nature chemistry

Article

https://doi.org/10.1038/s41557-024-01600-2

Discovery of the selenium-containing antioxidant ovoselenol derived from convergent evolution

Received: 19 February 2024

Accepted: 11 July 2024

Published online: 14 August 2024

Check for updates

Chase M. Kayrouz[1,4], Kendra A. Ireland[2,4], Vanessa Y. Ying[1], Katherine M. Davis[2] & Mohammad R. Seyedsayamdost[1,3]

Selenium is an essential micronutrient, but its presence in biology has been limited to protein and nucleic acid biopolymers. The recent identification of a biosynthetic pathway for selenium-containing small molecules suggests that there is a larger family of selenometabolites that remains to be discovered. Here we identify a recently evolved branch of abundant and uncharacterized metalloenzymes that we predict are involved in selenometabolite biosynthesis using a bioinformatic search strategy that relies on the mapping of composite active site motifs. Biochemical studies confirm this prediction and show that these enzymes form an unusual C–Se bond onto histidine, thus giving rise to a distinct selenometabolite and potent antioxidant that we have termed ovoselenol. Aside from providing insights into the evolution of this enzyme class and the structural basis of C–Se bond formation, our work offers a blueprint for charting the microbial selenometabolome in the future.

Low-molecular-weight (LMW) selenols are an emerging and intriguing class of biomolecules. They probably possess biological functions similar to the chemically related LMW thiols, whose versatile reactivities and ability to access several oxidation states allow them to play a central role in myriad biological processes[1,2]. However, unlike the LMW thiol family, which boasts many characterized members in both primary and secondary metabolism[3], the list of known LMW selenols is exceedingly short. Of the few LMW selenols that have been detected in biological samples, only two are known to originate from Se-specific biosynthetic pathways: selenosugars and selenoneine (SEN)[4] (Fig. 1a). Both pathways were recently discovered through a selenium-focused genome mining effort, which not only revealed a new biosynthetic route but also pointed to an uncharted chemical space of natural selenometabolites.

A number of approaches can be envisioned for exploring this new chemical space. One potentially fruitful strategy revolves around the thio/selenoimidazoles (TSIs), a small but enticing class of thiol/selenol-functionalized histidine derivatives in biology, consisting of ergothioneine (EGT), ovothiol (OVO) and SEN[5] (Fig. 1a). Typically synthesized by bacteria and fungi, these LMW thiols/selenols are thought to play important, redox-based roles in the lives of their producers, as well as in higher organisms for which TSIs are diet derived. This includes humans, who accumulate high concentrations of EGT in various tissues through uptake by the membrane transporter OCTN1 (refs. 6,7). TSIs are also interesting from a biosynthetic perspective, as they involve a unique metalloenzyme class that we have termed non-haem iron sulfoxide/selenoxide synthases (NHISS) responsible for forging C–S/Se bonds. In OVO and EGT biosynthesis, the NHISS homologues OvoA and EgtB, respectively, catalyse oxidative C–S bond formation between cysteine and histidine derivatives en route to the final products[8,9] (Fig. 1b and Supplementary Fig. 1). The recently discovered NHISS SenA carries out an analogous reaction using a selenosugar as a selenium source, in place of cysteine, to generate a selenoxide intermediate in the SEN biosynthetic pathway[4] (Fig. 1b and Supplementary Fig. 1).

[1]Department of Chemistry, Princeton University, Princeton, NJ, USA. [2]Department of Chemistry, Emory University, Atlanta, GA, USA. [3]Department of Molecular Biology, Princeton University, Princeton, NJ, USA. [4]These authors contributed equally: Chase M. Kayrouz, Kendra A. Ireland. e-mail: mrseyed@princeton.edu

https://doi.org/10.1038/s41557-024-01600-2

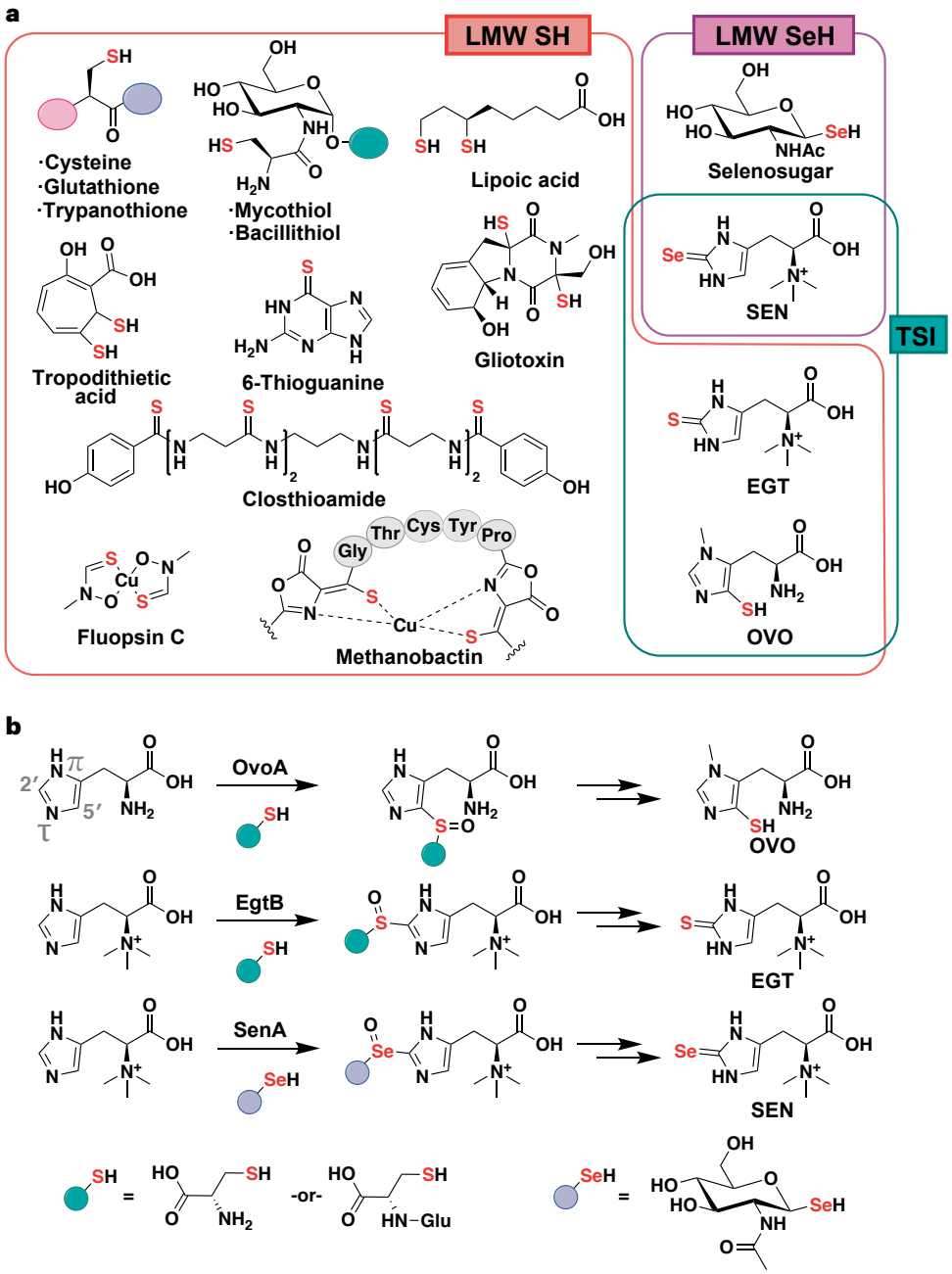

Fig. 1 | LMW thiols and selenols in biology. a, Select LMW thiols (SH), all known LMW selenols (SeH) formed by specific biosynthetic pathways, and the TSI family. LMW thiols are represented in primary and secondary metabolism, with tropodithietic acid, 6-thioguanine, gliotoxin, closthioamide, fluopsin C and methanobactin serving as members of the latter group. **b**, NHISS-catalysed C–S/Se bond forming reactions involved in the biosynthesis of TSIs.

The small number of characterized NHISS enzymes led us to wonder whether divergent homologues are involved in the biosynthesis of yet-unknown selenometabolites. To approach this question, we compiled the active site motifs of previously characterized family members and used these to traverse the phylogenetic tree of the NHISS family in search of homologues with divergent substrate binding residues. This analysis revealed a large and previously uncharacterized clade of NHISS enzymes, which we predicted may carry out previously unexplored chemistry. Using a combination of biochemical and structural biology techniques, we show they are involved in the production of a selenometabolite that we term ovoselenol (OVS). Aside from uncovering its complete biosynthetic pathway and providing insights into the structure and evolution of NHISS enzymes, our results reveal a previously unknown chemotype in the TSI family and highlight that the microbial selenometabolome is now ripe for exploration.

Results

Phylogenetic analysis of the NHISS enzyme family

We began with the hypothesis that members of the NHISS family that catalyse distinct reactions would have active site motifs divergent from those of previously characterized members. So far, five subfamilies of bacterial NHISS enzymes have been studied. Type I, II and IV EgtBs (hereafter referred to as EgtB-I, EgtB-II and EgtB-IV, respectively) catalyse C–S bond formation between cysteine and the imidazole 2′-carbon of hercynine (Nα-trimethylhistidine) during EGT biosynthesis[9–11] (Fig. 1b and Supplementary Fig. 1). OvoA, representing a fourth subfamily, introduces a C–S bond between cysteine and the 5′-carbon of the

https://doi.org/10.1038/s41557-024-01600-2

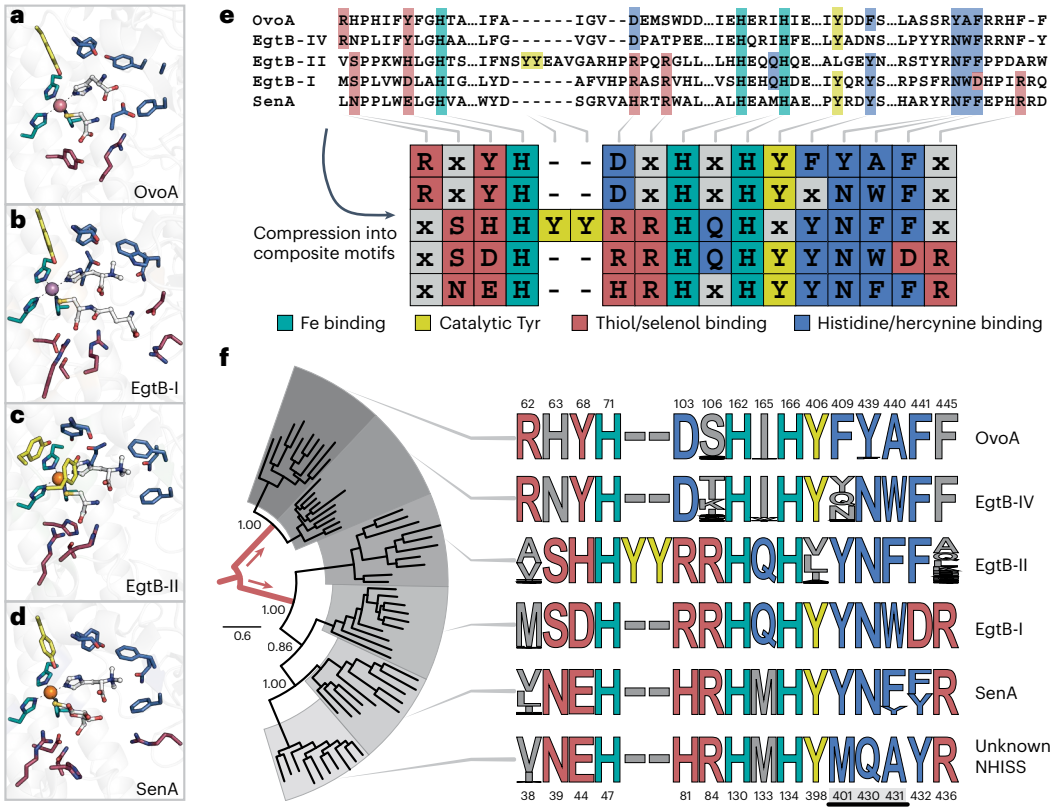

Fig. 2 | Identification of divergent NHISS clades with potentially distinct functions. a–d, Active site architectures of NHISS subfamilies, with important residues coloured according to function (PDB accession codes 8KHQ (ref. 14) (**a**), 4X8D (ref. 13) (**b**), 6O6L (ref. 10) (**c**) and 8K5I (ref. 15) (**d**)). See **e** for a legend. **e**, Compression of full-length MSAs into composite motifs, which encapsulate the relevant residues, coloured according to **a**–**d**. **f**, Mapping the composite motifs onto a phylogenetic tree of the NHISS family reveals a clade that features a unique composite motif and diverges recently from the SenA subfamily. The tree root is highlighted in red, representing the family's initial bifurcation. Bootstrap values (1,000 replicates) are displayed at key nodes. Residue position numberings for OvoA from *Hydrogenimonas thermophila* and the unknown NHISS from *H. utahensis* are shown. The key residues distinguishing the unknown NHISS from the other subfamilies are underlined.

histidine imidazole[8]. Finally, the recently identified subfamily containing SenA catalyses C–Se bond formation between a selenosugar and the imidazole 2′-carbon of hercynine[4]. Type III EgtBs are of fungal origin and will not be discussed further here[10,12].

Crystal structures of EgtB-I, EgtB-II, OvoA and SenA have been reported, revealing key active site residues involved in metal binding, substrate positioning and catalysis[10,13–15] (Fig. 2a–d). Examination of these residues by multiple sequence alignment (MSA) appears to be sufficient to predict the subfamily of known NHISS enzymes. We therefore reasoned that we could compress any NHISS protein sequence into a short, composite motif that would be predictive of its function (Fig. 2e). Applying this strategy, we retrieved all NHISS homologues from the National Center for Biotechnology Information (NCBI) database, extracted the relevant residues for each member by MSA, and mapped these composite motifs onto a phylogenetic tree of the retrieved protein sequences. A subset of this tree is displayed (Fig. 2f and Supplementary Fig. 2) and illustrates the family's evolutionary trajectory, highlighting an important early bifurcation between the OvoA-like enzymes (OvoA and EgtB-IV) and the EgtB-like enzymes (EgtB-I, EgtB-II and SenA). This analysis indicates that the vast majority of members fall into one of the previously characterized classes (Supplementary Table 4), as judged by their composite motifs. However, closer examination reveals a distinct clade, diverging most recently from the SenA family, which bears a unique composite motif and has yet to be characterized (Fig. 2f, 'Unknown NHISS'). We therefore predicted that this divergent clade, a sixth class of bacterial NHISS, may be responsible for unique chemistry, possibly involving selenium given its close homology to the SenA subfamily.

A widespread BGC encodes the biosynthesis of OVS

Within the current sequence database, the NHISS subfamily identified here contains 228 members derived almost exclusively from aquatic γ-proteobacteria (Fig. 3a,b). To elucidate their function, we began by analysing their genomic contexts to pinpoint surrounding genes that may be involved in the same biosynthetic pathway. Interestingly, many look conspicuously similar to the SEN (*sen*) biosynthetic gene clusters (BGCs), featuring a selenosugar synthesis cassette (*senBC*) in addition to the divergent NHISS. An additional fourth gene, a close homologue of the OVO Nπ-methyltransferase (Supplementary Fig. 1) that is not present in *sen* clusters, is also often co-localized with the new NHISS. These four genes consistently co-occur, albeit separated chromosomally in some cases (Fig. 3a and Supplementary Fig. 3). Furthermore, the majority of these organisms lack the histidine Nα-trimethylase gene, *egtD*, suggesting a substrate different from that of the SenA or EgtB subfamilies. We therefore speculated that these divergent NHISS enzymes may be catalysing C–Se bond formation between histidine and a selenosugar, followed by imidazole methylation by the associated methyltransferase.

To test this hypothesis, we recombinantly expressed and purified the identified NHISS enzyme and its adjacent methyltransferase from *Halomonas utahensis*. Upon joint incubation of these two enzymes with the Se donor from the *sen* pathway, *N*-acetyl-1-seleno-β-D-glucosamine (SeGlcNAc), and *S*-adenosylmethionine (SAM), the reactions were quenched with the thiol/selenol-derivatization reagent monobromobimane (mBBr) and analysed by high-performance liquid chromatography-coupled mass spectrometry (HPLC–MS)

https://doi.org/10.1038/s41557-024-01600-2

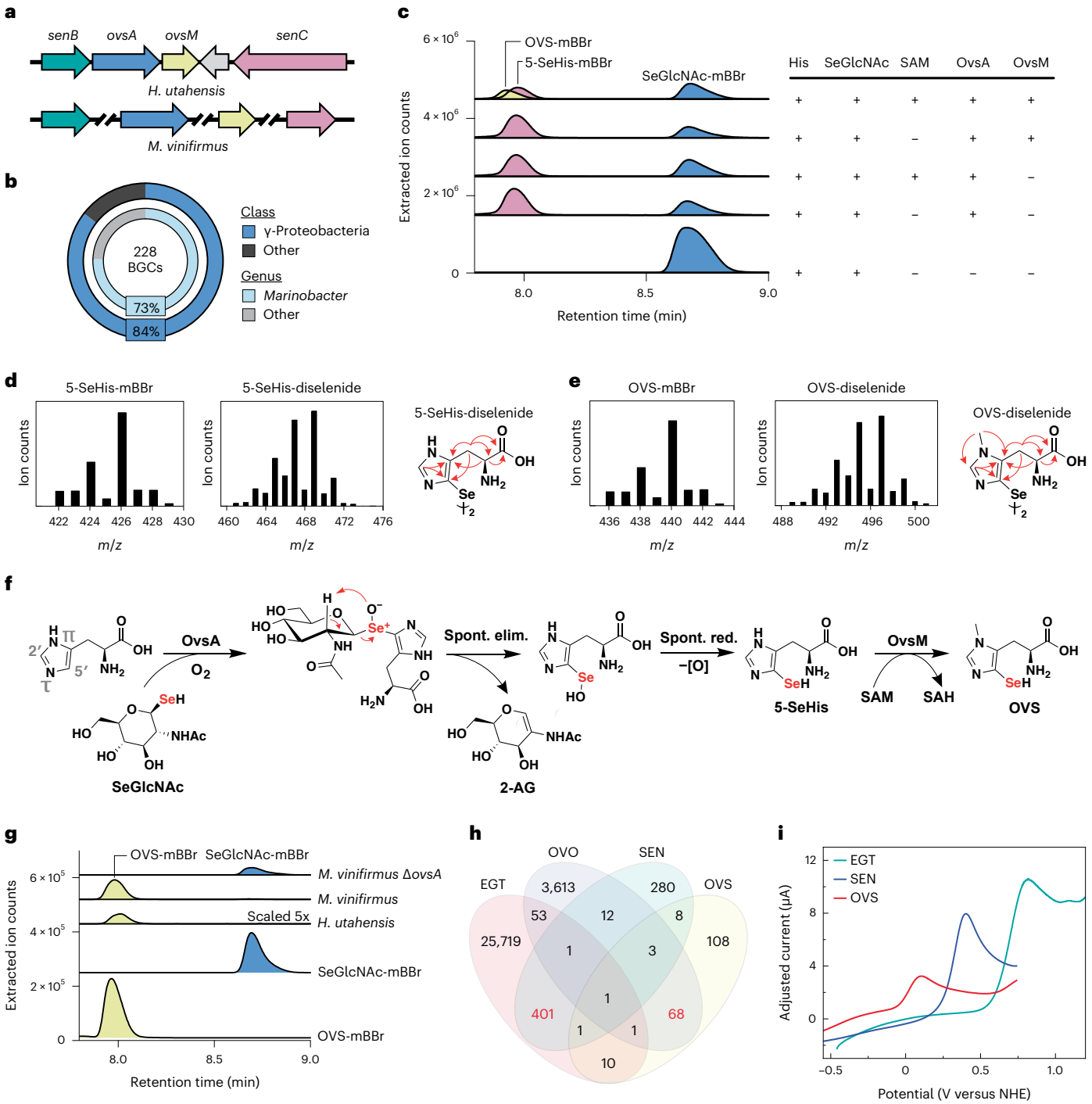

Fig. 3 | Discovery of OVS and its biosynthetic pathway. a, Genes *ovsA* (NHISS) and *ovsM* (methyltransferase) are clustered with selenosugar biosynthesis genes *senB/C* in *H. utahensis*. The grey arrow represents a hypothetical gene not conserved among *ovs* clusters. In other strains, such as *M. vinifirmus*, the genes co-occur but are chromosomally separated. **b**, *Ovs* genes are found predominantly in *Marinobacter* and other γ-proteobacteria species. **c**, In vitro reconstitution of OVS biosynthesis reveals 5′-selenylation of histidine by recombinant OvsA in the presence of SeGlcNAc, followed by SAM-dependent Nπ-methylation by OvsM. Extracted ion chromatograms of mBBr derivatives of SeGlcNAc, 5-SeHis and OVS are shown. **d**,**e**, Mass spectra of mBBr-derivatized and underivatized 5-SeHis (**d**) and OVS (**e**). Relevant $^{1}H–^{13}C$ HMBC NMR correlations (red arrows) used to solve the structure are shown. **f**, Complete biosynthetic pathway of OVS. Spont. elim. and spont. red. denote spontaneous elimination and reduction, respectively. See text for details. **g**, Identification of mBBr-derivatized OVS in crude extracts of wild-type *H. utahensis* and *M. vinifirmus*. Genetic deletion of *ovsA* in the latter results in a mutant that fails to produce OVS, accumulating SeGlcNAc as a result. Extracted ion chromatograms of mBBr derivatives of SeGlcNAc and OVS are shown, alongside mBBr derivatives of pure standards. **h**, Co-occurrence of BGCs for EGT, OVO, SEN and OVS in all sequenced microorganisms. Sequences from undefined taxa were omitted from this analysis. The combinations of EGT/SEN and OVO/OVS (shown in red) are disproportionally overrepresented. **i**, Peak oxidation potentials of OVS, and of EGT and SEN for comparison, measured by CV. See text and Supplementary Information for details.

(Fig. 3c). Remarkably, reaction of the NHISS with SeGlcNAc and histidine resulted in rapid formation of a species with a single selenium atom (m/z = 426), as judged by the characteristic MS isotope pattern of selenium-containing compounds (Fig. 3d). This intermediate was converted to a product 14 Da heavier (m/z = 440) upon the addition of the methyltransferase and SAM, indicating a single methyl transfer (Fig. 3e). Purification and nuclear magnetic resonance (NMR)-based structure elucidation of both underivatized species revealed 5-selenohistidine (5-SeHis) and Nπ-methyl-5-selenohistidine as products of the NHISS and the methyltransferase, respectively (Fig. 3d,e, Supplementary Tables 5 and 6, and Supplementary Figs. 4 and 5). The latter is the selenium isologue of OVO, which we term ovoselenol (OVS), representing the elusive fourth member of the TSI family that has been conceptualized in silico[16,17] but never observed in nature. Consequently, we designate these divergent NHISS enzymes and their associated methyltransferases as OvsA and OvsM, respectively.

Additional assays confirmed that OvsA does not accept Nπ-methylhistidine nor does OvsM methylate histidine, suggesting a clear order of biosynthetic events consisting of histidine selenylation followed by Nπ-methylation (Fig. 3f and Supplementary Fig. 6). Analogous to SenA reactivity[4], selenylation by OvsA probably proceeds through enzyme-catalysed, oxidative C–Se bond formation to furnish a short-lived selenoxide that rapidly eliminates to release the sugar moiety in the form of 2-acetamidoglucal, a species that we observe in the full reaction mixture (Fig. 3f and Supplementary Fig. 7). No reaction was observed upon replacement of SeGlcNAc with selenocysteine, cysteine or a thiosugar (1-thio-β-D-glucose), indicating a strict preference for a selenosugar substrate (Supplementary Fig. 8).

We next sought to demonstrate that OVS is the intended product of the *ovs* pathway by analysing the metabolites produced by organisms harbouring the *ovs* BGC (Fig. 3a,g). Indeed, we observed clear production of OVS upon culturing *H. utahensis* and the genetically tractable bacterium *Marinobacter vinifirmus* in the presence of sodium selenite. Furthermore, targeted gene disruption of *ovsA* in *M. vinifirmus*, a homologue of the *H. utahensis ovsA*, resulted in a mutant (Δ*ovsA*) unable to produce OVS (Fig. 3g and Supplementary Fig. 9). This mutant instead accumulated SeGlcNAc, confirming the identity of the selenosugar substrate and providing direct evidence for its production in bacteria. No OVO, 5-thiohistidine or thiosugars were observed in the samples, further demonstrating the Se specificity of the pathway. Together, these results establish OVS as a unique selenometabolite generated by the *ovs* BGC.

With OVO, EGT, SEN and now OVS, the TSI family encompasses four distinct chemotypes. Analysis of the frequency of these four BGCs reveals that, while microorganisms typically produce only one of the four, co-occurrence of *egt*/*sen* and *ovo*/*ovs* clusters are disproportionately high (Fig. 3h). In other words, there appears to be a selective advantage to producing both EGT and SEN, or OVO and OVS, as opposed to other combinations. This curious observation could point to the existence of EGT/SEN and OVO/OVS redox couples, with the Se-containing compound providing the more reducing metabolite. Consistent with this idea is the one-electron peak oxidation potential of OVS (Fig. 3i and Supplementary Fig. 10), which we measured by cyclic voltammetry (CV) to be ~110 mV versus the normal hydrogen electrode (NHE) and, thus, ~340 mV lower than that of OVO[18]. As OVS is substantially more reducing, it can act as a more effective antioxidant. The one-electron peak oxidation potentials of SEN (410 mV) and EGT (810 mV) were also determined, with the latter similar to that of thiourea[19]. These trends are mirrored by the two-electron reduction potentials reported previously for these molecules: −490 mV (ref. 20) and −60 mV (ref. 7), respectively. While we could not determine the two-electron reduction potential of OVS due to the complexity of the CV traces[21], the observed peak potentials are consistent with a model in which one TSI can reduce another, thereby forming a small-molecule redox chain. As noted above, many organisms appear to produce more than one TSI, with notable examples including *Limnobacter thiooxidans*, which encodes all four chemotypes (Supplementary Fig. 3). It is thus unlikely that these molecules are biologically redundant or interchangeable, but rather serve distinct purposes in the cell.

Structural characterization of OvsA

The reaction catalysed by OvsA points to an interesting convergence along the evolutionary trajectory of the NHISS family. Specifically, OvsA, despite sharing a recent common ancestor with the SenA/EgtB subfamilies, displays a regioselectivity similar to that of OvoA, whose N-terminal NHISS domain shares relatively little sequence similarity (20–30%; Supplementary Fig. 1) with OvsA. We therefore wondered how the evolved architecture of the OvsA active site effects this convergence. The most striking differences between the OvsA subfamily's composite motif and that of the other known subfamilies lie among the residues putatively involved in histidine/hercynine binding (Fig. 2f). For example, OvoA and OvsA, the two histidine-accepting members, feature a conserved alanine residue (A431 in OvsA) in place of the bulky aromatic residue common to hercynine-accepting subfamilies. To examine the function of these differences, we constructed an OvsA variant whose putative histidine-binding motif was altered to match that of SenA (OvsA-M401Y/Q430N/A431F, hereafter referred to as OvsA-YNF). These mutations introduced a strong preference for hercynine over histidine, as evidenced by competition assays in the presence of both histidine and hercynine (Fig. 4a,b). Moreover, selenylation of hercynine by OvsA-YNF was found to occur at the imidazole 2′-carbon, suggesting that the Nα-methylation pattern is linked to the substrate orientation within the enzymes' active sites (Fig. 4c). Thus, three simple mutations can convert OvsA into SenA.

To acquire a more precise understanding of the OvsA active site architecture, we determined its X-ray crystal structure to 2.0 Å resolution (Protein Data Bank (PDB) accession code 8U42; Supplementary Table 7). The crystallographic model reveals that OvsA's overall topology is very similar to other structurally characterized NHISS subfamilies, despite modest sequence identities of 25–35%. Alignment with its closest structural homologue, SenA, yields a Cα root-mean-square deviation of 1.8 Å (PDB accession code 8K5J; Supplementary Table 8 and Supplementary Fig. 11). Broadly speaking, OvsA displays the representative NHISS fold, composed of an N-terminal DinB domain[22] and a C-terminal domain with the formylglycine-generating enzyme subclass of the C-type lectin (CLec) fold[23,24] (see Supplementary Information for additional discussion of OvsA's overall architecture). The catalytic iron centre is coordinated by a three-His facial triad at the interface between these two domains. As in SenA, EgtB-I and OvoA, the active site of OvsA additionally features a prominent Tyr residue (Tyr398; Fig. 2a–f and Supplementary Fig. 12). Previous mechanistic studies on EgtB-I and OvoA have demonstrated that this conserved residue is crucial for sulfoxide synthase activity[25–28] and probably plays a similar role in the selenoxide synthases. Overlaying the OvsA active site with that of thioglucose-bound SenA (Supplementary Fig. 13) suggests a highly similar mode of selenosugar binding, as predicted by their composite motifs.

To further probe the structural basis for regioselectivity, we also determined the structure of OvsA complexed to histidine to 2.72 Å resolution (PDB accession code 8U41; Fig. 4d, Supplementary Table 7 and Supplementary Fig. 14). Remarkably, the 5′-imidazole carbon is oriented towards the catalytic Tyr residue (Tyr398), thus facilitating C–Se bond formation at this position. The enzyme structure is pre-organized to enable such reactivity and minimal changes to the overall scaffold are observed upon substrate binding, primarily in the loop-rich CLec domain and distant from the active site. This result is not entirely unexpected as SenA, EgtB-I and EgtB-II demonstrate similarly fixed conformations even in the presence of substrate. Correct orientation of histidine in OvsA is enabled via four key hydrogen bonding interactions between the substrate's amino and carboxylate groups and nearby

https://doi.org/10.1038/s41557-024-01600-2

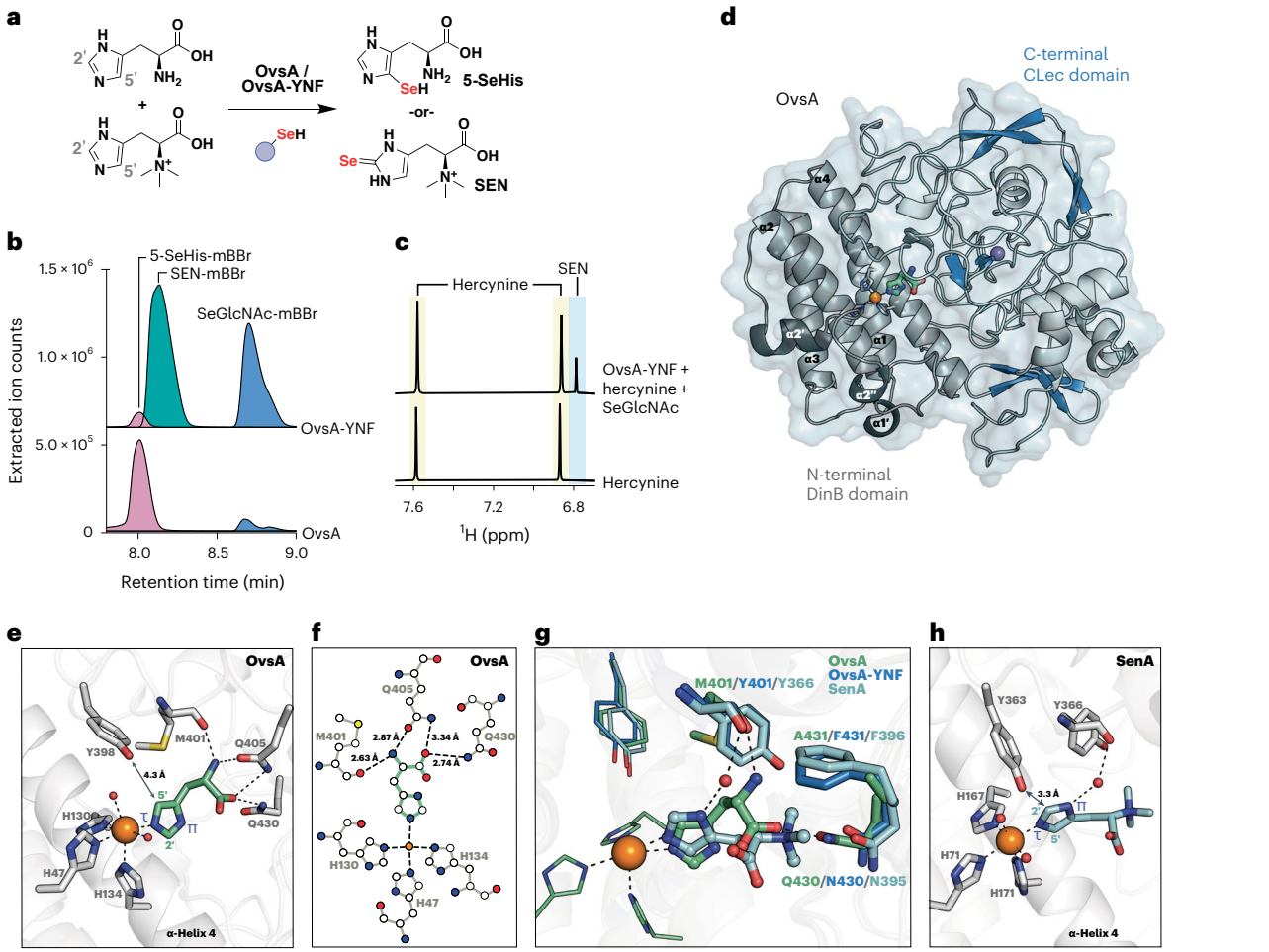

Fig. 4 | Mutagenesis and X-ray crystal structure analysis of OvsA reveal the basis for C–Se bond regioselectivity. a, Competition assays demonstrate substrate preference and regioselectivity of OvsA and OvsA-M401Y/Q430N/A451F (OvsA-YNF, mutated to match the hercynine binding motif of SenA). A 1:1 mixture of histidine and hercynine is presented to either OvsA or OvsA-YNF in the presence of SeGlcNAc, and product compositions are determined. **b**, Result of the competition assays in **a**. OvsA accepts histidine exclusively, while OvsA-YNF exhibits SenA-like activity, accepting mainly hercynine. **c**, ^{1}H-NMR analysis of the OvsA-YNF reaction with hercynine and SeGlcNAc, demonstrating 2′ C–Se bond formation to furnish SEN. **d**, X-ray crystal structure of OvsA in complex with iron (orange) and histidine (green). **e**, Active site architecture of OvsA reveals substrate interactions involved in the preference for 5′ C–Se bond formation. **f**, Two-dimensional representation of OvsA substrate binding interactions. **g**, Overlaid active sites of OvsA (green), OvsA-YNF (dark blue) and SenA (pale blue), illustrating the basis for imidazole reorientation. **h**, Active site architecture of SenA in complex with iron and hercynine (PDB accession code 8K5J), illustrating imidazole reorientation for 2′ C–Se bond formation.

residues: namely the backbone carbonyl of Met401 (2.6 Å), as well as the side chains of Gln405 (2.9 and 3.3 Å) and Gln430 (2.7 Å) (Fig. 4e,f). The substrate is additionally coordinated to the iron cofactor through its imidazole τ-nitrogen. In the absence of O_2 and a selenosugar, the remaining iron coordination sites are occupied by two water molecules.

In contrast to histidine binding by OvsA, hercynine binding in SenA is primarily accomplished through electrostatic interactions with the substrate's Nα-methyl groups, including dipolar contacts with the Asn395 side chain and cation–π interactions with Phe396 (Fig. 4g). These features are recapitulated in OvsA-YNF (Asn430 and Phe431, respectively) (Fig. 4h), the crystal structure of which we determined to 3.06 Å resolution (PDB accession code 8UX5; Supplementary Table 7). The resultant model closely aligns with that of wild-type OvsA, with the only discernible differences located in the active site. In SenA, hercynine is further stabilized by a water-mediated hydrogen bond between the backbone carbonyl of Tyr366 (Tyr401 in OvsA-YNF) and the imidazole π-nitrogen (Fig. 4g). This interaction orients hercynine such that its 2′-imidazole carbon faces the catalytic Tyr398, thus enabling substitution at this carbon. OvsA-YNF probably provides an additional level of stability through Gln405, the side chain of which reorients to interact with this same water molecule (Supplementary Fig. 15). In the absence of the Nα-methyl groups and associated dipolar interactions, the direct hydrogen bond between histidine's primary amino group and Met401 in wild-type OvsA results in rotation of the substrate backbone towards the N-terminal end of α-helix 4 (Fig. 4e,h and Supplementary Fig. 16). This reorganization of the active site forces rotation of the imidazole ring to maintain ideal bond geometry with iron, thereby enabling OvsA's observed 5′ regioselectivity.

Remarkably, OvsA shares a much more recent ancestor with SenA than with OvoA, despite displaying the regioselectivity of the latter. Consistent with the large evolutionary distance between the two, the histidine orientation is achieved in OvoA through a completely different set of binding interactions (Supplementary Fig. 17). Among these differences is a substrate hydrogen bond with D103, located on an extended loop segment in OvoA that is absent in OvsA (see Supplementary Information for a detailed comparison of OvsA and OvoA). Taken together, these structural observations indicate that regioselectivity of C–S/Se bond formation by NHISS enzymes is dictated by the orientation of the imidazole ring with respect to the catalytic Tyr residue, which is influenced by histidine's Nα-methylation pattern.

Discussion and conclusion

We report the discovery of the selenometabolite OVS, which completes a quartet of redox-active TSIs that are synthesized by diverse microorganisms. In addition to expanding the short list of selenium-containing small molecules, OVS also highlights an important evolutionary convergence within the NHISS family. The phylogeny of the NHISS family can be crudely described as an early bifurcation event leading to two main branches, namely the OvoA-like enzymes (OvoA and EgtB-IV) and the EgtB-like enzymes (EgtB-I, EgtB-II, SenA and OvsA). Despite the relatively low sequence similarity between the extant members of the two branches (Supplementary Fig. 1), nature has found ways to repurpose certain subbranches such that they exhibit catalytic features of another. The first reported example of this phenomenon is EgtB-IV, which, despite a remarkable similarity to OvoA, performs an EgtB-like reaction to furnish EGT[11]. Herein, we identify the second such example in which an enzyme from the EgtB-like branch was evolutionarily remodelled, in this case, to generate the selenium isologue of OVO. The *ovs* cluster also encodes a standalone Nπ-methyltransferase reminiscent of the C-terminal domain of OvoA[8,14,29] (Supplementary Fig. 1), suggesting another intriguing instance of gene exchange and repurposing among these biosynthetic pathways.

It is becoming increasingly evident that subtle differences among these enzymes are responsible for their distinct reactivities, which may not be obvious from sequence similarity alone. Important structural studies over the past decade have begun to illuminate the key residues responsible for engendering this catalytic diversity, affording us the ability to roughly predict the residues involved in substrate binding[10,13–15]. Informed by the structure of OvsA, we add an additional level of resolution to the prediction framework by compressing the sequences into composite motifs that can be used to classify and identify unique families of NHISS enzymes. Indeed, there remain many family members with divergent motifs whose functions have yet to be uncovered (Supplementary Fig. 2 and Supplementary Table 4).

Nature has clearly demonstrated a strong imperative towards generating all four chemotypes of histidine-derived thiols/selenols. Furthermore, in addition to repurposing of the NHISS scaffold, at least two other enzyme classes exist among anaerobic bacteria and archaea (EanB and MES) that assemble EGT by completely different mechanisms[30,31]. The ubiquity of these pathways underscores the biological importance of this family of molecules, and the remarkable specificity of the enzymes involved implies potentially non-redundant biological functions for each of the final products. Finally, we hope that the addition of OVS to the small-but-growing list of selenometabolites will inspire further studies to elucidate their roles in redox biology and, more broadly, catalyse further exploration of this emerging field.

Online content

Any methods, additional references, Nature Portfolio 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/s41557-024-01600-2.

References

1. Walsh, C. T. *The Chemical Biology of Sulfur* (The Royal Society of Chemistry, 2020).
2. Reich, H. J. & Hondal, R. J. Why nature chose selenium. *ACS Chem. Biol.* **11**, 821–841 (2016).
3. Dunbar, K. L., Scharf, D. H., Litomska, A. & Hertweck, C. Enzymatic carbon–sulfur bond formation in natural product biosynthesis. *Chem. Rev.* **117**, 5521–5577 (2017).
4. Kayrouz, C. M., Huang, J., Hauser, N. & Seyedsayamdost, M. R. Biosynthesis of selenium-containing small molecules in diverse microorganisms. *Nature* **610**, 199–204 (2022).
5. Cordell, G. A. & Lamahewage, S. N. S. Ergothioneine, ovothiol A, and selenoneine-histidine-derived, biologically significant, trace global alkaloids. *Molecules* **27**, 2673 (2022).
6. Gründemann, D. et al. Discovery of the ergothioneine transporter. *Proc. Natl Acad. Sci. USA* **102**, 5256–5261 (2022).
7. Cheah, I. K. & Halliwell, B. Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim. Biophys. Acta* **1822**, 784–793 (2012).
8. Braunshausen, A. & Seebeck, F. P. Identification and characterization of the first ovothiol biosynthetic enzyme. *J. Am. Chem. Soc.* **133**, 1757–1759 (2011).
9. Seebeck, F. P. In vitro reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **132**, 6632–6633 (2010).
10. 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).
11. Liao, C. & Seebeck, F. P. Convergent evolution of ergothioneine biosynthesis in cyanobacteria. *ChemBioChem* **18**, 2115–2118 (2017).
12. Hu, W. et al. Bioinformatic and biochemical characterizations of C–S bond formation and cleavage enzymes in the fungus *Neurospora crassa* ergothioneine biosynthetic pathway. *Org. Lett.* **16**, 5382–5385 (2014).
13. 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).
14. Wang, X. et al. Biochemical and structural characterization of OvoATh2: a mononuclear nonheme iron enzyme from *Hydrogenimonas thermophila* for ovothiol biosynthesis. *ACS Catal.* **13**, 15417–15426 (2023).
15. Liu, M. et al. Structural insights into a novel nonheme iron-dependent oxygenase in selenoneine biosynthesis. *Int. J. Biol. Macromol.* **256**, 128428 (2024).
16. Elder, J. B., Broome, J. A. & Bushnell, E. A. C. Computational insights into the regeneration of ovothiol and ergothioneine and their selenium analogues by glutathione. *ACS Omega* **7**, 31813–31821 (2022).
17. Wiebe, J., Zaliskyy, V., & Bushnell, E. A. C. A Computational investigation of the binding of the selenium analogues of ergothioneine and ovothiol to Cu(I) and Cu(II) and the effect of binding on the redox potential of the Cu(II)/Cu(I) redox couple. *J. Chem*. https://doi.org/10.1155/2019/9593467 (2019).
18. Marjanovic, B., Simic, M. G. & Jovanovic, S. V. Heterocyclic thiols as antioxidants: why ovothiol C is a better antioxidant than ergothioneine. *Free Radic. Biol. Med.* **18**, 679–685 (1995).
19. Kirchnerova, J. & Purdy, W. C. The mechanism of the electrochemical oxidation of thiourea. *Anal. Chim. Acta* **123**, 83–95 (1981).
20. Yamashita, M. & Yamashita, Y. in *Selenoneine in Marine Organisms* (ed. Kim, S.-K.) 1059–1069 (Springer, 2015).
21. Zhu, Q., Costentin, C., Stubbe, J. & Nocera, D. G. Disulfide radical anion as a super-reductant in biology and photoredox chemistry. *Chem. Sci.* **14**, 6876–6881 (2023).
22. Cooper, D. R., Grelewska, K., Kim, C.-Y., Joachimiak, A. & Derewenda, Z. S. The structure of DinB from *Geobacillus stearothermophilus*: a representative of a unique four-helix-bundle superfamily. *Acta Crystallogr. Sect. F* **66**, 219–224 (2010).
23. McMahon, S. A. et al. The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat. Struct. Mol. Biol.* **12**, 886–892 (2005).
24. Le Coq, J. & Ghosh, P. Conservation of the C-type lectin fold for massive sequence variation in a *Treponema* diversity-generating retroelement. *Proc. Natl Acad. Sci. USA* **108**, 14649–14653 (2011).

25. Chen, L. et al. Use of a tyrosine analogue to modulate the two activities of a nonheme iron enzyme OvoA in ovothiol biosynthesis, cysteine oxidation versus oxidative C–S bond formation. *J. Am. Chem. Soc.* **140**, 4604–4612 (2018).
26. Goncharenko, K. V. & Seebeck, F. P. Conversion of a non-heme iron-dependent sulfoxide synthase into a thiol dioxygenase by a single point mutation. *Chem. Commun.* **52**, 1945–1948 (2016).
27. Cheng, R. et al. OvoAMtht from *Methyloversatilis thermotolerans* ovothiol biosynthesis is a bifunction enzyme: thiol oxygenase and sulfoxide synthase activities. *Chem. Sci.* **13**, 3589–3598 (2022).
28. Chen, L. et al. Mechanistic studies of a nonheme iron enzyme OvoA in ovothiol biosynthesis using a tyrosine analogue, 2-amino-3-(4-hydroxy-3-(methoxyl) phenyl) propanoic acid (MeOTyr). *ACS Catal.* **9**, 253–258 (2019).
29. Naowarojna, N. et al. In vitro reconstitution of the remaining steps in ovothiol A biosynthesis: C–S lyase and methyltransferase reactions. *Org. Lett.* **20**, 5427–5430 (2018).
30. Burn, R., Misson, L., Meury, M. & Seebeck, F. P. Anaerobic origin of ergothioneine. *Angew. Chem.* **129**, 12682–12685 (2017).
31. Beliaeva, M. A. & Seebeck, F. P. Discovery and characterization of the metallopterin-dependent ergothioneine synthase from *Caldithrix abyssi*. *JACS Au* **2**, 2098–2107 (2022).

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 Twist Biosciences. DNA primers were purchased from Sigma. Synthesis of SeGlcNAc was carried out as previously described[4,32].

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 liquid chromatography (LC)–MS and HPLC experiments were water + 0.1% formic acid (Solvent A) and MeCN + 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 dual carbon proton (DCH) double resonance cryoprobe. NMR data were acquired using Bruker Topspin software and processed/analysed using MestReNova software.

Bioinformatics

All NHISS homologues were retrieved from the NCBI database by first performing initial blastp searches (*E*-value cut-off of 1×10^{-30}) using seed queries of previously characterized NHISS subfamilies (EgtB-I; EHI13262.1, EgtB-II; WP_014099805.1, SenA; WP_062361878.1, OvoA; CAO95049.1, EgtB-IV; WP_008197519.1). Next, ten random, divergent sequences from each BLAST result were selected to be seeds for further blastp searches to capture the full diversity of NHISS homologues in the NCBI database. Following dereplication by strain name and removal of partial sequences, 39,678 total NHISS sequences remained.

Next, MSAs were performed separately on the hits for each subfamily query using MAFFT[33] with either the L-INS-i algorithm (<1,000 sequences) or the FFT-NS-2 algorithm (>1,000 sequences). These MSAs were then used to extract the residue positions corresponding to iron-binding, histidine/hercynine-binding, thiol/selenol-binding and catalytic tyrosine motifs as reported in previous crystal structure analyses of EgtB-I[13], EgtB-II[10], OvoA[14] and SenA[15]. Residues at these 17 positions were then collapsed into a single composite motif for each retrieved NHISS. All sequences were then categorized into one of the following six subfamilies (or other) based on the following composite motif criteria, where 'x' denotes any amino acid and '–' denotes a gap in the MSA: EgtB-I, xSDH–RRHQHYYNWDR (12,211 total sequences); EgtB-II, xSHHYYRRHQHxYNFFx (16,304 total sequences); SenA, xNEH–HRHxHYYN(F/Y)(F/Y)R (1,514 total sequences); OvsA, xNEH–HRHxHYMQAYR (228 total sequences); OvoA, RxYH–DxHxHYF(Y/F)AFx (4,375 total sequences); EgtB-IV, RxYH–DxHxHYxNWFx (466 total sequences); and other (4,580 total sequences). Logo plots of composite motifs were generated using the Logomaker Python script[34].

Phylogenetic tree construction

For the tree in Fig. 2, all full-length sequences were clustered at 70% sequence identity using the CD-hit clustering tool[35], followed by random selection of ten representatives from each subfamily to be used for tree construction. For the tree in Supplementary Fig. 2, six of the most abundant uncharacterized subfamilies (grouped according to composite motifs) were selected for display, and five members of each of these subfamilies were used along with single members of the six characterized subfamilies for tree construction.

Phylogenetic trees were constructed by first generating MSAs using MAFFT with the L-INS-i algorithm, which were then used as input to FastTree[36] for maximum-likelihood phylogeny approximation with parameters -gamma, -mlacc 2 and -slownii. Trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree). The initial bifurcation between the OvoA-like (OvoA and EgtB-IV) and EgtB-like (EgtB-I, EgtB-II, SenA and OvsA) branches was selected as the tree root.

Strains, media and general culture conditions

Strains used in selenometabolite characterization and recombinant protein production experiments are listed in Supplementary Table 1. Bacto marine broth (Difco 2216) and agar plates were used for general maintenance and liquid cultures of *M. vinifirmus* DSM 17747. A growth medium consisting of 100 g l^{-1} NaCl, 5.4 g l^{-1} KCl, 10 g l^{-1} peptone, 2.5 g l^{-1} sodium acetate, 2 g l^{-1} yeast extract, 100 mM $MgSO_4$, 100 mM $MgCl_2$ and 5 mM $CaCl_2$ (+ 2 g l^{-1} agar for solid media), adjusted to pH 7.5 with Tris–HCl, was used for general maintenance and liquid cultures of *H. utahensis* DSM 3051.

Expression and purification of 6xHis-tagged OvsA, OvsM and OvsA-YNF

Genes encoding OvsA, OvsM and OvsA-M401Y/Q430N/A431F (OvsA-YNF) from *H. utahensis* DSM 3051 were obtained as synthetic DNA fragments, codon-optimized for expression in *Escherichia 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 (Supplementary Table 3). Ligation mixtures were transformed into chemically competent *E. coli* DH5α by heat shock and plated onto Luria–Bertani (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.

For protein expression, starter cultures were prepared by inoculating 15 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 at 200 rpm, starters were used to inoculate 2.8 l baffled Fernbach flasks containing 1.5 l LB broth supplemented with 50 mg l^{-1} kanamycin (1% inoculum) and incubated at 37 °C at 200 rpm. At OD_{600} of 0.5–0.6, protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside, and cultures were incubated at 18 °C at 200 rpm for an additional 12–24 h. Cells were pelleted by centrifugation (8,000*g*, 15 min, 4 °C). 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 g^{-1} cell paste), which consisted of 25 mM Tris–HCl (pH 8), 500 mM NaCl, 10 mM imidazole and 10% glycerol, supplemented with 1 mM phenylmethylsulfonyl fluoride. Once homogeneous, 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 (~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*, 15 min, 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 (pH 8), 500 mM NaCl, 300 mM imidazole and 10% glycerol. Eluted proteins were then buffer-exchanged using a 50 ml column of Sephadex G-25 (Cytiva) into storage buffer consisting of 50 mM Tris–HCl (pH 8) and 150 mM NaCl. For crystallography, OvsA and OvsA-YNF were further purified by fractionation on a Sephacryl S-200 HR HiPrep 16/60 column (Cytiva) with a running buffer consisting of 50 mM Tris–HCl (pH 8) and 150 mM NaCl. 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.

Enzyme activity assays

Assays were carried out in 100 µl reactions containing 50 mM Tris–HCl (pH 8) and freshly prepared dithiothreitol (DTT) at a final concentration

of 4 mM. Substrates were added at final concentrations of 1 mM. Enzymes were added at final concentrations of 10 µM. Reactions were initiated by the addition of SeGlcNAc (1 mM, monomer basis). Control assays were carried out in an identical fashion, except lacking one or more of the substrates/enzymes. Competition assays were performed similarly, except a mixture of 1 mM histidine and 1 mM hercynine was presented to either OvsA or OvsA-YNF.

After a 1 h incubation period at room temperature, 50 µl of each reaction mixture was quenched with 50 µl of MeOH, while another 50 µl was derivatized 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 LC–MS using a Synergi Hydro-RP column (Phenomenex, 250 × 4.6 mm, 4 µm) with a flow rate of 1 ml min^{-1}. Elution programs for MeOH-quenched samples consisted of 5% solvent B for 4 min, followed by a gradient of 5–80% solvent B over 1 min, followed by a gradient of 80–100% over 1 min and a hold at 100% for 4 min. Elution programs for mBBr-derivatized samples consisted of 5% solvent B for 3 min, followed by 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 4 min.

The assay to demonstrate 2′ C–Se bond formation by OvsA-YNF was carried out in 700 µl 50 mM Tris–HCl (pH 8) containing 4 mM DTT, 1 mM hercynine, 1 mM SeGlcNAc and 10 µM OvsA. After 1 h, protein was removed by centrifugal spin filter (10 kDa molecular weight cut-off (MWCO)) and the sample was lyophilized, redissolved in 130 µl D_2O, and analysed by ^1H NMR spectroscopy. Appearance of a single imidazole proton signal at 6.8 ppm indicates the formation of SEN, thus demonstrating the SenA-like regioselectivity of OvsA-YNF.

Purification and structural characterization of 5-SeHis diselenide

A 70 ml enzymatic reaction containing 50 mM Tris–HCl (pH 8), 10 mM DTT, 2 mM histidine, 4 mM SeGlcNAc (monomer basis) and 10 µM OvsA was incubated at room temperature for 3 h, followed by removal of protein by centrifugal spin filter (10 kDa MWCO). The crude reaction mixture was lyophilized, redissolved in water and fractionated on a Luna Omega Polar C18 column (Phenomenex, 150 × 21.2 mm, 5 µm) with a flow rate of 15 ml min^{-1} and 100% solvent A. Fractions containing 5-SeHis were further purified by fractionation on an XBridge BEH Amide OBD column (Waters, 10 × 250 mm, 5 µm) with a flow rate of 4 ml min^{-1} and 60% solvent B. Purified 5-SeHis diselenide was dissolved in D_2O and analysed by NMR spectroscopy. The selenium atom was confirmed to be positioned at the imidazole 5′-carbon, as evidenced by the absence of an imidazole 5′-proton and diagnostic ^1H–^{13}C heteronuclear multiple bond correlation (HMBC) crosspeaks. Full spectra, chemical shift assignments and select two-dimensional (2D) correlations can be found in Supplementary Fig. 4 and Supplementary Table 5.

Purification and structural characterization of OVS diselenide

A 70 ml enzymatic reaction containing 50 mM Tris–HCl (pH 8), 10 mM DTT, 2 mM histidine, 3 mM SAM, 4 mM SeGlcNAc (monomer basis), 10 µM OvsA and 10 µM OvsM was incubated at room temperature for 3 h, followed by removal of protein by centrifugal spin filter (10 kDa MWCO). The sample was acidified to pH 2 with concentrated HCl and loaded onto a Dowex 50WX8 cation exchange column pre-equilibrated with water. The column was washed with water until the elution pH was neutral, after which the column was eluted with 50 ml portions of 1%, 5%, 10%, 15% and 20% NH_4OH, sequentially. The 5% NH_4OH fraction containing OVS diselenide was lyophilized, redissolved in water and fractionated on a Luna Omega Polar C18 column (Phenomenex, 150 × 21.2 mm, 5 µm) with a flow rate of 15 ml min^{-1} and 100% solvent A. Fractions containing OVS diselenide were further purified by fractionation on an XBridge BEH Amide OBD column (Waters, 10 × 250 mm, 5 µm) with a flow rate of 4 ml min^{-1} and 60% solvent B. Purified OVS diselenide was dissolved in D_2O and analysed by NMR spectroscopy. Full spectra, chemical shift assignments and select 2D correlations can be found in Supplementary Fig. 5 and Supplementary Table 6.

Disruption of the *ovsA* gene in *M. vinifirmus*

The *ovsA* gene in *M. vinifirmus* was disrupted by replacement of an internal region of the gene with a chloramphenicol-resistant marker (Cm^R) by homologous recombination. Genomic DNA from *M. vinifirmus* DSM 17747 was isolated using the Wizard Genomic DNA Purification Kit (Promega) following the manufacturer's instructions. From genomic DNA, two ~2 kb regions flanking the *ovsA* gene were amplified using primer pairs Vini-Left-F/Vini-Left-R and Vini-Right-F/Vini-Right-R. The Cm^R gene was amplified from plasmid pGro7 (Takara Bio) using the primer pair Cm-Vini-F/Cm-Vini-R. Primers were designed such that amplicons contain overhangs to allow for assembly into pEX18Tet-SacB, a conjugative suicide vector encoding the *sacB* gene for counterselection[37] (Supplementary Table 2). Flanking regions and Cm^R gene fragments were assembled into pEX18Tet-SacB linearized with HindIII and BamHI using HiFi DNA Assembly Master Mix. Ligations were transformed into chemically competent *E. coli* DH5α by heat shock and plated onto LB agar containing 25 mg l^{-1} Cm. After confirmation by Sanger sequencing, assembled plasmids were transformed into the conjugation donor *E. coli* JV36 (ref. 38).

For conjugation, single colonies of *M. vinifirmus* and *E. coli* JV36 containing the gene disruption vector were inoculated into 5 ml marine broth and 5 ml LB supplemented with 25 mg l^{-1} Cm, respectively, and placed at 37 °C at 200 rpm. After 12 h, both strains were pelleted by centrifugation and resuspended in 300 µl marine broth. A conjugation mixture containing 50 µl donor and 50 µl recipient was spotted onto marine agar and placed at 37 °C. After 24 h, the spot was scraped from the plate and resuspended in 500 µl marine broth, and 1 µl of the mixture was plated onto marine agar supplemented with 25 mg l^{-1} Cm and 10% sucrose to select for double-crossover mutants. Single colonies were restreaked once more on the same medium. The mutant genotype was verified by PCR with the primer pair Vini-KOcheck-F/ViniKOcheck-R (Supplementary Table 2 and Supplementary Fig. 9).

Selenometabolite production screens

For each strain tested, a single colony from an agar plate was inoculated into a sterile culture tube containing 5 ml of liquid medium and incubated at 37 °C at 200 rpm. The starter cultures were then used to inoculate 25 ml liquid cultures supplemented with 50 µM of filter-sterilized Na_2SeO_3 and incubated at 37 °C at 200 rpm. Production cultures of *H. utahensis* were grown for 3 days and *M. vinifirmus* (wild type and Δ*ovsA*) for 16 h. Following incubation, cultures were pelleted by centrifugation, resuspended in 300 µl MeCN containing 10 mM mBBr and sonicated for 30 min at room temperature to facilitate cell lysis and selenol derivatization. Samples were pelleted by centrifugation, and supernatants were analysed by HPLC–MS. Analytes were separated on a Synergi Fusion-RP column (Phenomenex, 100 × 4.6 mm, 4 µm) with a flow rate of 0.5 ml min^{-1} and an elution program consisting of 5% solvent B for 3 min, followed by a gradient of 5–75% solvent B over 6 min, then a gradient of 75–100% over 1 min, and a hold at 100% for 4 min.

Enzymatic preparation of SEN diselenide

To obtain pure material for electrochemical measurements, SEN was purified from a large-scale reaction containing SenA and its substrates hercynine and SeGlcNAc. SenA from *Variovorax paradoxus* was expressed and purified as previously described[4]. A 60 ml enzymatic reaction containing 50 mM Tris–HCl (pH 8), 20 mM DTT, 5 mM hercynine, 6 mM SeGlcNAc (monomer basis) and 20 µM SenA was incubated at room temperature for 24 h, followed by removal of protein by centrifugal spin filter (10 kDa MWCO). The sample was acidified to pH 2 with concentrated HCl and loaded onto a Dowex 50WX8 cation exchange column pre-equilibrated with water. The column was washed with water until the elution pH was neutral, after which the column was eluted

with 50 ml portions of 1%, 5%, 10%, 15% and 20% NH_4OH, sequentially. The 5% NH_4OH fraction containing SEN diselenide was lyophilized, redissolved in water and purified on a Luna Omega Polar C18 column (Phenomenex, 150 × 21.2 mm, 5 µm) with a flow rate of 15 ml min^{-1} and 100% solvent A. NMR characterization: ^{1}H NMR (500 MHz, D_2O) δ 7.02 (s, 1H), 3.80 (dd, J = 11.7, 3.9 Hz, 1H), 3.11–3.24 (m, 2H), 3.17 (s, 9H); ^{13}C NMR (126 MHz, D_2O) δ 170.6, 134.9, 131.2, 120.8, 78.1, 52.1, 24.9.

Peak oxidation potential measurements

CV measurements were performed with a CH Instruments 760C potentiostat running CHI600E software using a glassy carbon (3 mm diameter) working electrode, a Pt wire counter electrode and an Ag wire pseudo-reference electrode in a conventional three-electrode cell. All cyclic voltammograms were collected in potassium phosphate buffer (0.1 M, pH 7.0). Solution containing the compound of interest was prepared at 2 mg/10 ml. The scan rate was 100 mV s^{-1} unless otherwise noted. All measurements were conducted at room temperature under an argon atmosphere. The glassy carbon working electrode was polished between measurements with an aluminium slurry on a microcloth polishing pad, followed by solvent rinses and drying under a stream of nitrogen.

To reduce the diselenide form of OVS and SEN, a solution containing either molecule was degassed under argon for 10 min and kept under a positive argon atmosphere. It was then subjected to programmed electrolysis (constant potential −0.55 V versus saturated calomel electrode (SCE), 30 min) using the same electrode assembly. Immediately following bulk electrolysis, linear sweep voltammetry data were collected. The potential of the pseudo-reference electrode was determined using the ferrocenium/ferrocene redox couple as an internal standard in anhydrous CH_3CN, previously distilled and kept over molecular sieves and K_2CO_3. Tetrabutylammonium hexafluorophosphate (nBu$_4$NPF$_6$, 0.1 M in CH_3CN) was used as the supporting electrolyte. Measured potentials were adjusted to the SCE scale (with $E_{1/2}$ taken to be 0.40 V versus SCE in CH_3CN) and then converted to NHE.

Crystallization of OvsA

Crystals of OvsA were grown using the sitting drop vapour diffusion method at room temperature. All crystals were obtained with 27 mg ml^{-1} OvsA in 50 mM Tris (pH 8) and 150 mM NaCl. Crystals of substrate-free OvsA were found to diffract more strongly when grown anaerobically in the ferrous oxidation state. Therefore, to obtain crystals of substrate-free OvsA, 2 molar equivalents (eq.) of $(NH_4)_2Fe(SO_4)_2$ was added to the protein solution, which was then sonicated in an ultrasonic water bath for 15 min to remove O_2 and left in an anaerobic chamber for an additional 4 h to ensure anaerobicity. The anaerobic protein solution was reduced with 10 mM sodium dithionite and crystallized by mixing 1:1 with a precipitant solution of 0.1 M sodium acetate (pH 4.8), 3.5 M sodium formate. To obtain crystals of OvsA bound to histidine, 5 molar eq. of $(NH_4)_2Fe(SO_4)_2$ and 20 mM histidine were added to OvsA, which was then mixed 1:1 with a precipitant solution of 0.1 M sodium acetate (pH 4.8) and 3.9 M sodium formate. To obtain crystals of OvsA-M401Y/Q430N/A431F (YNF), 5 molar eq. of $(NH_4)_2Fe(SO_4)_2$ and 20 mM histidine were added to the protein solution, which was then mixed 1:1 with a precipitant solution of 0.1 M sodium acetate (pH 4.6), 3.5 M sodium formate and 10 mM sarcosine.

Clear hexagonal rods appeared within 1 week for all conditions and were fully formed within 2 weeks. During crystal collection, the crystals were looped and briefly transferred into cryoprotectant before flash freezing in liquid N_2. For the substrate-free crystals, the cryoprotectant was composed of the precipitant with 28.3% (v/v) ethylene glycol. For the histidine-containing wild-type OvsA crystals and OvsA-YNF crystals, the cryoprotectant was composed of the precipitant with 25% (v/v) ethylene glycol and 20 mM histidine.

X-ray data collection and processing

All crystals were maintained at 100 K to minimize X-ray-induced damage during data collection. Diffraction images for wild-type OvsA were indexed and integrated using iMosflm[39], while XDS[40] was used for OvsA-YNF. The intensity data were then merged, scaled and converted to structure-factor amplitudes using POINTLESS, AIMLESS and CTRUNCATE within the CCP4 suite[41–43]. Model building was conducted in Coot[44], the structures refined in Phenix[45] and model quality assessed using Molprobity[46]. Data processing and refinement statistics can be found in Supplementary Table 7. Figures depicting the structures were generated with PyMOL, using chain A from all structures[47]. The 2D interaction diagram for histidine binding was generated with LigPlot+[48].

OvsA in complex with iron + histidine. Diffraction data for the crystals with iron and histidine were collected at beamline 21-ID-F of the Advanced Photon Source (APS) at Argonne National Laboratory using a Rayonix MX300 detector. Images were collected sequentially ($\Delta\varphi$ = 1°) with an incident wavelength of 0.9787 Å. The structure of OvsA was solved via molecular replacement with PHASER using the AlphaFold2-predicted structure as the search model[49], after removing all residues from the model with per-residue confidence scores (pLDDT) below 90%. The histidine-bound OvsA structure (PDB accession code 8U41) is in the hexagonal P 6_5 space group, contains two molecules in the asymmetric unit and was refined to 2.72 Å resolution.

OvsA in complex with iron. Diffraction data for the substrate-free crystals were collected at beamline 23-ID-B of the APS using an Eiger X 16M (Dectris) detector. Images were collected sequentially ($\Delta\varphi$ = 0.2°) with an incident wavelength of 1.0332 Å. The structure was solved via molecular replacement using the histidine-bound structure as the search model in PHASER. The substrate-free OvsA structure (PDB accession code 8U42) is in the orthorhombic P 2_1 2_1 2_1 space group, contains two molecules in the asymmetric unit and was refined to 2.0 Å resolution.

OvsA-M401Y/Q430N/A431F (OvsA-YNF) in complex with iron. Diffraction data for OvsA-YNF crystals were collected at beamline ID7B2 of the Cornell High Energy Synchrotron Source using an Eiger2 16M (Dectris) detector. Images were collected sequentially ($\Delta\varphi$ = 0.25°) with an incident wavelength of 0.9686 Å. Crystals grown with hercynine diffracted poorly, so data were instead collected on crystals grown in the presence of histidine, which resulted in better crystal growth and higher-resolution diffraction. Density was not observed, however, for bound histidine in the OvsA-YNF active site. The structure was solved via molecular replacement using the high-resolution substrate-free OvsA structure as the search model in PHASER. The structure (PDB accession code 8UX5) is in the hexagonal P 6_5 space group, contains two molecules in the asymmetric unit and was refined to 3.06 Å resolution.

Statistics and reproducibility

All enzyme activity assays, electrochemical measurements and selenometabolite production screens are representative of at least three independent experiments. PCR verification of gene disruption experiments was performed on three biological replicates (three separate exconjugants) with identical results.

Reporting summary

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

Data availability

Protein crystal structure coordinates have been deposited with the PDB (https://www.rcsb.org/) under accession numbers 8U42 (OvsA), 8U41 (Ovs+His) and 8UX5 (OvsA-YNF). Other referenced crystal structures are available in the PDB under accession numbers 8KHQ, 4X8D, 6O6L, 8K5I and 8K5J. Experimental data supporting the conclusions of this study are available within the Article and its Supplementary Information. Source data are provided with this paper and include

sequences retrieved from the NCBI Non-redundant Protein Database (https://www.ncbi.nlm.nih.gov/protein/), bioinformatic analyses and raw experimental data from main text figures. Due to large file sizes, additional raw data from Supplementary Information will be made available upon request from the corresponding author.

References

32. 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).
33. Katoh, K. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
34. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2019).
35. 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).
36. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
37. Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J. & Schweizer, H. P. A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* **212**, 77–86 (1998).
38. Blodgett, J. A. V. et al. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc. Natl Acad. Sci. USA* **107**, 11692–11697 (2010).
39. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D* **67**, 271–281 (2011).
40. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
41. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D* **67**, 282–292 (2011).
42. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution. *Acta Crystallogr. Sect. D* **69**, 1204–1214 (2013).
43. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235–242 (2012).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).
45. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221 (2010).
46. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
47. The PyMOL molecular graphics system, version 2.5.5. *Schrödinger* https://www.pymol.org/ (2023).
48. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).
49. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

Acknowledgements

We thank the Eli Lilly-Edward C. Taylor Fellowship in Chemistry (to C.M.K.), the National Science Foundation (Graduate Research Fellowship Program no. 1937971 to K.A.I. and NSF CAREER award no. 184786 to M.R.S.) and the National Institutes of Health (grant R35 GM147557 to K.M.D. and grant R01 GM140034 to M.R.S.) for financial support. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This research used resources of the APS and the Center for High-Energy X-ray Sciences (CHEXS). APS is a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). GM/CA@ APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. CHEXS is supported by the NSF award DMR-1829070, and the MacCHESS resource is supported by NIGMS award 1-P30-GM124166-01A1 and NYSTAR.

Author contributions

C.M.K. and M.R.S. conceived of the idea for the study. C.M.K. performed all bioinformatic and biochemistry experiments. K.A.I. performed all structural biology experiments. V.Y.Y. synthesized SeGlcNAc and performed electrochemical measurements. C.M.K., K.A.I., V.Y.Y., K.M.D. and M.R.S. analysed 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/s41557-024-01600-2.

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

Peer review information *Nature Chemistry* 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): June 21, 2024

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 version 10.0 Agilent OpenLab version A.01.05 Bruker Topspin version 4.4.0 CH Instruments Electrochemical Analyzer/Workstation version CHI600E
Data analysis	Agilent MassHunter Workstation Qualitative Analysis 10.0 MestReNova x64 14.1 Mosflm 7.4.0 Phenix 1.20.1_4487 Coot 0.9.8.4 AIMLESS (CCP4 Software Suite version 4) POINTLESS (CCP4 Software Suite version 4) CTRUNCATE (CCP4 Software Suite version 4) XDS version 6.30.2023 PyMOL 2.5.5 LigPlot+ 2.2.8 MAFFT version 7 Logomaker version 0.8 CD-HIT version 4.6.1 FastTree version 2.1 FigTree version 1.4.4

Molprobity version 4.5.2
AlphaFold version 2

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

Protein crystal structure coordinates have been deposited with the Protein Data Bank (PDB; https://www.rcsb.org/) under accession numbers 8U42 (OvsA), 8U41 (Ovs+His), and 8UX5 (OvsA-YNF). Other referenced crystal structures are available in the PDB under accession numbers 8KHQ, 4X8D, 6O6L, 8K5I, and 8K5J. Experimental data supporting the conclusions of this study are available within the article and its Supplementary Information. Source data provided with this paper include sequences retrieved from the NCBI Non-redundant Protein Database (https://www.ncbi.nlm.nih.gov/protein/), bioinformatic analyses, and raw experimental data from main text figures. Due to large file sizes, additional raw data from the Supplementary Information will be made available upon request from the corresponding author.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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 sampling-based experiments or statistical tests were performed. All averages reported (average log[E-values], RMSD values, etc.) are based on entire populations. A minimum of three independent replicates was performed for all experiments, consistent with previous publications in the field.
Data exclusions	No data were excluded from any analyses.
Replication	All enzyme activity assays, electrochemical measurements, and selenometabolite production screens are representative of at least three independent experiments. PCR verification of gene disruption experiments was performed on three biological replicates (three separate exconjugants) with identical results.
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
☒	☐ Plants

Methods

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

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable