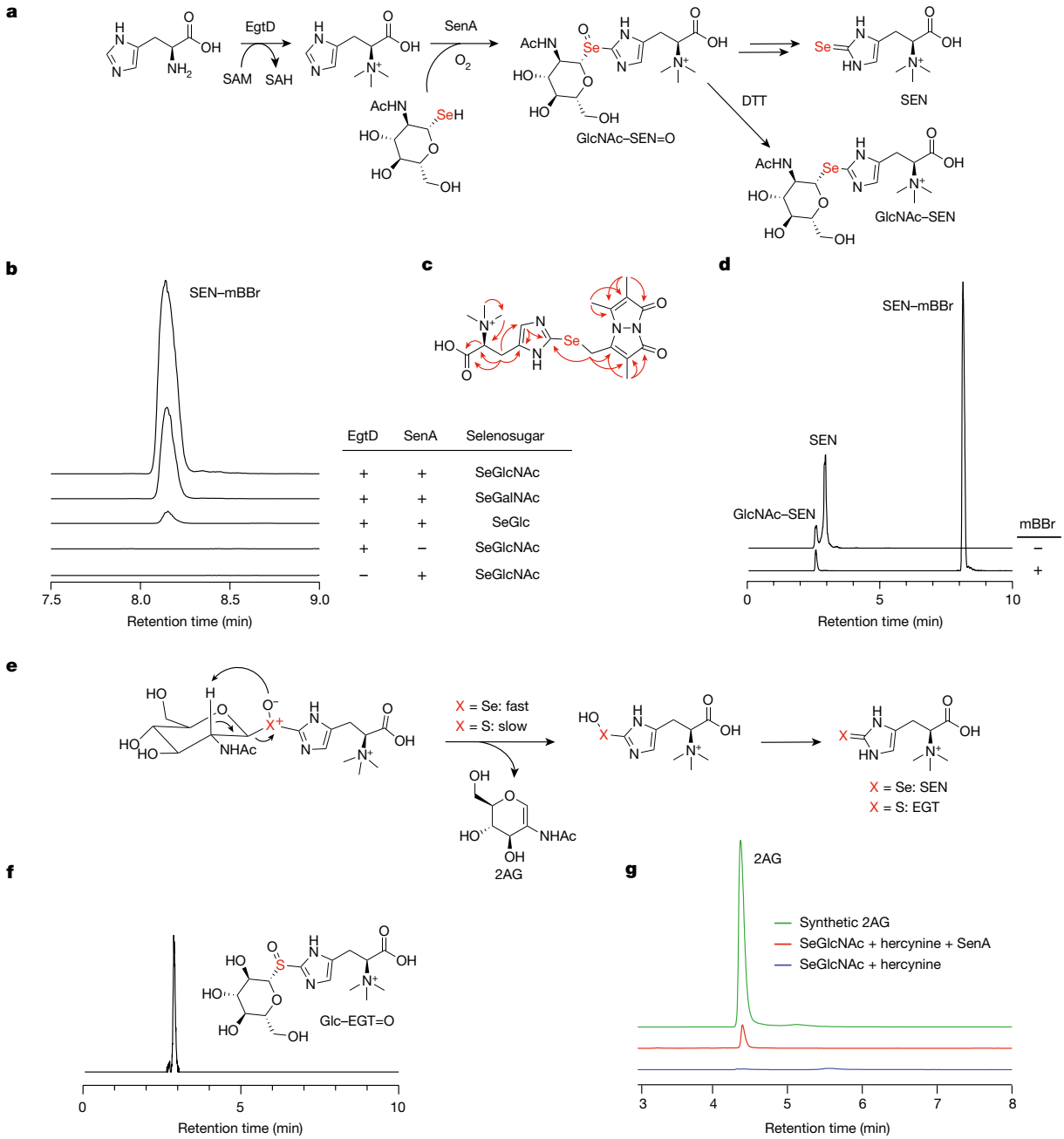

Fig. 4 | SenA, a novel selenoneine synthase. a, Biosynthetic pathway for SEN with the reactions elucidated in this work. Hercyncyl-SeGlcNAc selenoxide (GlcNAc–SEN=O) can spontaneously convert to SEN (see **e**) or be reduced to hercynyl-SeGlcNAc selenoether (GlcNAc–SEN). **b**, Reaction of SenA monitored by HPLC–MS. EICs of SEN–mBBr (*m/z* 468.1) are displayed. The reaction content for each trace in the chromatogram is shown in the table on the right. Among sugars tested, SenA shows preference for SeGlcNAc. **c**, Structure of SEN–mBBr. Relevant ^{1}H–^{13}C HMBC correlations are shown (red arrows). **d**, SEN can be detected directly by HR-MS in the absence of mBBr derivatization. The top trace displays a combined EIC of GlcNAc–SEN (*m/z* 481.1) and SEN. The bottom trace displays a combined EIC of GlcNAc–SEN and SEN–mBBr. **e**, Proposed internal elimination to form SEN. Selenoxide eliminations are considerably faster than related sulfoxide eliminations. **f**, HPLC–MS detection of trace hercynyl-thioglucose sulfoxide (Glc–EGT=O; *m/z* 408.1) when the SenA reaction is carried out with a thiosugar donor. **g**, HPLC–MS detection of 2AG in the SenA product mixture, verified by comparison of retention time with a synthetic standard. EICs of 2AG (*m/z* 226.1, Na^+ adduct) are displayed. No 2AG was observed in the absence of SenA. Results of each reaction are representative of at least three independent assays.

standard (Fig. 4g and Supplementary Fig. 13). These studies confirm the unprecedented reaction catalysed by SenA.

With the substrate in hand, we examined the requirements for the reaction further. Assays in an inert atmosphere in the presence of SeGlcNAc yielded only trace product, whereas rapid conversion was observed under aerobic, but otherwise identical, conditions (Supplementary Fig. 14). No product was observed when the assay was carried out in the presence of EDTA. Furthermore, a triple His-to-Ala mutant (H71A;H167A;H171A), targeting the metal-binding facial triad conserved across all SenA and EgtB proteins, was completely inactive. Supplementation of EDTA-treated wild-type SenA with Zn, Ni, Mn and Fe only yielded appreciable product with the latter (Supplementary Fig. 15). We therefore conclude that SenA is an Fe-dependent enzyme that requires O_2 for activity.

We also probed the requirement for DTT and found that the reductant substantially enhances product yields (Supplementary Fig. 16). Aside from generating reduced Fe enzyme for reaction with O_2, we suspect that DTT also maintains reduced levels of SeGlcNAc, which dimerizes easily under aerobic conditions. Comparison of reactions in the presence and absence of DTT allowed us to locate GlcNAc–SEN only in the former, thus lending support to our model in Fig. 4a, in which GlcNAc–SEN=O either spontaneously converts to SEN via selenoxide elimination or is reduced to give GlcNAc–SEN. Finally, we examined the substrate preferences of SenA and *V. paradoxus* EgtB by reacting each with hercynine and either SeGlcNAc, thioglucose, Cys, Sec or GGC. SenA only gave appreciable product with SeGlcNAc, whereas EgtB only accepted Cys as a substrate (Supplementary Fig. 17). These studies further corroborate that the SEN pathway is specific for Se, whereas the bacterial EGT pathway is specific for sulfur.

Conclusions

Our results open new avenues in selenobiology, notably the incorporation of Se into other natural products and secondary metabolites. These molecules typically comprise a limited set of elements, usually consisting of C, N, O, H, P and S. Demonstration of a widespread pathway for Se incorporation into natural products provides impetus for interrogating other selenometabolites and their biosynthetic origins that are likely to be encoded in microbial genomes. The reactions of SenA (selenoneine synthase) and SenB (selenosugar synthase) reported here double the number of previously known Se–C bond-forming enzymes in nature and at the same time stimulate future studies into their reaction mechanisms. Seminal studies by Stadtman and Böck previously demonstrated pathways for incorporation of Se into protein and nucleic acid biopolymers[9,11–13]. SenA, SenB and SenC provide the first pathway for biosynthesis of selenosugars, which could have yet unknown roles in primary and secondary metabolism. More generally, selenometabolomes provide an untapped chemical space in bacteria that can now be mined for novel natural products and their associated bioactivities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05174-2.

1. Reich, H. J. & Hondal, R. J. Why nature chose selenium. *ACS Chem. Biol.* **11**, 821–841 (2016).
2. Yamashita, Y. & Yamashita, M. Identification of a novel selenium-containing compound, selenoneine, as the predominant chemical form of organic selenium in the blood of bluefin tuna. *J. Biol. Chem.* **285**, 18134–18138 (2010).
3. Cheah, I. K. & Halliwell, B. Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim. Biophys. Acta* **1822**, 784–793 (2012).
4. Rotruck, J. T. et al. Selenium: biochemical role as a component of glutathione peroxidase. *Science* **179**, 588–590 (1973).
5. Zhong, L. & Holmgren, A. Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations. *J. Biol. Chem.* **275**, 18121–18128 (2000).
6. Aachmann, F. L. et al. Insights into function, catalytic mechanism, and fold evolution of selenoprotein methionine sulfoxide reductase B1 through structural analysis. *J. Biol. Chem.* **285**, 33315–33323 (2010).
7. Berry, M. J., Kieffer, J. D., Harney, J. W. & Larsen, P. R. Selenocysteine confers the biochemical properties characteristic of the type I iodothyronine deiodinase. *J. Biol. Chem.* **266**, 14155–14158 (1991).
8. He, S. H. et al. EPR studies with ^{77}Se-enriched (NiFeSe) hydrogenase of *Desulfovibrio baculatus*. *J. Biol. Chem.* **264**, 2678–2682 (1989).
9. Wittwer, A. J., Tsai, L., Ching, W. M. & Stadtman, T. C. Identification and synthesis of a naturally occurring selenonucleoside in bacterial tRNAs: 5-[(methylamino)methyl]-2-selenouridine. *Biochemistry* **23**, 4650–4655 (1984).
10. Weekley, C. M. & Harris, H. H. Which form is that? The importance of selenium speciation and metabolism in the prevention and treatment of disease. *Chem. Soc. Rev.* **42**, 8870–8894 (2013).
11. Ehrenreich, A., Forchammer, K., Tormay, P., Veprek, B. & Böck, A. Selenoprotein synthesis in *E. coli*. Purification and characterisation of the enzyme catalysing selenium activation. *Eur. J. Biochem.* **206**, 767–773 (1992).
12. Forchhammer, K. & Böck, A. Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J. Biol. Chem.* **266**, 6324–6328 (1991).
13. Wolfe, M. D. et al. Functional diversity of the rhodanese homology domain. *J. Biol. Chem.* **279**, 1801–1809 (2004).
14. Stone, M. J. & Williams, D. H. On the evolution of functional secondary metabolites (natural products). *Mol. Microbiol.* **6**, 29–34 (1992).
15. Lin, J. et al. Comparative genomics reveals new candidate genes involved in selenium metabolism in prokaryotes. *Genome Biol. Evol.* **7**, 664–676 (2015).
16. Seebeck, F. P. In vitro reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **132**, 6632–6633 (2010).
17. Yamashita, Y. et al. Selenoneine, total selenium, and total mercury content in the muscle of fishes. *Fish. Sci.* **77**, 679–686 (2011).
18. Yamashita, M. & Yamashita, Y. in *Springer Handbook of Marine Biotechnology* (ed. Kim, S. K.) 1059–1069 (Springer, 2015).
19. Lim, D., Gründemann, D. & Seebeck, F. P. Total synthesis and functional characterization of selenoneine. *Angew. Chem. Int. Ed.* **58**, 15026–15030 (2019).
20. Pluskal, T., Ueno, M. & Yanagida, M. Genetic and metabolomic dissection of the ergothioneine and selenoneine biosynthetic pathway in the fission yeast, *S. pombe*, and construction of an overproduction system. *PLoS ONE* **9**, e97774 (2014).
21. Goncharenko, K. V. et al. Selenocysteine as a substrate, an inhibitor and a mechanistic probe for bacterial and fungal iron-dependent sulfoxide synthases. *Chem. Eur.* **26**, 1328–1334 (2020).
22. Goncharenko, K. V., Vit, A., Blankenfeldt, W. & Seebeck, F. P. Structure of the sulfoxide synthase EgtB from the ergothioneine biosynthetic pathway. *Angew. Chem. Int. Ed.* **54**, 2821–2824 (2015).
23. Naowarojna, N. et al. Crystal structure of the ergothioneine sulfoxide synthase from *Candidatus Chloracidobacterium thermophilum* and structure-guided engineering to modulate its substrate selectivity. *ACS Catal.* **9**, 6955–6961 (2019).
24. Stampfli, A. R. et al. An alternative active site architecture for O_2 activation in the ergothioneine biosynthetic EgtB from *Chloracidobacterium thermophilum*. *J. Am. Chem. Soc.* **141**, 5275–5285 (2019).
25. Vit, A., Mashabela, G. T., Blankenfeldt, W. & Seebeck, F. P. Structure of the ergothioneine-biosynthesis amidohydrolase EgtC. *ChemBioChem* **16**, 1490–1496 (2015).
26. Song, H. et al. Mechanistic studies of a novel C-S lyase in ergothioneine biosynthesis: the involvement of a sulfenic acid intermediate. *Sci Rep.* **5**, 11870 (2015).
27. Reich, H. J., Renga, J. M. & Reich, I. L. Organoselenium chemistry. Conversion of ketones to enones by selenoxide syn elimination. *J. Am. Chem. Soc.* **97**, 5434–5447 (1975).

Methods

Materials

All materials were purchased from Millipore-Sigma or Fisher Scientific unless otherwise specified. Cloning reagents were purchased from New England BioLabs (NEB). Codon-optimized gene fragments were purchased from Genewiz. DNA primers were purchased from Sigma.

General experimental procedures

HPLC–MS was performed on an Agilent instrument equipped with a 1260 Infinity Series HPLC, an automated liquid sampler, a photodiode array detector, a JetStream ESI source and the 6540 Series Q-tof mass spectrometer. HPLC–MS data were acquired and analysed with Agilent MassHunter software. HPLC purifications were carried on an Agilent 1260 Infinity Series HPLC system equipped with a photodiode array detector and an automated fraction collector. Solvents for all HPLC–MS and HPLC experiments were water and 0.1% formic acid (solvent A), and MeCN and 0.1% formic acid (Solvent B). HPLC data were acquired and analysed with Agilent OpenLab software. NMR spectra were collected at the Princeton University Department of Chemistry's NMR facility on a Bruker Avance III 500 MHz NMR spectrometer equipped with a DCH double resonance cryoprobe. NMR data were acquired using Bruker Topspin software and processed and/or analysed using MestReNova software. Anaerobic enzyme assays were carried out in an MBraun glovebox under a N_2 atmosphere maintained at <0.1 ppm O_2.

Bioinformatics

All SelD protein sequences were retrieved from the NCBI database using a combination of bioinformatic tools. First, the NCBI E-Direct toolkit was used to extract all proteins from the conserved domain database (CDD) with a match to the SelD position-specific scoring matrix 'COG0709'. This resulted in 205,276 retrieved sequences. Next, the CD-hit clustering tool was used to cluster the sequences at 98% sequence identity to dereplicate and reduce redundancy in the dataset resulting from over-sequencing bias[28]. This reduced the dataset to 31,280 sequences, illustrating the amount of sequence redundancy. These were then aligned against a short list of representative bacterial SelD sequences using the blastp algorithm with an E-value cut-off of 1×10^{-20} to remove any non-SelD protein families, such as thiamine-phosphate kinase, which has slight homology to SelD[29,30]. This resulted in a final, strictly dereplicated set of 10,947 SelD protein sequences from the NCBI database.

Next, the E-Direct toolkit was used to extract all annotated open reading frames (ORFs) within a window 4 kb upstream or downstream of each *selD* gene. From this dataset, a total of 2,960 ORFs with a minimum of 1 bp overlap with a *selD* ORF were obtained and re-annotated using a combination of the NCBI CDD-search tool and manual examination. The protein families corresponding to the ten most common *selD*-overlapping genes are shown in Supplementary Table 1.

To further examine the *tigr04348* gene, first a blastp search was performed against the NCBI non-redundant protein database using a short list of TIGR04348 protein sequences from various phyla and an E-value cut-off of 1×10^{-20}. This resulted in a list of 2,353 TIGR04348 sequences after dereplication by strain name. Next, the E-Direct toolkit was used to extract 5 kb upstream and downstream of each *tigr04348* gene. These genetic regions were then used to construct a BLAST database and interrogated using the tblastn algorithm for the presence of *egtB* and *selD* homologues, resulting in a total of 827 *egtB–tigr04348–selD* BGCs, henceforth referred to as the *sen* BGC consisting of sen*A*, *senB* and *senC*. A select few examples of *sen* BGCs are depicted in Supplementary Fig. 1, noting particularly the presence of colocalized *egtD* and a putative selenate transporter gene in a subset of them.

For amino acid multiple sequence alignments of EgtB and SenA, the top 500 SenA sequences most similar to SenA from *V. paradoxus* and the top 4,000 EgtB sequences most similar to EgtB from *Mycobacterium thermoresistibile* were retrieved using BLAST and aligned separately using MAFFT with the FFT-NS-2 algorithm[31]. These alignments were then used to generate logo plots of the important motifs, as determined by Seebeck et al.[22] for *M. thermoresistibile*, using the Logomaker Python script[32] (Supplementary Fig. 7).

Strains, media and general culture conditions

Strains used in selenometabolite characterization and recombinant protein production experiments are listed in Supplementary Table 2. ISP medium 2 agar plates and tryptic soy broth were used for general maintenance and liquid cultures of *A. palatopharyngis* DSM 44832 and *S. rimosus* ATCC 10970. Nutrient agar plates and nutrient broth were used for general maintenance and liquid cultures of *V. paradoxus* DSM 30034. LB agar and broth (supplemented with 50 mM MOPS, pH 7.0) were used for general maintenance and liquid cultures of *B. thailandensis* E264.

Selenometabolite production screens

For each strain, a single colony from an agar plate was inoculated into a sterile culture tube containing 5 ml of liquid medium and incubated at 30 °C/250 r.p.m. The starter cultures were then used to inoculate 5 ml liquid cultures supplemented with 100 μM of filter-sterilized Na_2SeO_3 and incubated at 30 °C/250 r.p.m. Production cultures of *B. thailandensis* were grown for 24 h, *V. paradoxus* and *S. rimosus* for 48 h, and *A. palatopharyngis* for 72 h. Following incubation, 2 ml of each culture was transferred to an Eppendorf tube, pelleted by centrifugation and supernatants were removed. Cell pellets were resuspended in 0.4 ml of ddH_2O and subjected to hot-water extraction by heating to 95 °C for 15 min. After cooling to room temperature, the mixtures were pelleted by centrifugation and supernatants were analysed by HPLC–MS. Analytes were separated on a Synergi Fusion-RP column (Phenomenex; 100 x 4.6 mm, 4 μm) with a flow rate of 0.5 ml min^{-1} and an elution programme consisting of 0% solvent B wash for 5 min, a linear gradient from 0% to 100% B over 10 min, followed by a hold at 100% B for 3 min.

Construction of protein expression plasmids

Genomic DNA from *V. paradoxus* DSM 30034 was isolated using the Wizard Genomic DNA Purification Kit (Promega) following the manufacturer's instructions. From genomic DNA, *senC*, *egtD* and *egtB* genes were PCR-amplified using Q5 High Fidelity DNA polymerase (NEB) with primers Vpa-SenC-F/R, Vpa-EgtD-F/R and Vpa-EgtB-F/R, respectively, which have overhangs that allowed for assembly into pET28b(+) (Supplementary Table 3). *senA*, *senA-H71A;H167A;H171A* and *senB* from *V. paradoxus* DSM 30034 were obtained as synthetic DNA fragments, codon-optimized for expression in *E. coli* and with overhangs to allow assembly into pET28b(+). Protein expression plasmids were assembled from gene fragments and vector pET28b(+), linearized with NdeI and XhoI (NEB), using HiFi DNA Assembly Master Mix (NEB) following the manufacturer's instructions. Ligation mixtures were transformed into chemically competent *E. coli* DH5α by heat shock and plated onto LB agar containing 50 mg l^{-1} kanamycin. After confirmation by Sanger sequencing, assembled plasmids were transformed into *E. coli* BL21(DE3) for protein expression.

Expression and purification of 6×His-tagged SenA, SenB, SenC, EgtB and EgtD proteins

All six proteins were produced separately in *E. coli* BL21(DE3) cells grown in two 4-l flasks, each containing 2 l Terrific Broth supplemented with 50 mg l^{-1} kanamycin at 37 °C/170 r.p.m. Small cultures were prepared by inoculating 40 ml of LB medium containing 50 mg l^{-1} kanamycin with a single colony of *E. coli* BL21(DE3) carrying the desired plasmid. After overnight growth at 37 °C/170 r.p.m., 4 l of TB medium plus 50 mg l^{-1} kanamycin were inoculated with the 40 ml small culture and incubated at 37 °C/170 rpm. At OD_{600} = 0.5–0.6, protein expression was induced with 0.2 mM IPTG, and cultures were incubated at 37 °C/170 r.p.m. for an additional 12–24 h. Cells were pelleted by centrifugation (8,000*g*

for 15 min at 4 °C), yielding approximately 7 g of cell paste per litre. The cell pastes were stored at −80 °C until purification.

All purification steps were carried out in a cold room at 4 °C. Cells were resuspended in lysis buffer (5 ml per gram cell paste), which consisted of 25 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.7, supplemented with 1 µl ml^{-1} protease inhibitor cocktail (Sigma) and 1 mM phenylmethylsulfonyl fluoride. Once homogenous, 0.1 mg ml^{-1} deoxyribonuclease I (Alfa Aesar) was added, and the cells were lysed by the addition of 5 mg ml^{-1} lysozyme followed by sonication using 30% power (approximately 150 W) in 15 s on–15 s off cycles for a total of 4 min. This process was repeated twice. The lysate was then clarified by centrifugation (17,000*g* for 15 min at 4 °C) and loaded onto a 5-ml Ni-NTA column pre-equilibrated in lysis buffer. The column was washed with lysis buffer, and His-tagged proteins were eluted with elution buffer consisting of 25 mM Tris-HCl, 300 mM NaCl, 300 mM imidazole and 10% glycerol, pH 7.7. Eluted proteins were then buffer-exchanged using a 50-ml column of Sephadex G-25 (Cytiva) into storage buffer consisting of 25 mM Tris-HCl, 150 mM NaCl and 10% glycerol, pH 7.7. Purified proteins were stored at −80 °C. Protein concentrations were determined spectrophotometrically on a Cary 60 UV-visible spectrophotometer (Agilent) using calculated molar extinction coefficients at 280 nm. From 4 l cultures, the following yields were obtained: 105 mg SenA, 198 mg SenB, 152 mg SenC, 63 mg EgtD, 43 mg SenA-H71A;H167A;H171A, and 63 mg EgtB.

In vitro reconstitution of SenC activity

The selenophosphate synthetase activity of SenC was characterized according to previous methods[33–35]. Under anaerobic conditions, a 700-µl reaction containing 2 mM DTT, 1 mM ATP, 1.5 mM Na_2Se and 20 µM SenC was prepared in buffer consisting of 50 mM tricine, 20 mM KCl, 5 mM $MgCl_2$ and 10% D_2O, pH 7.2. Control reactions were prepared in an identical manner, lacking either SenC or Na_2Se, or with Na_2S in place of Na_2Se. After a 1-h incubation period at room temperature, the reactions were transferred to NMR tubes, removed from the glovebox and immediately analysed by ^{31}P-NMR. Results were consistent with previous reports of selenophosphate synthetase enzymes, with conversion to SeP only observed in the presence of all components. No production of thiophosphate was observed in the presence of Na_2S (Supplementary Fig. 2).

In vitro reconstitution of SenB activity

Under anaerobic conditions, 200-µl reactions containing 2 mM DTT, 2 mM ATP, 1 mM Na_2Se, 20 µM SenC, 20 µM SenB and 2 mM of NDP-sugar were prepared in buffer consisting of 50 mM tricine, 20 mM KCl and 5 mM $MgCl_2$, pH 7.2. Control reactions were prepared in an identical manner, lacking either SenB, SenC or ATP. After a 6-h incubation period at room temperature, reactions were removed from the glovebox and exposed to oxygen for 30 min to oxidize any unreacted Na_2Se. Next, 50 µl of each reaction mixture was quenched with 50 µl of MeOH, whereas another 50 µl was quenched with 50 µl of 10 mM mBBr in MeCN. Reactions were incubated for an additional 30 min at room temperature in the dark to allow for complete derivatization with mBBr. Reactions quenched with MeOH were filtered and analysed by HPLC–MS using a Kinetex Polar C18 column (Phenomenex, 150 × 4.6 mm, 2.6 µm) with a flow rate of 0.4 ml min^{-1} and an elution programme consisting of 0–20% solvent B over 5 min, followed by 20–100% solvent B over 3 min, and a final step of 3 min at 100%. This assay demonstrated the production of underivatized selenosugar diselenides (Supplementary Fig. 4). Reactions quenched with mBBr were filtered and analysed by HPLC–MS using a Synergi Hydro-RP column (Phenomenex, 250 × 4.6 mm, 4 µm) with a flow rate of 1 ml min^{-1} and an elution programme consisting of a 5% solvent B wash for 3 min, a gradient of 5–75% solvent B over 6 min, followed by a gradient of 75–100% solvent B over 1 min, and a final hold at 100% for 5 min. The remaining 100 µl of each reaction mixture was lyophilized for use as a crude selenosugar substrate in downstream assays with SenA.

Structural characterization of SenB products

SenB enzymatic assays were carried out anaerobically as described above on a 10-ml scale with UDP-Glc and UDP-GlcNAc. Reactions were quenched with 3 ml of 10 mM mBBr in MeCN and incubated on a platform rocker for 1 h in the dark to facilitate complete derivatization. SeGlc–mBBr was purified by semi-preparative HPLC using a Synergi Fusion-RP column (Phenomenex, 250 × 10 mm, 4 µm) with a flow rate of 2.5 ml min^{-1} and elution consisting of a gradient of 0–40% solvent B for 15 min, followed by 40–100% solvent B for 3 min and a final hold at 100% B for 5 min. SeGlcNAc–mBBr was purified by semi-preparative HPLC using a Luna C18 column (Phenomenex, 250 × 10 mm, 5 µm) with a flow rate of 2.5 ml min^{-1} and elution consisting of a gradient of 10–50% solvent B for 10 min, followed by a gradient of 50–100% solvent B for 10 min, followed by a hold at 100% for 3 min. Purified compounds were dissolved in DMSO-d_6 and analysed by NMR spectroscopy, confirming the presence of a selenium atom bound to the anomeric carbon, which was determined to be β-configured for both selenosugars, as shown by the large axial–axial coupling constants between protons on the sugar C-1 and C-2 (9.2 Hz for SeGlc–mBBr and 10.5 Hz for SeGlcNAc–mBBr). Full spectra, chemical shift assignments and select 2D correlations can be found in Supplementary Figs. 3 and 5 and Supplementary Tables 5 and 6.

SenB UDP-sugar competition assays to deduce substrate preference

SenB reactions were carried out as described above with the following modifications. Single-substrate reactions containing either UDP-Glc, UDP-GalNAc or UDP-GlcNAc were incubated for 2 h, followed by quenching with mBBr. The competition assay, containing a mixture of 1 mM UDP-Glc, 1 mM UDP-GalNAc and 1 mM UDP-GlcNAc, was quenched with mBBr after 5, 10, 30, 60 and 120 min. All quenched reactions were analysed by HPLC–MS as described above.

In vitro reconstitution of EgtD activity

Under aerobic conditions, 100-µl reactions containing 0.5 mM histidine, 2 mM SAM and 20 µM EgtD were prepared in buffer consisting of 50 mM Tris-HCl and 150 mM NaCl, pH 8.0. A control reaction lacking EgtD was prepared in an identical manner. After a 30-min incubation period at room temperature, reactions were quenched with 100 µl MeOH, filtered and analysed by HPLC–MS using a Synergi Hydro-RP column (Phenomenex, 250 × 4.6 mm, 4 µm) with a flow rate of 1 ml min^{-1} and an elution programme consisting of a 0% solvent B wash for 3 min, followed by a gradient to 100% solvent B over 2 min and a hold at 100% for 5 min. Results were consistent with previous reports of EgtD enzymes, with conversion of histidine to hercynine observed only in the presence of EgtD[36] (Supplementary Fig. 8).

In vitro reconstitution of SenA activity

Under aerobic conditions, crude selenosugar substrates from lyophilized SenBC assay mixtures were first dissolved in buffer consisting of 50 mM Tris-HCl and 150 mM NaCl, pH 8. Next, 0.5 mM histidine, 2 mM SAM, 2 mM DTT, 0.2 mM $(NH_4)_2Fe(SO_4)_2$, 20 µM EgtD and 20 µM SenA were added to a final volume of 100 µl. Control assays were prepared in an identical manner, except lacking either EgtD or SenA. Assays containing sodium β-D-thioglucose (Fisher) in place of selenosugar substrate were also prepared in a similar manner. However, thioglucose was found to be a very poor substrate for SenA, and thus 80 µM enzyme was used to facilitate appreciable conversion. After a 7-h incubation period at room temperature, 50 µl of each reaction mixture was quenched with 50 µl of MeOH, whereas another 50 µl was quenched with 50 µl of 10 mM mBBr in MeCN. Reactions were incubated for an additional 30 min at room temperature in the dark to allow for complete derivatization with mBBr. The samples were then filtered and analysed by HPLC–MS using a Synergi Hydro-RP column (Phenomenex, 250 × 4.6 mm, 4 µm) with a flow rate of 1 ml min^{-1} and elution consisting of 5% solvent B for

3 min, a gradient of 5–75% solvent B over 6 min, followed by a gradient of 75–100% over 1 min and a hold at 100% for 5 min.

Structural characterization of SenA products

A 10-ml SenA enzymatic assay was carried out aerobically as described above using a crude SeGlcNAc substrate from a lyophilized, 10-ml SenBC reaction mixture. Reactions were quenched with 5 ml of 10 mM mBBr in MeCN and incubated on a platform rocker for 1 h in the dark to facilitate complete derivatization. SEN–mBBr was purified by semi-preparative HPLC using a Luna C18 column (Phenomenex; 250 × 10 mm, 5 µm) with a flow rate of 2.5 ml min^{-1} and an elution programme consisting of a gradient of 7.5–80% solvent B for 12 min, followed by a hold at 80% for 4 min. Purified SEN–mBBr was dissolved in D_2O and analysed by NMR spectroscopy. The Se atom was confirmed to be positioned at the imidazole C2 carbon, as evidenced by the absence of an imidazole C-2 proton and diagnostic ^{1}H–^{13}C HMBC correlations. Full spectra, chemical shift assignments and select 2D correlations can be found in Supplementary Fig. 9 and Supplementary Table 7. The structures of other SenA products (hercynyl-SeGlcNAc selenoether (GlcNAc–SEN), hercynyl-SeGlc selenoether (Glc–SEN), hercynyl-thioglucose sulfoxide (Glc-EGT=O) and SEN) were confirmed by tandem HR-MS/MS analysis (Supplementary Fig. 10 and Supplementary Table 8).

Chemical synthesis of SeGlcNAc and 2AG

Syntheses of SeGlcNAc and 2AG were carried out as previously described[37,38]. Synthetic schemes and NMR characterization are detailed further in the Supplementary Information.

SenA oxygen-dependence experiments

Under anaerobic conditions, SenA was first exchanged into degassed buffer consisting of 50 mM Tris-HCl and 150 mM NaCl, pH 8 using a 30-kDa MWCO centrifugal filter. Next, a 100-µl reaction was prepared in the same buffer containing 1 mM DTT, 1 mM hercynine (Sigma), 0.2 mM $(NH_4)_2Fe(SO_4)_2$, 1 mM synthetic SeGlcNAc (monomer basis) and 20 µM SenA. The reaction mixture was immediate split into two 50-µl portions. The first portion was diluted with an equal volume of anaerobic buffer. The second portion was removed from the anaerobic chamber and diluted with an equal volume of aerobic buffer. Both reactions were incubated at room temperature for 1 h, followed by mBBr derivatization and analysis by HPLC–MS as described above.

SenA metal and reductant dependence experiments

To assess the metal requirement, a 1.25 ml aliquot of 333 µM SenA was incubated with 20 mM EDTA on ice for 30 min. EDTA was then removed by gel filtration using a 50-ml column of Sephadex G-25, and the protein was concentrated using a 30-kDa MWCO centrifugal filter. Using this sample, 100-µL reactions in 50 mM Tris-HCl and 150 mM NaCl, pH 8 were prepared containing 1 mM DTT, 1 mM hercynine, 1 mM synthetic SeGlcNAc (monomer basis), 20 µM SenA, and 0.2 mM of either $(NH_4)_2Fe(SO_4)_2$, $MnCl_2$, $ZnCl_2$, $NiCl_2$ or 1 mM EDTA. Control reactions with as-purified SenA and as-purified SenA-H71A;H167A;H171A (mutated metal-binding motif) were prepared in an identical manner, except without addition of metal.

To assess the reductant requirement, reactions in 50 mM Tris-HCl and 150 mM NaCl, pH 8, with as-purified SenA were prepared containing 1 mM hercynine, 1 mM synthetic SeGlcNAc (monomer basis), 20 µM SenA, and either 1 mM or 0 mM DTT. All reactions were incubated at room temperature for 30 min, followed by mBBr derivatization and analysis by HPLC–MS as described above.

SenA and EgtB selenol or thiol substrate preference experiments

Substrate preferences were determined by incubating SenA and EgtB separately with hercynine and an excess of either SeGlcNAc, thioglucose, Sec, Cys or GGC. Under aerobic conditions, 100-µl reactions in 50 mM Tris-HCl and 150 mM NaCl, pH 8 were prepared containing 1 mM DTT, 0.4 mM hercynine, 0.2 mM $(NH_4)_2Fe(SO_4)_2$, 2 mM selenol or thiol, and 20 µM SenA or EgtB. Reactions were incubated for 3 h, quenched with an equal volume of MeOH, and analysed by HPLC–MS using a Synergi Hydro-RP column (Phenomenex, 250 × 4.6 mm, 4 µm) with a flow rate of 1 ml min^{-1} and an isocratic elution at 2% solvent B for 3 min. Conversion was visualized by the extent of hercynine consumption.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Experimental data supporting the conclusions of this study are available within the article and its Supplementary information. Sequences were retrieved from the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/) and the NCBI Non-redundant Protein Database (https://www.ncbi.nlm.nih.gov/protein/). NCBI accession numbers of analysed proteins from *V. paradoxus* DSM 30034 are as follows: WP_062361878.1 (SenA), WP_080642484.1 (SenB), WP_062361881.1 (SenC), WP_062366250.1 (EgtD) and WP_062366249.1 (EgtB). Raw experimental data and complete bioinformatic datasets can be made available on reasonable request.

28. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**, 3150–3152 (2012).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
30. McCulloch, K. M., Kinsland, C., Begley, T. P. & Ealick, S. E. Structural studies of thiamin monophosphate kinase in complex with substrates and products. *Biochemistry* **47**, 3810–3821 (2008).
31. Katoh, K. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
32. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in python. *Bioinformatics* **36**, 2272–2274 (2019).
33. Veres, Z., Kim, I. Y., Scholz, T. D. & Stadtman, T. C. Selenophosphate synthetase. Enzyme properties and catalytic reaction. *J. Biol. Chem.* **269**, 10597–10603 (1994).
34. Kim, I. Y., Veres, Z. & Stadtman, T. C. *Escherichia coli* mutant SELD enzymes. The cysteine 17 residue is essential for selenophosphate formation from ATP and selenide. *J. Biol. Chem.* **267**, 19650–19654 (1992).
35. Glass, R. S. et al. Monoselenophosphate: synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. *Biochemistry* **32**, 12555–12559 (1993).
36. Vit, A., Misson, L., Blankenfeldt, W. & Seebeck, F. P. Ergothioneine biosynthetic methyltransferase EgtD reveals the structural basis of aromatic amino acid betaine biosynthesis. *ChemBioChem* **16**, 119–125 (2014).
37. Kumar, A. A., Illyes, T. Z., Kover, K. E. & Szilagyi, L. Convenient syntheses of 1,2-trans selenoglycosides using isoselenuronium salts as glycosylselenenyl transfer reagents. *Carbohydrate Res.* **360**, 8–18 (2012).
38. Pravdic, N. & Fletcher, H. G. The behavior of 2-acetamido-2-deoxy-D-mannose with isopropenyl acetate in the presence of p-toluenesulfonic acid. I. Isolation and identification of derivatives of 2-amino-D-glucal (2-amino-1,2-dideoxy-D-arabino-hex-1-enopyranose) and of other products. *J. Org. Chem.* **32**, 1806–1810 (1967).

Acknowledgements We thank Andy K. L. Nguy for helpful discussions and the Edward C. Taylor 3rd Year Fellowship in Chemistry (to C.M.K.), the Life Sciences Research Foundation Postdoctoral Fellowship sponsored by the Open Philanthropy Project (to J.H.), the Swiss National Science Foundation Early "Postdoc Mobility" Fellowship (no. P2EZP2_187995 to N.H.), the National Science Foundation CAREER Award (no. 1847932 to M.R.S.), and the US National Institutes of Health (GM129496 to M.R.S.) for financial support.

Author contributions C.M.K. and M.R.S. conceived the idea for the study. C.M.K. designed and performed the bioinformatic search and all experiments described in the paper. J.H. synthesized and purified SeGlcNAc and 2AG. N.H. conceived the selenoxide elimination. C.M.K. and M.R.S. analysed the data and prepared the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05174-2.

Correspondence and requests for materials should be addressed to Mohammad R. Seyedsayamdost.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Mohammad R. Seyedsayamdost
Last updated by author(s): Jul 20, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Agilent MassHunter Workstation Data Acquisition Agilent OpenLab Bruker Biospin Topspin
Data analysis	Agilent MassHunter Workstation Qualitative Analysis 10.0 MestReNova x64 14.1 NCBI blast+ suite version 2.7.1 NCBI EDirect toolkit NCBI CDD search web tool CD-hit version 4.8.1 MAFFT version 6.717b Logomaker version 0.8

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Experimental data supporting the conclusions of this study are available within the article and its Supplementary information. Sequences were retrieved from the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/) and the NCBI Non-redundant Protein Database (https://www.ncbi.nlm.nih.gov/protein/). NCBI accession numbers of analyzed proteins from V. paradoxus DSM 30034 are as follows: WP_062361878.1 (SenA), WP_080642484.1 (SenB), WP_062361881.1 (SenC), WP_062366250.1 (EgtD), WP_062366249.1 (EgtB). Raw experimental data and complete bioinformatic datasets can be made available upon reasonable request.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical tests were relevant to our analyses, and therefore no sample size calculations were performed. All assays were performed a minimum of three independent times to ensure reproducibility, concordant with other publications in this field.
Data exclusions	No data were excluded from any analyses.
Replication	Metabolite production screens consisted of at least three biological replicates and enzyme assays were repeated at least three independent times. All results were reliably reproduced.
Randomization	No experimental groups were used in this study, and randomization was therefore not relevant.
Blinding	Group allocation was not relevant to the experiments in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging