Discovery and structure of a widespread bacterial ABC transporter specific for ergothioneine

L-Ergothioneine (ET), the 2-thioimidazole derivative of trimethylhistidine, is biosynthesized by select fungi and bacteria, notably Mycobacterium tuberculosis, and functions as a scavenger of reactive oxygen species. The extent to which ET broadly functions in bacterial cells unable to synthesize it is unknown. Here we show that spd_1642-1643 in Streptococcus pneumoniae, a Gram-positive respiratory pathogen, encodes an ET uptake ATP-binding cassette (ABC) transporter, designated EgtU. The solute binding domain (SBD) of EgtU, EgtUC, binds ET with high affinity and exquisite specificity in a cleft between the two subdomains, with cation-π interactions engaging the betaine moiety and a network of water molecules that surround the thioimidazole ring. EgtU is highly conserved among known quaternary amine compound-specific transporters and widely distributed in Firmicutes, including the human pathogens Listeria monocytogenes, as BilEB, Enterococcus faecalis and Staphylococcus aureus. ET increases the chemical diversity of the low molecular weight thiol pool in Gram-positive human pathogens and may contribute to antioxidant defenses in the infected host.

Cell-abundant low molecular weight (LMW) thiols maintain the reducing environment of the cytoplasm of bacterial cells and the cytosol of eukaryotic cells, and include the ubiquitous tripeptide glutathione (GSH). Bacteria unable to access glutathione often synthesize other thiols, including bacillithiol and mycothiol found in some Firmicutes and Actinomycetes, respectively. These cell-abundant LMW thiols provide protection against endogenous or exogenous reactive oxygen and nitrogen species (ROS, RNS) as ROS scavengers to create thiol disulfides which are subsequently reduced, regenerating the free thiol and thus maintaining redox balance. Bacterial LMW thiols are known to play key roles in oxidative and reductive stress responses in the infected host.

Ergothioneine (ET) is a LMW thiol and trimethylamine (betaine) derivative of histidine with sulfur installed at the imidazole C2 position. Unlike other LMW thiols that function as cellular redox buffers, ET is found in the thione tautomer rather than the thiol tautomer at physiological pH. This thione-thiol tautomerization significantly increases the thiol-disulfide reduction potential of ET relative to other LMW thiols thus endowing ET with properties of a highly effective scavenger of myriad ROS and ET also chelates transition metals including Cu+, Cu2+, and Fe2+ and thus may impact metal and metal oxidation state speciation in cells.

ET biosynthesis has been reported to occur only in select filamentous fungi including Neurospora crassa, certain cyanobacteria, and bacteria.
Methylobacterium spp., Burkholderia spp., and in Actinomycetes, including the causative agent of tuberculosis, Mycobacterium tuberculosis, and its soil saprophyte, Mycobacterium smegmatis. There have been no reports of ET biosynthesis in plants or animals.

In humans, ET is obtained from the diet and accumulates in tissues via an ergothioneine-specific transporter ETT, previously named organic cation/carnitine transporter 1 (OCTN1; SLC22A4) and a member of the Major Facilitator Superfamily (MFS). The expression level of ETT in various tissues has been used as proxy for the abundance and distribution of ET in animals. ETT expression is high in the small intestine and the kidney which reflects dietary ET uptake and ET recovery from the urine, respectively. High ETT expression has also been identified in blood cells, including erythrocytes in bone marrow, granulocytes, monocytes, and neutrophils, while detectable expression occurs in other tissues including the lung. These studies suggest that ET is bioavailable in vertebrates and could be exploited by both resident commensals and pathogenic organisms to provide protection against host oxidative stressors; however, no widespread bacterial transporter for ET is known.

Here, we describe the discovery and structural characterization of a bacterial ET transporter belonging to the ATP-binding cassette (ABC) superfamily. ABC transporters use the binding and hydrolysis of ATP to drive substrate translocation across the membrane, via two transmembrane domains (TMDs) and two cytoplasmic nucleotide-binding domains (NBDs). Prokaryotic ABC importers rely on high-affinity substrate-binding domains (SBDs) that dictate the specificity of the transporter. In Gram-negative bacteria, SBDs are typically soluble periplasmic proteins, while in Gram-positive bacteria, they are either anchored to the membrane by a covalently attached lipid or fused to periplasmic proteins, while in Gram-positive bacteria, they are either anchored to the membrane by a covalently attached lipid or fused to periplasmic proteins. Bacterial ABC importers can be subdivided into two classes, type I and type II according to the topology of their TMDs, and the two types appear to function by distinct mechanisms.

We show here that spd_1642-1643 in the Gram-positive commensal and respiratory pathogen Streptococcus pneumoniae encodes a type II ABC transporter that is highly selective for ET. The cytoplasmic ATPase, denoted EgtUA, is encoded by spd_1642. The TMD (EgtUB) and SBD (EgtUC) are fused into a single chain, denoted EgtUBC encoded by spd_1643. Quantitative LMW thiol profiling reveals that both functional EgtUA and EgtUBC are required for ET accumulation in S. pneumoniae. The ET-bound crystal structure of EgtUC, coupled with extensive NMR studies, provides novel insights into ET affinity, binding mechanism, and specificity. Bioinformatics analyses and accompanying biophysical studies reveal that EgtU is widely distributed in Firmicutes including the human pathogens Enterococcus faecalis, Staphylococcus aureus, and Listeria monocytogenes, the latter as a bile acid exclusion and electrophile stress-associated cation oxidase. This discovery expands the diversity of the LMW thiols to include ET in an important pathogen where it may contribute to antioxidant defenses in the infected host.

**Results**

**Spd_1642-1643 encodes an ABC transporter specific for ET**

We recently identified an uncharacterized operon in Streptococcus pneumoniae D39, spd_1642-1645, that is highly conserved in Streptococcus and is regulated in part by a quinone-sensing Rrf2 family transcriptional regulator, SrrA. The SrrR regulator allows access to a host-derived nutritional catabolism source, while avoiding oxidative and electrophile stress-associated catabol oxidation. This operon encodes an uncharacterized MarR family transcriptional regulator (spd_1643), a putative SnaO2 family polyketide cyclase/hydrolase (spd_1644) and an ABC transporter annotated as an osmoprotectant uptake system (Opu or Pro) that transports quaternary amines, e.g., glycine betaine (GB) or L-proline (spd_1642-1643) (Fig. 1a). Given the connection of SIRr to redox stress and iron assimilation, we hypothesized that spd_1642 is involved in ET uptake and thus named this spd_1642-1643 cluster egtU (ergothioneine uptake), where spd_1642 encodes EgtUBC and spd_1643 encodes EgtUA.

To test this hypothesis, we used a mass spectrometry-based thiol profiling strategy to quantify LMW thiols present in lysates obtained from exponentially growing S. pneumoniae cells (Fig. 1b, c; Supplementary Figs. 1–4). With this approach, isotopically labeled LMW thiol standards are spiked in at a known concentration and used to quantify the concentration of LMW thiols in the cell lysates. We find that glutathione (GSH), cysteine (Cys) and ET are major LMW thiols in a wild-type S. pneumoniae D39 strain cultured in a brain-heart infusion (BHI) rich growth medium (Fig. 1d). Moreover, a markerless ΔegtUBC strain lacks detectable ET, while ET levels are restored in an egtUBC-repaired strain (Fig. 1d). In strong contrast, cellular levels of GSH and cysteine are unaffected by the loss of egtUBC. We next quantified pneumococcal thiol levels when grown in a chemically defined medium to which variable ET was added (0.05–5 µM). These studies reveal a concentration-dependent increase in cellular ET that is lost in the ΔegtUBC strain, with no impact on Cys levels and no detectable GSH (Fig. 1e). Furthermore, two independent egtUA mutant strains derived from a ΔegtUA parent strain that express mutant EgtUA with no intron ATPase activity (Fig. 1f) fail to import ET into cells (Fig. 1f) in a way that can be rescued by reintroduction of the wild-type egtU allele.

This experiment reveals that ATP hydrolysis is required to concentrate ET against a concentration gradient, to ~1 mM in cells (Fig. 1e; Supplementary Fig. 5). These findings collectively show that EgtU is an ergothioneine-specific uptake ABC transporter in S. pneumoniae with spd_1642 encoding a transmembrane permease domain-solute binding domain (TMD-SBD) fusion protein, EgtUBC, and spd_1643 encoding the ATPase EgtAU required to power uptake.

The EgtU SBD binds ET with high affinity

EgtUBC is predicted to contain six transmembrane helices in residues 1-230 (EgtUB), followed by the soluble, extracellular substrate-binding domain (EgtUC). Recombinant EgtUBC can be expressed alone (residues 233-506), with high yield, purity, and hydrodynamic homogeneity (Supplementary Fig. 7a). This construct is thermally stable, with a melting temperature of 52 °C (Supplementary Fig. 8a, b). Perturbation of intrinsic tyrosine fluorescence (EgtUC has no Trp residues) is a convenient technique for measuring ligand binding in vitro, and has been previously applied to other QAC-binding proteins. The intrinsic tyrosine fluorescence of purified EgtUC increases upon addition of ET, accompanied by a slight red shift in the emission spectrum (Fig. 2a). These data confirm that EgtUC binds ET as a 1:1 complex, and reveal a K_d of ~2.0 × 10^5 M^{-1} (Table 1), comparable to that of E. coli HisG for histidine and a number of other SBP-ligand complexes. The perturbation of the tyrosine fluorescence suggests that EgtU may engage ET by trimethylamine cation interactions in a manner analogous to SBPs specific for osmoprotectants GB and choline.

Crystal structure of the EgtUC-ET complex

To identify the molecular determinants of ET binding by the EgtUC domain, we determined the atomic structure using X-ray crystallography (Fig. 2b). Two structures of EgtUC-ET complexes were independently obtained, at 1.82 and 2.44 Å resolution (Supplementary Table 1). The 2.44 Å structure of the holo EgtUC contains the wild-type EgtUC residues 233-506, while the 1.82 Å structure, termed EgtUC(ΔTrp), has the C-terminal five residues, GLLKK, replaced by the pair of amino acids VC. Isothermal titration calorimetry shows that EgtUC(ΔTrp) has ET binding affinity and thermodynamics that are identical to the wild-type protein (Supplementary Fig. 9a, Table 1). The structures are virtually

Nature Communications | (2022)13:7586

https://doi.org/10.1038/s41467-022-35277-3
 identical with a pairwise heavy-atom RMSD of 0.206 Å over the common regions (residues 233–501; Supplementary Fig. 9b), and we therefore use the higher resolution EgtUC-CT-E161Q structure to describe its features. The structure of EgtUC includes two globular subdomains connected by a hinge consisting of two strands ~10 residues long, with the ligand bound in the cleft between subdomains. The domain D1 consists of residues 233–331 and C-terminal residues 445–501, while D2 encompasses residues 341–432 (Fig. 2b). Each domain is characterized by a five-stranded β-sheet surrounded by five or six α-helices. The additional electron density found in the cleft between the two domains corresponds precisely to that of E161Q (Supplementary Fig. 10a). WT, wild-type; ΔegtU, markerless deletion of most of the egtU gene, which was then repaired via insertion of a wild-type egtU allele (ΔegtU repaired). a Normalized content of ET and CYS in the indicated strains of S. pneumoniae D39 grown in BHI media. WT, wild-type; ΔegtUBC, ΔegtUA ΔegtUBC repaired. **, not detected (≤0.0001 nmol thiol/mg protein). b Normalized content of ET and CYS in the indicated strains of S. pneumoniae D39 grown in CDM supplemented with indicated concentration of ET in biological triplicate. *, not detected (≤0.0001 nmol/mg protein). Molar thiol concentrations estimated from nmol/mg protein in S. pneumoniae as described elsewhere (see also Supplementary Fig. 5). f Same as d but with the indicated egtU mutant strains. g ATPase activity of WT vs. mutant EgtU proteins in vitro. **p < 0.01, ***p < 0.001 in a one-sided t-test. Data for d, e, f, and g are shown as the mean and standard deviation of three independent replicates, with individual measurements shown as red circles. Source data for d, e, f, and g are provided as a Source Data file.

The long, two-stranded hinge between domains identifies EgtUC as a type II SBP44,45, and more specifically as a member of cluster F44,45. EgtUC belongs to subcluster F-II, which employs a conserved set of aromatic residues arranged in a cage-like structure to coordinate the quaternary amine of the histidine betaine moiety. As in other subcluster F-II quaternary ammonium compound (QAC) binding proteins, notably Archaeoglobus fulgidus ProX (specific for GB)46, B. subtilis OpuBC (choline)47 and B. subtilis OpuCC (broad QAC substrate specificity)19, the aromatic residues Y341, Y419, Y443 and F293 form four sides of a pentagon and contribute cation-π interactions, while the base of the pentagon is formed by N339 (Fig. 2d). Three of these five residues, N339, Y341, and Y443, are within the two interdomain linkers. The carboxylate group of ET makes electrostatic interactions with K242, T296 and R379 (Fig. 2e, f).

The thioimidazole moiety of ET protrudes from the pentagonal cage toward the opening of the ET binding cleft, with the imidazole ring aligned roughly parallel to the domain interface. Here, E375 engages N339 in a hydrogen bonding interaction (Fig. 2e), while the hydroxyl group of T275 is close to N339 (Fig. 2f). The side chain of I243, the His and methyl group of T274, and the aliphatic region of the K242 side chain make van der Waals contact with the thione S of ET and all appear to meet the definition of a C-H-S hydrogen bond (Supplementary Table 2, Supplementary Fig. 9c). Except for T274, these residues are strongly conserved among EgtU sequences. The orientation of ET in the binding pocket is strikingly similar to that of histidine in the HisS binding pocket, despite a lack of similarity in interacting residues (Supplementary Fig. 10a). The cation-π interactions stabilize the trimethyl ammonium moiety in a similar part of the binding pocket as those of GB or choline in ATPαX19, or OpuBC19 and OpuCC19 from B. subtilis, although the latter two rely on hydrogen bonding interactions with the protein backbone rather than a salt bridge to a conserved lysine sidechain to orient the carboxylate (Supplementary Fig. 10b). In contrast, the carboxylate of ET is oriented in an entirely
different direction from that of GB in SBDs that use Trp sidechains for cation-π interactions with the trimethylamine moiety of GB, such as OpuAC from *L. lactis* or *B. subtilis* or ProX from *E. coli* (Supplementary Fig. 10c)\(^3\)\(^9\),\(^4\)^\(^5\)\(^6\). A string of highly ordered, high occupancy water molecules appears to surround the thioimidazole ring, making close contacts with the thione S and imidazole N\(^\varepsilon\)\(^2\), while also bridging conserved tyrosines Y419 and Y341 from each of the two domains (Fig. 2e – g; Supplementary Table 3). This network is connected to surface waters positioned in the cleft between the two domains. G244 is near these water molecules (Fig. 2f) and is invariant in EgtU SBDs (see below). In the GB-binding SBP *Af*ProX G244 is replaced with phenylalanine, which would severely disrupt the buried water molecules (Fig. 2h). While all *Af*ProX-family GB-specific SBPs use four Tyr to create the pentagonal cage, all EgtUs have F293 in place of *Af*ProX Y63, which is accompanied by a switch of Y337 for F107 in *Af*ProX (Fig. 2h).

**Fig. 2 | Structure of ET-bound SpEgtUC.** a Titration of ET into 1.0\(\mu\)M SpEgtUC monitored by change in intrinsic Tyr fluorescence. Each titration point is shown as the mean and standard deviation of three independent replicates. *Inset*, tyrosine emission spectra of SpEgtUC in the absence (black) and presence (red) of saturating ET. b Crystal structure of ET-bound EgtUC\(_{CTT}\) shown as ribbon, with D1 shaded light blue (residues 232 – 331) and dark blue (445-503), D2 shaded gray (341-432) and linkers colored red. ET is shown as cyan sticks. c Electron density map of ET and surrounding residues in the ligand binding pocket. d Quaternary amine region of the ET binding pocket, with residues in the aromatic pentagon shown as sticks, with polar and cation-π interactions shown as yellow dashed lines, with distances shown in Å. e ET binding pocket of D2, with backbone ribbon colored as in a. Water molecules within 4 Å of heavy atoms are shown as red spheres. Side chains in contact with water molecules or ET are shown as sticks. Polar interactions less than 4 Å are shown as yellow dashed lines. f The D1 ET binding pocket displayed as in d, with C-H-π hydrogen bonds shown as gray dashed lines (Supplementary Fig. 9c). g High-occupancy water molecules (a subset are labeled 1-9 in panels d and e; see Supplementary Table 3) lining the binding pocket and interdomain cleft, shown as red spheres. h Overlay of ligand binding pockets from the GB SBP *Af*ProX (magenta, PDB 1SW2) and SpEgtUC (blue/red/cyan), showing that a conserved G244 in SpEgtUC SBD provides space for a chain of water molecules (red spheres). The F-to-Y switch between EgtU and *Af*ProX homologs is also labeled, with F293 and Y337 in SpEgtUC and Y63 and F107 in *Af*ProX shown as sticks.

**Mutations of key residues impact ET binding affinity and thermodynamics**

In some SBPs, water molecules that line the binding pocket have been proposed to contribute to ligand promiscuity\(^7\), while in others\(^8\), they are thought to contribute to an enthalpic driving force for binding via H-bonding while also enhancing ligand selectivity. We therefore used isothermal titration calorimetry (ITC) to measure the thermodynamics of ET binding to EgtUC (Fig. 3a) and to assess the impact of mutations in perturbation of the global energetics of binding. ET binding is strongly enthalpically driven, with a \(\Delta H\) comparable to \(\Delta G\), and a small unfavorable \(\Delta T S\) value (Fig. 3a, Table 1). These thermodynamic parameters are rather similar to those previously found for the histidine-Hisj complex\(^9\), which suggests that trimethylamine cation-π interactions with Tyr/Phe are not necessarily a major net contributor to the \(\Delta H\) term in SBP-QAC ligand complexes\(^10\). The G244F substitution mutant binds ET – 100-fold more weakly than wild-type EgtUC, with a
The thermal stability of E375Q EgtUC is identical to wild-type EgtUC and is reduces the binding of ET (Fig. 2e). We bridge with R379, the side chain of which becomes strongly ordered with ET (Supplementary Fig. 8a). However, the thermal stability of G244F EgtUC is far less favorable than that of WT EgtUC (Supplementary Fig. 8b). This model closely resembles the apo-state structure of the homolog Listeria monocytogenes BiLEB (PDB 4Z7E)\(^*\), which is 60% identical to EgtUC. The individual subdomains D1 and D2 are nearly identical in the apo model and the ET-bound structure. Differences between the apo model and the ET-bound structure. Differences between the apo model and the ET-bound structure.

**Table 1 | Summary of parameters obtained for the binding of ET and other ligands to wild-type and mutant SpEgtUC**

| SBD       | ligand | Method\(^*\) | \(K_b (\text{M}^{-1})\) | \(\Delta G (\text{kcal mol}^{-1})\) | \(\Delta H (\text{kcal mol}^{-1})\) | \(\Delta S (\text{kcal mol}^{-1})\) | \(n\) |
|-----------|--------|---------------|----------------|-------------------------------|--------------------------------|--------------------------------|-----|
| SpEgtUC   | ET     | TYR           | \(2.0 \pm 0.1 \times 10^7\) | \(-10.0 \pm 0.1\) | - | - | - |
| F277W/L374C- bimane SpEgtUC | ET | PIFQ | \(2.0 \pm 0.4 \times 10^7\) | \(-9.9 \pm 0.1\) | - | - | - |
| SpEgtUC   | ET     | ITC           | \(1.7 \pm 0.1 \times 10^7\) | \(-9.9 \pm 0.1\) | \(-10.4 \pm 0.1\) | 0.5 \pm 0.1 | 0.99 \pm 0.01 |
| SpEgtUC-ET | ET | ITC | \(1.9 \pm 0.1 \times 10^7\) | \(-9.9 \pm 0.1\) | \(-10.7 \pm 0.4\) | 0.8 \pm 0.4 | 0.95 \pm 0.04 |
| SpEgtUC-GFP | ET | GFP | \(1.2 \pm 0.2 \times 10^7\) | \(-9.7 \pm 0.1\) | - | - | - |
| SpEgtUC-GFP | HER | GFP | \(5.2 \pm 0.5 \times 10^7\) | \(-5.1 \pm 0.1\) | - | - | - |
| SpEgtUC   | HER    | NMR          | \(6 \pm 1 \times 10^7\) | \(-3.8 \pm 0.1\) | - | - | - |
| SpEgtUC   | GB     | NMR          | <30                      | <2.0                         | - | - | - |
| G244F SpEgtUC | ET | ITC | \(2.0 \pm 0.3 \times 10^7\) | \(-7.2 \pm 0.1\) | \(-8.2 \pm 0.3\) | 1.0 \pm 0.2 | 0.88 \pm 0.01 |
| Y419F SpEgtUC | ET | ITC | \(1.6 \pm 0.1 \times 10^7\) | \(-8.5 \pm 0.1\) | \(-7.3 \pm 0.5\) | \(-1.2 \pm 0.4\) | 0.91 \pm 0.02 |
| F293Y SpEgtUC | ET | ITC | \(2.0 \pm 0.1 \times 10^7\) | \(-8.6 \pm 0.3\) | \(-9.9 \pm 0.2\) | 1.3 \pm 0.2 | 1.06 \pm 0.03 |
| E375Q SpEgtUC | ET | ITC | \(4.0 \pm 0.2 \times 10^5\) | \(-6.3 \pm 0.1\) | - | - | 0.75 \pm 0.03 |
| EfEgtUC  | ET     | TYR           | \(1.6 \pm 0.3 \times 10^7\) | \(-9.8 \pm 0.1\) | - | - | - |
| SaEgtUC  | ET     | TYR           | \(4.6 \pm 1.6 \times 10^7\) | \(-9.1 \pm 0.3\) | - | - | - |
| LmEgtUC | ET     | TYR           | \(1.8 \pm 0.8 \times 10^7\) | \(-9.9 \pm 0.4\) | - | - | - |

\(^*\)Methods: TYR, measured using intrinsic Tyr fluorescence enhancement; PIFQ, measured by monitoring the quenching of bimane fluorescence by a nearby Trp residue; ITC, isothermal titration calorimetry; GFP, measured using the SpEgtUC-GFP fusion protein; NMR, measured by monitoring chemical shift perturbations by NMR spectroscopy. Conditions for all other methods: 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA, 25.0°C, with the parameter values from two or three independent experiments shown as the mean and standard deviation.

**Fig. 3 | Thermodynamics of ET binding to wild-type and mutant SpEgtUCs. a–e**

Thermograms and representative ITC-derived ET binding curves obtained for WT and indicated mutant EgtUCs. The continuous lines show the best fit to a single-site binding model. Thermodynamic parameters are compiled in Table 1. Each titration shown is representative of at least two independent replicates. Source data are provided as a Source Data file.

In order to validate the model and to assess whether ET drives an open-to-closed transition in solution, we prepared a F277W/L374C...
and the binding of EgtUC (Fig. 4d; Table 1), supporting the model of ET-mediated domain closure of the SBD.

Results in a significant change in the spectrum (Fig. 4f). Backbone chemical shift perturbations (CSPs) upon binding ET for each residue in EgtUC. Assignments are missing for residues H310 and V485 in the ET-bound state (shaded pink). Prolines are shaded gray. h CSPs of ET binding painted onto the crystal structure of ET-bound SpEgtUC, with large chemical shift changes shown as thick, red tubes. i B-factors plotted on the crystal structure of SpEgtUC, with high values shown as thick, red tubes, revealing low B-factors in the interdomain linkers, and comparable to the B-factors of high occupancy solvent molecules. j Overlay of $^1$H, $^1$N TROSY spectra of apo and ET-bound EgtUC, zoomed to the region where the arginine side chain peaks are folded into this spectral window. k H-bond networks of arginine side chains (R320, R379, R404) found in the crystal structure. L $^1$H,$^1$N heteronuclear NOEs for these three arginine side chains. Heteronuclear NOE data were recorded as one replicate, with error bars indicating the uncertainty derived from spectral noise. Source data for those data shown in panel d are provided as a Source Data file.

**Fig. 4** Conformational and dynamic changes in SpEgtUC upon ET binding. a AlphaFold model of “open” apo SpEgtUC (gray D2, green D1, magenta linkers). F277W and the L374C site of bimane labeling are shown as an open hexagon and a cyan circle, respectively. b The “closed” crystal structure of ET-bound SBD (gray D2, blue D1, red linkers), with F277W and L374C as in a. c Bimane fluorescence emission spectra of bimane-labeled L374C/F277W SpEgtUC with (red) and without (black) ET. d Normalized change in fluorescence emission at 480 nm upon titration of bimane-labeled L374C/F277W SpEgtUC with ET in a position-induced fluorescence quenching (PiFQ) experiment. Each titration point is shown as the mean and standard deviation of two independent replicates. See Table 1 for fitted parameters. e $^1$H,$^1$N TROSY spectrum of apo EgtUC ( residue-specific assignments in Supplementary Fig. 12). f $^1$H,$^1$N TROSY spectrum of EgtUC bound to equimolar ET (Supplementary Fig. 13). g Backbone chemical shift perturbations (CSPs) upon binding EgtUC with ET for each residue in EgtUC. Assignments are missing for residues H310 and V485 in the ET-bound state (shaded pink). Prolines are shaded gray. h CSPs of ET binding painted onto the crystal structure of ET-bound SpEgtUC, with large chemical shift changes shown as thick, red tubes. i B-factors plotted on the crystal structure of SpEgtUC, with high values shown as thick, red tubes, revealing low B-factors in the interdomain linkers, and comparable to the B-factors of high occupancy solvent molecules. j Overlay of $^1$H,$^1$N TROSY spectra of apo and ET-bound EgtUC, zoomed to the region where the arginine side chain peaks are folded into this spectral window. k H-bond networks of arginine side chains (R320, R379, R404) found in the crystal structure. L $^1$H,$^1$N heteronuclear NOEs for these three arginine side chains. Heteronuclear NOE data were recorded as one replicate, with error bars indicating the uncertainty derived from spectral noise. Source data for those data shown in panel d are provided as a Source Data file.

Double mutant of EgtUC, introducing a nonnative Trp in D1, while attaching a bimane group to C374 in D2. Because the EgtU lacks native Trp residues, the bimane fluorescence should be quenched by Trp only when these two residues are in close proximity, termed position-induced fluorescence quenching (PiFQ) (Fig. 4a, b).31 Titration with ET results in a significant quenching of the bimane fluorescence (Fig. 4c), and the binding affinity of this construct is identical to that of wild-type EgtUC (Fig. 4d; Table 1), supporting the model of ET-mediated domain closure of the SBD.

**NMR studies of ligand-induced conformational change in EgtUC**

In order to understand the differences between the apo- and ligand-bound states of EgtUC in more detail, we turned to NMR spectroscopy. The 2D $^1$H,$^1$N TROSY spectrum of apo EgtUC shows broad chemical shift dispersion, with uniform crosspeak intensities, consistent with a globular domain with an α/β fold (Fig. 4e). Addition of equimolar ET causes significant changes in the spectrum (Fig. 4f). Backbone chemical shift assignments of the apo- and ET-bound states of $^{15}$N, $^{1}$H, $^{3}$H-SpEgtUC (Supplementary Figs. 12 and 13) give rise to chemical shift-based secondary structure predictions of the ET-bound state that are identical to the crystal structure, and are strikingly similar in the apo state, consistent with our structural model that shows minimal changes in the individual subdomains. Closer inspection reveals that the apo state has an extended β-strand at the end of the first linker and in a neighboring strand in D2 (Supplementary Fig. 11b). This small structural change in the linker is consistent with the difference between the ET-bound structure and apo model, sufficient to describe the opening and closing of the entire SBD. Chemical shift perturbations (CSPs) caused by equimolar ET binding are dramatic, but nearly exclusively localized to the domain interface (Fig. 4g, h).

We next used NMR spectroscopy to investigate the extent to which ligand binding affects protein flexibility, and in particular whether apo EgtUC samples both open and closed states, in binding ligand via a conformational selection mechanism. As expected, the bound
form features low crystallographic B-factors (Fig. 4i) and high $^{15}$N[H] heteronuclear nuclear Overhauser enhancements (hNOE) throughout, including the linkers, revealing that the linkers are strikingly rigid when bound to ET. Moreover, high hNOEs strongly suggest that the linker is also rigid in the apo state (Supplementary Fig. 11c). Mobility in both states is largely restricted to the termini and to a long loop within D1, also rigid in the apo state (Supplementary Fig. 11c). These data reveal that the binding of ET to EgtUC has strikingly little impact on sub-nm backbone dynamics throughout the molecule.

$^{13}$N, R$_1$, and R$_2$ longitudinal and transverse relaxation rates (Supplementary Fig. 11f) are largely similar in both the apo and ET-bound states, revealing relatively slow, anisotropic tumbling. The most notable difference between the two states is in the D2 helix spanning residues 375–390, which has lower R$_1$ and higher R$_2$ values in the ET-bound state. R$_1$ and R$_2$ rates are sensitive to N-H bond vector orientation as well as sub-nm flexibility, and HYDRONMR$^{52}$ can be used to distinguish whether the difference is due to a change in mobility or simply a change in conformation, by computing theoretical relaxation rates or simply a rigid body of known structure tumbling in solution with no internal mobility. HYDRONMR was therefore used to predict backbone relaxation parameters for the ET-bound crystal structure as well as for several models of the apo state. At one extreme, we examined a model identical to the crystal structure for the D1 and D2 domains (Supplementary Fig. 11i) than to the ET-bound crystal structure or to the model with fully uncoupled tumbling of the D1 and D2 domains (Supplementary Fig. 11j). A residue-by-residue analysis reveals the main difference between the two states is in the D2 helix spanning residues 301–312 already identified by low hNOEs, with line broadening in the 2D $^{13}$N, $^{1}H$ TROSY spectra as well as reduced R$_2$/R$_1$ values (Supplementary Fig. 11k).

As expected, HYDRONMR predictions for the R$_2$/R$_1$ ratio derived from the ET-bound crystal structure correlate well to the experimental parameters for ET-bound EgtUC in solution (Supplementary Fig. 11g), better than to the experimental parameters for apo EgtUC (Supplementary Fig. 11h). This result is consistent with a rigid backbone structure with mobility largely limited to loops. The apo experimental data correlate significantly better to the fully rigid AlphaFold2 model (Supplementary Fig. 11i) than to the ET-bound crystal structure or to the model with fully uncoupled tumbling of the D1 and D2 domains (Supplementary Fig. 11j). A residue-by-residue analysis reveals the main site of flexibility in the bound state is the long D1 loop (residues 301–312) already identified by low hNOEs, with line broadening in the 2D $^{13}$N, $^{1}H$ TROSY spectra as well as reduced R$_2$/R$_1$ values (Supplementary Fig. 11k). The apo state predictions match strikingly well to the experimental values (Supplementary Fig. 11i); the differences in R$_1$ observed in the helix in the middle of D2 (Supplementary Fig. 11f, upper panel) appears to derive from a reorientation of the bond vectors relative to the long axis of the molecule. These data strongly suggest that ET binding triggers an induced fit, rigid-body transition from a conformationally narrow open state to another conformationally narrow closed state, in striking contrast to expectations of a conformational selection model.

Although the backbone relaxation parameters in EgtUC are strikingly insensitive to ET binding, side chains in the binding pocket are strongly affected. Arginine guanidino protons are rarely observable in $^{13}$N, $^{1}H$ TROSY spectra acquired at pH 7.0 due to their high rate of solvent exchange, particularly when not involved in hydrogen bonds. Three slowly exchanging guanidino protons are observable (Fig. 4j), and all form hydrogen bonds in the crystal structure (Fig. 4k). R379, in particular, is sandwiched between the conserved E375 and D381 side chains, and forms a H-bond to the carboxylate oxygen of ET. The hNOEs of the side chains of R320 and R404 are unaffected by ET binding, but the hNOE of R379 is low in the absence of ligand and is dramatically increased upon binding to ET, indicating that motional disorder on the sub-nm timescale is quenched in the presence of ET (Fig. 4i).

**EgtUC binding to ET is highly specific**

We next wished to critically evaluate the specificity of ligand binding by EgtUC since this is a key feature of the function of EgtU as an ET transporter in cells. We first used differential scanning fluorimetry to show that L-hercynine induces only a small, concentration-dependent shift in $T_m$, 0.5 °C at 1 mM, which is far less than that of ET, which increases the EgtUC $T_m$ by nearly 7 °C (Supplementary Fig. 8a, b). Another high-throughput method for exploring the ligand specificity of EgtUC is also a first step to the development EgtUC as a genetically-encoded biosensor, involving the insertion of a circularly-permuted green fluorescent protein (GFP) sequence into EgtUC (Fig. 5a). Analysis of binding to the sensing domain indicates a conformational change in the GFP at the insertion site, which has been engineered to be near the chromophore, inducing a change in GFP fluorescence$^{33,34}$.

We find that our EgtUC-GFP fusion protein exhibits fluorescence that is strongly quenched upon ET binding (Fig. 5b), and that it binds ET with an affinity similar to WT EgtUC (Fig. 5c; Table 1). The low volume and high sensitivity of this experiment permit quantitative measurement of very low-affinity interactions with minimal consumption of protein and ligand, and we find that L-hercynine binds $\sim$10,000-fold less tightly than ET (Fig. 5d), consistent with the thermal unfolding results. We then used this assay to screen the ability of other potential ligands (Supplementary Fig. 14) to quench the fluorescence of the EgtUC-GFP fusion protein and find that none do so, at concentrations 1000-fold higher than the $K_d$ for ET (Fig. 5e), nor do they negatively impact the ability of ET to quench the fluorescence of EgtUC-GFP at 100-fold molar excess ligand relative to ET (Fig. 5f). ITC reveals no detectable change in global heat observed for selected other ligands, even for L-hercynine which binds weakly despite lacking only the thione sulfur atom of ET (Supplementary Fig. 14d). These experiments establish that our EgtUC-based sensor is highly specific for ET, suggesting that such a fusion protein could be used to monitor ET concentrations inside cells after further optimization$^{44}$.

NMR was next used to probe the binding of low-affinity ligands to EgtUC in more detail. A titration of L-hercynine into $^{15}$N-labeled EgtUC shows that the ligand-bound and free conformations are in fast-to-intermediate chemical exchange on the $^{1}H$ NMR timescale, with most peaks generally moving towards the corresponding resonance frequency of the ET-bound residue (Fig. 5g), while many vanish entirely. Only a few resonances, e.g., Q273 and T274, shift in a direction that is opposite to ET (Fig. 5h); these residues are in close proximity to the thione sulfur atom of ET (Fig. 2f). A large molar excess of ligand shows clear evidence of specific binding, with CSPs localized to the same interfacial loops that respond to ET (Fig. 5i; Supplementary Fig. 15a), but 2 mM hercynine was insufficient to saturate EgtUC. Fitting the chemical shift perturbations for several residues as a function of ligand concentration gives an affinity estimate of 600 M$^{-1}$ (Table 1, Fig. 5k). Titration of GB reveals only fast chemical exchange behavior, consistent with even weaker binding, as 30 mM GB fails to reach saturation (Supplementary Fig. 15b). Largely the same binding pocket residues are affected in the ET and GB complexes (Fig. 5j, i), but with an affinity estimated to be less than 30 M$^{-1}$ (Table 1).

**EgtU homologs are widely distributed across the genomes of Firmicutes**

We next asked if EgtUCs cluster in a global sequence analysis, while also elucidating conserved features of an EgtUC and how this differs from other osmoprotectant transporters. To do this, we used spd 1642 as a query to construct a sequence similarity network (SSN) using genomic enzymology tools$^{35}$ to visualize the relationships among EgtU homologs in the context of the entire superfamily of osmoprotectant uptake (opp) SBPs/SBDs (Supplementary Fig. 16). We find that EgtU is representative of a distinct subcluster of closely related sequences within SSN cluster 2 (Fig. 6a) that are characterized by the largest neighborhood connectivity of the entire SSN map (Fig. 6b; Supplementary Fig. 17). Remarkably, $^{39}$EgtU homologs are found nearly exclusively in Firmicutes and include gastrointestinal tract-resident bacteria, notably Lactococcus lactis, and a wide range of human
opportunistic pathogens beyond S. pneumoniae, including pathogenic Bacillus spp., B. cereus and B. infantalis (previously Opuf<sup>30</sup>), Enterococcus faecalis, Neisseria mucosa, Staphylococcus aureus and Listeria monocytogenes (Supplementary Fig. 18). A sequence logo representation of the multiple sequence alignment (Fig. 6c) of the SB cluster reveals that all functional features described above, including the aromatic cage and residues that interact with the imidazole and thione sulfur moieties in the α-EgtUC-ET complex, are highly conserved. Plotting the sequence conservation from this alignment onto an AlphaFold2 model of an EgtUC dimer with a single SB identifies several conserved residues at the interface that likely facilitate SB docking onto the TMD (Supplementary Fig. 19a). On the other hand, EgtU homologs found in other SSN cluster 2 subclusters do not appear to conserve key ET-specificity determinants defined here (Supplementary Fig. 19b–e)<sup>31</sup>, specifically the Y-to-F switch and G244, each of which contribute significantly to ET affinity in SpEgtUC (Fig. 2h and 3). For example, Clostridioides difficile Opuf<sup>30</sup> in the middle subcluster and others may be specific for another QAC, or exhibit relaxed QAC specificity. Indeed, very recent work in Helicobacter pylori reveals that some SSN cluster 2 transporters contribute to cellular ET uptake, but are characterized by a significantly lower affinity for ET relative to SpEgtUC<sup>31</sup>.

As a direct test of our functional grouping of proposed ET transporters, we purified and characterized candidate EgtUCs from F. faecalis, S. aureus and L. monocytogenes. We find that F/EgtUC binds ET with an affinity comparable to that of SpEgtUC (Supplementary Fig. 20a; Table 1), while NMR spectra of apo- and ET-bound F/EgtUC show similar features that are broadly consistent with comparable conformational changes to those described for SpEgtUC (Supplementary Fig. 20d, e). Since ET is obtained in the diet in animals, these findings with F. faecalis EgtU might suggest a competition for ET among resident microbiota and opportunistic pathogens in the GI

---

**Fig. 5** | The SpEgtUC binds ET with high selectivity. **a** Upper, schematic representation of the EgtUC-GFP fusion protein construct; lower, ribbon representation of an AlphaFold2 model of the EgtUC-GFP fusion, with domains indicated. **b** GFP fluorescence emission spectra, λ<sub>ex</sub> = 470 nm with or without 10 μM ET. **c** ET binding to the EgtUC-GFP fusion protein, monitored by quenching of the GFP fluorescence. Each data point is shown as the mean and standard deviation of three independent replicates. Continuous curve, fit to a 1:1 binding model; see Table 1 for binding parameters. **d** same as e except L-hercynine was added to EgtUC-GFP fusion protein (Table 1). **e** Quenching of GFP fluorescence of the EgtUC-GFP fusion protein following addition of 1, 10, or 100 μM of the indicated ligand (see Supplementary Fig. 14a for chemical structures). HIS, L-histidine; PB, proline-betaine; CHO, choline; ECT, ectoine; CAR, carnitine; DMSP, dimethylpropiothetin. Each bar represents triplicate measurements with each data point represented by a filled circle.

---

<p>Article</p>

https://doi.org/10.1038/s41467-022-35277-3

---

Nature Communications | (2022)13:7586 8
tract for some as yet unknown physiological advantage. Similar experiments were carried out with SaEC and we find a similar binding affinity, and no detectable binding by ITC to L-hercynine (Table 1; Supplementary Fig. 20b). The EgtU homolog from L. monocytogenes, denoted BilEB, has long been known to be associated with bile acid resistance\(^2\) and early work ruled out a role for BilEB in the uptake of choline, carnitine or GB\(^3\).

**Discussion**

In this work, we show that SpEgtU possesses characteristics of a bacterial ABC transporter that is specific for the low molecular weight thiol/thione L-ergothioneine (ET). We show that deletion of S. pneumoniae spd_1642 or introduction of an ATPase-inactive allele of spd_1643 creates a strain that is unable to accumulate ET either when grown in a vertebrate tissue-derived growth medium that contains significant endogenous ET, or on a chemically defined medium to which ET has been added. Our studies reveal that LmBilEB binds ET with an affinity identical to that of authentic SpEgtU (Table 1; Supplementary Fig. 20c), which argues strongly that BilEB is an ET uptake transporter.

A comparative sequence analysis suggests that EgtU is broadly distributed among Firmicutes known to colonize the vertebrate gastrointestinal (GI) tract, including commensals and pathogens, as well as pathogens known to infect other tissues but also capable of replicating in immune cells. Indeed, a recent report shows that the gut commensal bacterium, \textit{Lactobacillus reuteri}, takes up extracellular ET, although the mechanism of uptake was not defined in that study\(^6\). It is well established that commensals resist colonization by pathogens in the gut by depleting essential nutrients and remodeling resource allocation in this niche\(^6\). It is also known that ET can be catabolized by various bacteria, impacting ET bioavailability in their respective niches\(^5,6\). The work reported here raises the possibility of a competition between commensals and pathogens for a nutrient that may well
be protective against oxidative and antibiotic stressors, for which there is now evidence in *Helicobacter pylori*.

Our findings clearly establish a mechanism by which a bacterium need not synthesize ET in order to access its potential antioxidant properties in *E. coli*. In this case, ET is likely obtained in the diet of the vertebrate host where GI-resident bacteria that express EgtU would have initial access to this metabolite. As described above, the human ET transporter (ETT) is expressed in a wide range of tissues and cells, including neutrophils and macrophages; this suggests that ET may be bioavailable to both extracellular and intracellular pathogens, e.g., those phagocytosed by neutrophils. This is not yet known with certainty since concentrations of ET itself have not been comprehensively mapped in a wide variety of tissues or cells in an infected host using the analytical approaches we describe here. However, bacteria found in either the intracellular or extracellular lifestyle may well be capable of capturing ET, given that ETT is reported to transport ET with a Kₘ of 20 μM, which is ~500-fold weaker than the Kₘ for bacterial EgtUC described here. Of course, the Kₘ for transport by EgtU may not correlate with the Kₘ for substrate binding by EgtUC, and this is an important focus of future work.

What role ET plays in bacterial cell physiology can be hypothesized from literature published prior to the knowledge that EgtU homologs encode an ET transporter. For example, deletion of *egtUBC* gives rise to a fitness defect in a lung infection model in *S. pneumoniae* D39 strain and is thus a virulence factor. We provide evidence to suggest that ET is the long-sought metabolite that is transported by *Listeria monocytogenes* BIEB, required to promote an adaptive response to bile acid stress during gastrointestinal transit. How ET protects *L. monocytogenes* from bile acid stress is unknown but oxidative stress resistance is a strong possibility. In methicillin-resistant *S. aureus*, *egtUBC* expression is induced ≥15-fold after long exposures to human neutrophil-derived azurophilic granule proteins, but with no significant response to peroxide and hypochlorous acid stress at the same time points; this suggests an as yet unknown ET-dependent phagocytosis resistance mechanism to killing by these effectors. In *E. faecalis*, egtU and egtBC are among the most highly upregulated genes in a mouse model of colitis when colonized with a simplified human microbiome, but this increase in expression is lost when *E. faecalis* is monoclonized. This finding suggests that competition for this thiol may be physiologically important. Finally, although recent studies show that EgtU (denoted OpuF) from *Bacillus subtilis* strain (BDH9286), Imidazole from Chem-Impex (00184), dithiothreitol from Chem-Impex (00127), Tris-HCl from VWR Chemicals (B9800), HEPES from Chem-Impex (00174), and monobromobimane (mBBr) from Sigma-Aldrich (B4380). Other chemical reagents include IPTG from GoldBio (2481C100), TCEP from Chem-Impex (00127), Tris-HCl from MP Biomedicals (B3501), ectoine from Sigma-Aldrich (81619), proline-betaine from VWR Chemicals (TCP3055), L-carnitine from Sigma Aldrich (C0283), and dimethypropionothiohydrochloride from Sigma Aldrich (S0825). These chemicals were used without further purification. Other chemicals include ITPG from GoldBio (2481C100), TCEP from Chem-Impex (00194), dihydrothreitol from Chem-Impex (00127), Tris-HCl from MP Biomedicals (B3501), HEPS from Chem-Impex (00174), EDTA from VWR Chemicals (BDH9232), NACl from VWR Chemicals (BDH9286), Imidazole from Chem-Impex (00184), Sypro Orange from Sigma (S5692), and monobromobimane (mBBr) from Sigma (B4380). BHI broth was obtained from BD (57300, lot 1159859) while Luria broth was obtained from BDH (9286), Imidazole from Chem-Impex (00184), dithiothreitol from Chem-Impex (00127), Tris-HCl from VWR Chemicals (B9800), HEPES from Chem-Impex (00174), and monobromobimane (mBBr) from Sigma-Aldrich (S0825). Bacterial cell pellets for LMW thiol pro...
microcentrifuged at top speed for 20 min. 50 μL supernatants were transferred and filtered with 0.2 μm cutoff micro-centrifuge filter tubes. Then 1 μM heavy Δc-HPE-IAM-derivatized LMW thiol standards (ET, cysteine, GSH) were added into the flow through with total volume brought up to 100 μL with Milli-Q water. Both light and heavy Δc-HPE-IAM labeled standards were prepared by capping 100 μM reduced LMW thiolos (ET, cysteine, GSH) with 3 mM Dc-HPE-IAM in triethyl buffer at 37 °C for 1 h (see Supplementary Fig. 1). The samples were analyzed by a C18 (YMC-Triart C18) LC system coupled to a Waters SYNAPT G2S high-resolution MS using a mobile phase A (0.25% acetic acid, 10% methanol) and mobile phase B (0.25% acetic acid, 90% methanol) with the following LC elution gradient: 0-3 min, 100% A, 0% B; 3-7 min, linear gradient to 75% A, 25% B; 7-9 min, 75% A, 25% B; 9-12 min, linear gradient to 25% A, 75%B; 12-14 min, linear gradient to 0% A, 100% B; 14-20 min, 0% A, 100% B. The resulting total ion chromatogram (TIC) was searched for positively charged ions (z = 1; M’ or M + H+) (mass tolerance of ±0.02 m/z; Supplementary Fig. 1) using Waters Masslynx software and the extracted ion chromatograms of each light (H4) and heavy (Dc) HPE-IAM-capped thiol identified in the LMW obtained, peak areas quantified, and identity confirmed by LC-MS/MS by comparison to the corresponding authentic compound standard (Supplementary Figs. 2–4). The ratio of the light and heavy MSI features was used to calculate the concentration of each thiol using the known concentration heavy standard spiked into the mixture. The remaining 1 mL culture cell pellets were analyzed by Bradford Assay to quantify the total protein concentration of each sample. The LMW thiol concentration is presented as nmol thiol/mg total protein, and where indicated, used to estimate the cellular concentrations (μM) as described in the legend to Supplementary Fig. 5.

Cloning, protein expression, and purification of EgtUCs from S. pneumoniae, Enterococcus faecalis, Staphylococcus aureus, and Listeria monocytogenes

The region of the gene encoding the soluble, extracellular EgtUC domain of S. pneumoniae D39 EgtUB (locus tag spd_1642) from residue E233 was PCR-amplified from the genomic DNA. The same was done for the candidate EgtUCs of E. faecalis OG1RF EgtU (locus tag OGRF_RS02210) beginning at residue K233, S. aureus FRP3757 USA300 (locus tag sausa300_0707) beginning at residue G233, and L. monocytogenes strain 10403S (locus tag Imo1422) beginning at residue S231. The primers used in the cloning are listed in Supplementary Table 5. Each gene was inserted into the pSUMO expression vector with an N-terminal hexa-histidine tag. All mutants were prepared by PCR-based site-directed mutagenesis. The SpEgtUC-GFP expression construct was prepared using primers (Supplementary Table 5) largely following a published procedure8. In brief, the PCR fragment F1 containing SpEgtUC and pSUMO plasmid (6.5 kb) was amplified using SpEgtUC pSUMO expression vector as template and primers SpEgtU_S_P1 and SpEgtU_S_P2. PCR fragments containing the CTD or NTD of “superfolder” GFP (GFP) with linkers were prepared by using primer pairs SpEgtU_S_P3/SpEgtU_S_P4 and SpEgtU_S_P5/SpEgtU_S_P6, with the genomic DNA of I. guttatus containing sfGFP DNA sequence as a template9. The sfGFP CTD and NTD fragments were ligated together by Gibson assembly protocol10. Expression vectors were amplified in E. coli DH5α and sequences verified.

The sequence-verified expression vectors were transformed into E. coli BL21(DE3) and grown in either LB (S. pneumoniae, E. faecalis) or an M9 minimal medium (L. monocytogenes, S. aureus) supplemented with 30 μg/mL kanamycin, 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce protein expression at OD600 = 0.8. Following overnight expression at 18 °C, the cells were pelleted by centrifugation. The cell pellet was resuspended in Buffer A (25 mM Tris-HCl, pH 8), 500 mM NaCl, 10% glycerol, 20 mM imidazole), and lysed by sonication on ice. The crude lysate was clarified by centrifugation. 70% ammonium sulfate was applied to precipitate the protein and the pellet was collected by centrifugation. The precipitated pellet was resuspended in Buffer A and the solution subjected to Ni(II) immobilized affinity chromatography using a 5 mL HiTrap FF column (GE Healthcare Life Sciences) with a gradient from 100% Buffer A to 100% Buffer B (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol, 500 mM imidazole). The fractions containing the His-tagged SUMO fusion protein were pooled and digested by SUMO protease (20 μg/mL) while dialyzing in buffer A with 2 mM dithiothreitol (DTT) at room temperature. The digested protein fractions were applied to a HiTrap FF column in Buffer A. The flow-through fractions were pooled and concentrated by centrifugation with a 10 kDa cutoff and subjected to size exclusion chromatography on a Superdex 200 column in Buffer C (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA) and monomeric fractions pooled. The concentration of purified protein was measured using the molar extinction coefficients at 280 nm (ε280) (Supplementary Table 6). Purified protein fractions were pooled and stored at −80 °C until use.

Intrinsic tyrosine fluorescence titration analysis

Data were acquired on a PCl spectrofluorometer with λex 285 nm (2 mm slit) and the emission intensity recorded through a 305 nm cut-off filter. The ligand was prepared in titration buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA). All proteins were buffer exchanged into the same titration buffer and ligands were titrated into 3 mL 1 μM protein. The titrations were carried out with continuous stirring at 25.0 (±0.1) °C and resulting data corrected for dilution and the inner filter effect and fit to a 1:1 protein:ligand binding model to estimate Ka using DynaFit11, assuming a linear relationship of fluorescence signal change to fractional occupancy of EgtUC with ET.

Isothermal calorimetry titration

ITC experiments were carried out using a MicroCal VP-ITC calorimeter at 25 (±0.1) °C by titrating 20 or 30 μM SpEgtUC or the indicated mutant in the sample chamber in 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA with the indicated ligand (ET, L-hercynine, L-histidine or GB) in the syringe in same buffer. For the ET titration, the ligand concentration in the syringe was typically 375 μM with 30 μM protein in the sample chamber. For other ligands, the ligand concentration was 600 μM, and 20 μM protein in chamber. The raw ITC data were integrated, concentration normalized, and plotted as heat versus ligand/protein ratio using Origin. All data were fit to a single site binding model included in the data analysis package provided by MicroCal.

Sypro Orange differential scanning fluorimetry

Sypro Orange thermal denaturation assays were carried out for each mutant in triplicate using a 96-well plate StepOne Plus RT-PCR machine (Applied Biosystems). Each well contained 20 μL solution with 10 μM protein, 10× Sypro Orange dye, 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA. 100 μM ET was used for wells containing ET. Hercynine concentrations ranged from 10 μM to 1000 μM. The temperature was increased from 25 °C to 95 °C at a ramp rate of 1.5 °C per minute. Apparent melting temperatures (Tm) were determined from the maximum of the first derivative of the fluorescence intensity curve12.

SpEgtUC crystallography and data analysis

The purified protein was buffer exchanged into crystallography buffer, 50 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA pH 7.5. A 3-fold excess of ET was added to the purified protein and excess ligand removed by chromatography on a HiLoad 16/600 Superdex 200 size exclusion column (Cytiva). The main peak corresponding to monomeric SpEgtUC was pooled and concentrated for protein crystallography screening. SpEgtUC_S11-ET (15 mg/mL) crystals grew in sodium citrate, pH 3.6, 0.2 M potassium sodium tartrate and 1.8–2.0 M ammonium
sulfate at 20 °C using the hanging-drop vapor-diffusion method. SpEgtUC-ET (15 mg/mL) crystals grew in 1.6 M sodium citrate, pH 6.5, at 20 °C using the hanging-drop vapor-diffusion method. Crystals were harvested, cryo-protected in a reservoir solution supplemented with 25% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K at the Beamline station 4.2.2 at the Advanced Light Source (Berkeley National Laboratory, CA) and were initially indexed, integrated, and scaled using XDSD. Molecular replacement was used to estimate phases using Phaser and PDB code 4Z7E as search model. Successive cycles of automatic building in Autobuild (PHENIX) and manual building in Coot, as well as refinement (PHENIX Refine) led to complete models. MolProbity software was used to assess the geometric quality of the models, and Pymol was used to generate molecular images. Data collection and refinement statistics are indicated (Supplementary Table 1).

Structure modeling
Apo SpEgtUC was modeled using ColabFold on https://colab.research.google.com/github/skorytkin/ColabFold/blob/main/beta/AlphaFold2-advanced.ipynb with AlphaFold2 downloaded from https://github.com/deepmind/alphafold on Sep 2, 2021. SpEgtUC-GFP and the SpEgtUC dimer were modeled using colabfold version 1.3, using localcolabfold downloaded from https://github.com/YoshitakaMo/localcolabfold and installed on 4 July 2022 with default parameters. The GFP chromophore was added to the image using PDB 7S7V and Pymol. The rank 1 model of SpEgtUC was submitted to https://consurf.tau.ac.il/consurf.php with a custom multiple sequence alignment from the sequence similarity network analysis to color by sequence conservation.

Position-induced quenching of bimane fluorescence
F277W/L374C SpEgtUC was prepared as described above except that buffer for protein purification was degassed and 2 mM TCEP was added to all purification buffers. The purified protein was then buffer exchanged into degassed labeling buffer, 200 mM sodium phosphate, pH 7.4 without reducing reagent. A mBBr stock solution was prepared in DMSO at 20 mM and stored at -80 °C until used. 20 µM protein was mixed with a 30-fold molar excess of mBBr in labeling buffer at 37 °C for 1 h and excess mBBr removed by eight rounds of washing through a 10 kDa cut-off centrifugation filter with labeling buffer. The concentration of the labeled protein was measured by absorption at 280 nm using the molar extinction coefficient shown (Supplementary Table 6). The conjugated bimane concentration was measured using an εBim of 5000 M⁻¹ cm⁻¹. Data were acquired in 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA on a PC1 spectrofluorometer with a 120 ppm were used for this calculation.

NMR backbone assignments
Uniformly 15N, 13C, 1H-labeled SpEgtUC was expressed in E. coli BL21 (DE3) cells in M9 minimal medium containing 1 kg D2O, as well as 1.0 g of 15NH4Cl and 2 g 13C6-glucose as the sole nitrogen and carbon sources, respectively. Uniformly 15N-labeled protein was expressed in E. coli BL21 (DE3) cells in M9 minimal medium containing 1.0 g of 15NH4Cl as the sole nitrogen source. Further expression, isolation, and purification of these isotope-labeled proteins was performed as described above for unlabeled protein. To facilitate exchange of deuterated amides back to protons, the purified protein was incubated with 2.5 M guanidinium-HCl and 5 mM EDTA for 3 h, then dialyzed into NMR buffer (10 mM sodium phosphate, pH 7.0, 150 mM NaCl). 15N TROSY spectra on samples labeled with only 15N were used to confirm nearly complete back-exchange of the deuterated sample. NMR spectra were recorded at 35 °C on a 600 MHz Bruker Avance Neo spectrometer equipped with a cryogenic probe in the METACyt Biomolecular NMR Laboratory at Indiana University, Bloomington.

NMR samples for backbone assignment contained 0.2 mM 15N, 13C, 1H-labeled protein, with or without 0.75 mM ET, in 10 mM sodium phosphate pH 7.0, 150 mM NaCl, and 10% v/v D2O, with 0.3 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as an internal reference. Backbone chemical shifts were assigned for each state using TROSY versions of the following standard triple-resonance experiments: HNCACB, HNCOCA, HNCA, HNCO, and HNCOCA, using non-uniform sampling with Poisson gap schedules. Data were collected using Topspin 4.1.3 (Bruker) and processed using NMRPipe and isihms, and analyzed using CARA and Sparky, all on NMRbox as described. TALOS-N was used for chemical shift-based secondary structure predictions. Chemical shift perturbations (CSP) of the backbone upon ligand binding were calculated using ‘H and 15N chemical shifts with Δδ = (ΔδNH + 0.2(ΔδCa))2. Chemical shift perturbations upon interaction with L-hercynine were monitored using 0.2 mM 15N EgtUC and concentrations ranging up to 2 mM. A total of 30 mM GB was titrated into 0.15 mM 15N-labeled EgtUC.

1H spin relaxation experiments
NMR samples for relaxation experiments contained 0.75 mM 15N-labeled protein, with or without 0.75 mM ET, in 10 mM sodium phosphate pH 7.0, 150 mM NaCl, and 10% v/v D2O, with 0.3 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as an internal reference. The 1H spin relaxation rates R1 and R2, and 15N heteronuclear NOE (hNOE) values were measured using TROSY pulse sequences. The relaxation delays used were 0.05, 0.20, 0.50, 0.80, 1.2, 1.6, 2.0, and 2.5 s for R1 and 0.017, 0.034, 0.051, 0.068, 0.085, 0.102, 0.119, 0.136, 0.170, and 0.204 s for R2. Residue-specific R1 and R2 values were obtained from fits of peak intensities vs. relaxation time to a single exponential decay function, while hNOE ratios were calculated directly from intensities in experiments recorded with (2 s relaxation delay followed by 3 s saturation) and without saturation (relaxation delay of 5 s). Errors in hNOE values were calculated by propagating the error from the signal to noise. Hydrogen atoms were added to the crystal structure coordinates for ET-bound WT EgtUC and to the AlphaFold2 model of the apo state using the PDB utilities at http://spin.niddk.nih.gov/bax/nmrserver/pdbutil in order to obtain structure-based predictions for relaxation rates using HYDRONMR2. A value for the atomic radius element of 3.8 Å, the known viscosity for water at 35 °C, and CSA of ≤120 ppm were used for this calculation.

Ligand specificity analysis using the SpEgtUC-GFP titration assay
To measure the ET binding affinity with SpEgtUC-GFP, the fluorescence change upon ET titration was acquired on a PC1 spectrofluorometer with excitation at 485 nm (2 mm slit) and total emission recorded through a 510 nm cut-off filter in titration buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA) with 2 mM TCEP. ET was titrated into 3 mL 1 mM protein in the same buffer until saturation of the protein was reached. The titration was done with continuous stirring at 25.0 ± 0.1 °C and the resulting data fit to a 1:1 protein:ligand binding model to estimate Kd using DynaFit. The emission spectrum from 400 nm to 650 nm was measured before and after the titration. The initial emission intensity at 510 nm was set to 1, emission intensity at 650 nm set to 0.

To analyze the ligand specificity of SpEgtUC-GFP, triplicate 1 µM protein samples were mixed with 0, 1.0, 10, and 100 µM of the indicated ligand in 100 µL titration buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA) with 2 mM TCEP added in a 96-well plate at 25 °C. Ligands include L-ergothionine (ET), L-hercynine (HER), L-histidine (His),...
(HIS), glycine betaine (GB), proline betaine (PB), choline (CHO), ectoine (ECO), L-carnitine (CAR) and dimethylsulfoxonipropionate (DMSP). Fluorescence was obtained by excitation at 485 nm and emission at 510 nm. After the fluorescence intensity was determined, ET was added into samples to 1.0 μM with 100 μM of the indicated ligand, with the fluorescence intensity of those samples measured again. The change in fluorescence intensity, ΔF, between ET-added samples (Fs) and ET-free samples (Fo) were normalized to the ratio R defined as (Fo-Fs)/Fo.

EFI-GNN analysis

A sequence similarity network was generated using the sequence BLAST option with SpEgtUBC as the query sequence of the Uniprot database using the default Uniprot BLAST E-value of 5 using the Enzyme Function Institute–Enzyme Similarity Tool (EFI-EST; https://efi.igb.illinois.edu/efi-est/). All of the resulting sequences belonged to the pfam protein family PF04069 and were retrieved in December 2021 using the UniRef90 option. This option takes sequences that share ≥90% sequence identity over 80% of the sequence length, groups them together and represents them by a sequence known as the cluster ID. The resulting sequence file was subjected to SSN analysis using an alignment score of 120 and a minimum and maximum sequence length of 250 and 650 residues in an effort to eliminate truncation artifacts. The resulting SSN was colored and found to contain 19,991 metanodes and 57,649 unique accession IDs that segregate into 2044 non-singleton clusters and 2458 singletons and displayed as a repnode (representative node). 60% of the resulting sequences belonged to the pfam protein family PF04069 and were retrieved in December 2021 using the UniRef90 option. This option takes sequences that share ≥90% sequence identity over 80% of the sequence length, groups them together and represents them by a sequence known as the cluster ID. The resulting sequence file was subjected to SSN analysis using an alignment score of 120 and a minimum and maximum sequence length of 250 and 650 residues in an effort to eliminate truncation artifacts. The resulting SSN was colored and found to contain 19,991 metanodes and 57,649 unique accession IDs that segregate into 2044 non-singleton clusters and 2458 singletons and displayed as a repnode (representative node).

Statistical analysis methods

The number of biological or independent replicates (n) is indicated for each experiment and whenever possible all experimental data points are shown along with the standard deviation. No statistical method was used to predetermine the sample size.

Reporting summary

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

Data availability

The data that support this study are available from the corresponding author upon request. The crystallographic structures have been deposited in the Protein Data Bank under accession codes 7TXL (SpEgtUBC structure) and 7TXK (SpEgtUBC_CTT structure). NMR data are available from the BMRB under accession codes 51423 (apo-SpEgtUBC) and 51424 (ET-bound SpEgtUBC). AlphaFold2 models are available at https://modelarchive.org/ under the accession codes ma-xwg27 (apo-SpEgtUBC), ma-8pa8 (SpEgtUBC-CFF), and ma-42n23 (SpEgtUBC dimer with a single docked EgtUB). Source data are provided with this paper.

References

1. Fahey, R. C. Glutathione analogs in prokaryotes. Biochim. Biophys. Acta 1830, 3182–3198 (2013).
2. Gaballa, A. et al. Biosynthesis and functions of bacillithiol, a major low-molecular-weight thiol in Bacilli. Proc. Natl Acad. Sci. USA 107, 6482–6486 (2010).
3. Reyes, A. M. et al. Chemistry and redox biology of mycothiol. Antioxid. Redox Signal 28, 487–504 (2018).
4. Ulrich, K. & Jakob, U. The role of thiols in antioxidant systems. Free Radic. Biol. Med. 140, 14–27 (2019).
5. Reniere, M. L. Reduce, induce, thrive: bacterial redox sensing during pathogenesis. J. Bacteriol. 200, e00128–18 (2018).
6. Barger, G. & Ewins, A. J. The constitution of ergothioneine a betaine related to histidine. J. Am. Chem. Soc. 99, 2336–2341 (1911).
7. Carlson, J., Kierstan, M. P. & Brocklehurst, K. Reactions of L-ergothioneine and some other aminothiones with 2,2′-and 4,4′-dipyridyl disulfides and of L-ergothioneine with iodocadetamide. 2-Mercaptoimidazoles, 2- and 4-thiopyridines, thiourea and thiocetamide as highly reactive neutral sulphur nucleophiles. Biochem. J. 139, 221–235 (1974).
8. Cumming, B. M., Chinta, K. C., Reddy, V. P. & Sten, A. J. C. Role of ergothioneine in microbial physiology and pathogenesis. Antioxid. Redox Signal 28, 431–444 (2018).
9. Jenny, K. A., Mose, G., Haupt, D. J. & Hondal, R. J. Oxidized forms of ergothioneine are substrates for mammalian thioredoxin reductase. Antioxidants 11, 185 (2022).
10. Hanlon, D. P. Interaction of ergothioneine with metal ions and metalloenzymes. J. Med. Chem. 14, 1084–1087 (1971).
11. Motohashi, N., Mori, I., Sugiura, Y. & Tanaka, H. Metal complexes of ergothioneine. Chem. Pharm. Bull. 22, 654–657 (1974).
12. Akamnu, D., Cecchini, R., Arumoa, O. I. & Halliwell, B. The antioxidant action of ergothioneine. Arch. Biochem. Biophys. 286, 10–16 (1991).
13. Genghof, D. S. Biosynthesis of ergothioneine and hercynine by fungi and Actinomycetales. J. Bacteriol. 103, 475–478 (1970).
14. Pfeiffer, C., Bauer, T., Surek, B., Schomig, E. & Grundemann, D. Cyanobacteria produce high levels of ergothioneine. Food Chem. 126, 1766–1769 (2011).
15. Alamgir, K. M., Masuda, S., Fujitani, Y., Fukuda, F. & Tani, A. Production of ergothioneine by Mycobacterium species. Front. Microbiol. 6, 1185 (2015).
16. Garnage, A. M. et al. The proteobacterial species Burkholderia pseudomallei produces ergothioneine, which enhances virulence in mammalian infection. FASEB J. 32, 10070–10076 (2018).
17. Genghof, D. S. & Vandamme, O. Biosynthesis of ergothioneine and hercynine by mycobacteria. J. Bacteriol. 87, 852–862 (1964).
18. Genghof, D. S., Inamine, E., Kovalenko, V. & Melville, D. B. Ergothioneine in microorganisms. J. Biol. Chem. 223, 9–17 (1956).
19. Melville, D. B. & Eich, S. The occurrence of ergothioneine in plant material. J. Biol. Chem. 218, 647–651 (1956).
20. Cheah, I. K. & Halliwell, B. Ergothioneine, recent developments. Redox Biol. 42, 101868 (2021).
21. Grundemann, D., Hartmann, L. & Flogel, S. The ergothioneine transporter (ETT): substrates and locations, an inventory. FEBS Lett. 596, 1252–1269 (2021).
22. Tamai, I. et al. Cloning and characterization of a novel human pH-dependent organic cation transporter, OCTN1. FEBS Lett. 419, 107–111 (1997).
23. Grundemann, D. et al. Discovery of the ergothioneine transporter. Proc. Natl Acad. Sci. USA 102, 5256–5261 (2005).
24. Harwood, M. D., Zhang, M., Pathak, S. M. & Neuhoff, S. The pH-regional-specific relative and absolute expression of gut transporters in adult caucasians: a meta-analysis. Drug Metab. Dispos. 47, 854–864 (2019).
25. Monaco, G. et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep. 26, 1627–1640.e7 (2019).
26. Hartman, P. E. Ergothioneine as antioxidant. Methods Enzymol. 186, 310–318 (1990).
27. Berg, T. et al. Expression of MATE1, P-gp, OCTN1 and OCTN2, in epithelial and immune cells in the lung of COPD and healthy individuals. Respir. Res. 19, 68 (2018).
28. Berntsson, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J. & Poolman, B. A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–2617 (2010).
68. Goncharenko, K. V. et al. Selenocysteine as a substrate, an inhibitor and a mechanistic probe for bacterial and fungal iron-dependent sulfoxide synthases. Chemistry 26, 1328–1334 (2019).

69. Kayrouz, C. M., Huang, J., Hauser, N. & Seyedsayamdost, M. R. Biosynthesis of selenium-containing small molecules in diverse microorganisms. Nature 610, 199–204 (2022).

70. De Luna, P., Bushnell, E. A. & Gauld, J. W. A density functional theory investigation into the binding of the antioxidants ergothioneine and ovothiol to copper. J. Phys. Chem. A 117, 4057–4065 (2013).

71. Dreib, A., Brewer, M. I. & Bayse, C. A. DFT modeling of the prevention of Fe(II)-mediated redox damage by imidazole-based thiones and selenones. J. Inorg. Biochem. 193, 9–14 (2019).

72. Chalana, A. et al. Interplay of the intermolecular and intramolecular interactions in stabilizing the thione-based copper(I) complexes and their significance in protecting the biomolecules against metal-mediated oxidative damage. Polyhedron 215, 115647 (2022).

73. Lanie, J. A. et al. Genome sequence of Avery’s virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. J. Bacteriol. 189, 38–51 (2007).

74. Zhang, Y. et al. The pneumococcal iron uptake protein A (PiuA) specifically recognizes tetradentate Fe(III) bis- and mono-catechol complexes. J. Mol. Biol. 432, 5390–5410 (2020).

75. Aprianto, R., Slager, J., Holsappel, S. & Veening, J. W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Res. 46, 9990–10006 (2018).

76. Abo, M., Li, C. & Weerapana, E. Isotopically-labeled iodoacetamide-alkyne probes for quantitative cysteine-reactivity profiling. Mol. Pharm. 15, 743–749 (2018).

77. Hamid, H. A. et al. Polysulfide stabilization by tyrosine and hydroxyphenyl-containing derivatives that is important for a reactive sulfur metabolomics analysis. Redox Biol. 21, 101096 (2019).

78. Perez, A. J. et al. Movement dynamics of divisome proteins and PBP2x:FtsW in cells of Streptococcus pneumoniae. Proc. Natl Acad. Sci. USA 116, 3211–3220 (2019).

79. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 15, 343–345 (2009).

80. Kuzmic, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV protease. Anal. Biochem. 237, 260–273 (1996).

81. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212–2221 (2007).

82. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26, 795–800 (1993).

83. Echols, N. et al. Graphical tools for macromolecular crystallography in PHENIX. J. Appl. Crystallogr. 45, 581–586 (2012).

84. Headd, J. J. et al. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. Acta Crystallogr. D 68, 381–390 (2012).

85. Maciejewski, M. W. et al. NMRbox: a resource for biomolecular NMR computation. Biophys. J. 112, 1529–1534 (2017).

86. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J. Biomol. NMR 56, 227–241 (2013).

87. Tumescheit, C., Firth, A. E. & Brown, K. CIAlign: a highly customizable command line tool to clean, interpret and visualise multiple sequence alignments. PeerJ 10, e12983 (2022).

88. Jacobsen, F. E., Kazmierzczak, K. M., Lisher, J. P., Winkler, M. E. & Giedroc, D. P. Interplay between manganese and zinc homeostasis in the human pathogen Streptococcus pneumoniae. Metallomics 3, 38–41 (2011).

Acknowledgements
This research was supported by a US National Institutes of Health grant R35 GM118157 (D.P.G.). We thank Prof. Malcolm Winkler (Indiana University) for assistance in S. pneumoniae D39 strain construction and growth curves, Dr. Hongwei Wu (Indiana University) for help in some of the NMR data acquisition, and Prof. Michelle Reniere (University of Washington) for the gift of L. monocytogenes genomic DNA and comments on the manuscript.

Author contributions
Conceptualization, Y.Z., D.P.G.; Investigation: Y.Z., G.G.-G., K.A.L., C.M.P, B.J.C.W., K.A.E.; Writing: Y.Z., K.A.E., D.P.G., Supervision: Y.Z., K.A.E., D.P.G.

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/s41467-022-35277-3.

Correspondence and requests for materials should be addressed to David P. Giedroc.

Peer review information Nature Communications thanks Dirk Gruen- demann, Ioannis Manolaridis, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022