Materials and methods

Generation of Δ*sodA*Δ*sodB* expression strains of *E. coli* BL21(DE3) and pLysSRARE2 (DE3) pLysS

The Fe/Mn SOD gene knock out (KO) strain of *E. coli* BL21(λ DE3) was generated following a standard protocol[1] with minor modifications. First, the sequences of *sodA* and *sodB* genes in the BL21 genome (GCF_000022665.1, ASM2266v1) were compared to that of the *E. coli* K-12 strain (GCF_000005845.2, ASM584v2) where successful gene KO strains were previously generated[2]. Specificity of designed primers with homology to the targeted genes was verified using local BlastN (-task blastn-short, evalue 0.1)[3] against the whole BL21 genome. *E. coli* BL21(λDE3) cells expressing λ Red recombinase (BL21-RED) from pKD46 plasmid were induced with 10 mM arabinose at OD_{600}=0.2 and made electrocompetent at OD_{600}=0.6 by three washes in ice cold 10 % (v/v) glycerol. A kanamycin resistance cassette with recombination overhangs was amplified from pKD4 plasmid template with Q5® Hot Start DNA Polymerase (NEB), purified with Monarch® DNA gel extraction kit (NEB), and electroporated (500 ng DNA, 250 ng/µL) into 90 µL electrocompetent BL21-RED cells at 2.5 kV, 200 Ohm, 0.025 F for 4.5 ms (BTX ECM 630) in 0.1 cm electroporation cuvettes (VWR), followed by 2 h recovery in SOC medium. All bacterial cultures (except recovery following electroporation in SOC medium) were grown in LB medium supplemented with 0.2 % (w/v) glucose, with 1.5 % (w/v) agar for solid medium. The remaining steps of the protocol were performed as described in the original publication[1]. The BL21(λDE3) Δ*sodB* mutant was used for subsequent generation of the BL21(λDE3) Δ*sodA*Δ*sodB* double mutant following the same protocol. Correct antibiotic resistance cassette insertions were verified using diagnostic PCR with OneTaq® Quick-Load® 2X Master Mix (NEB). The gene deletion was verified using PCR with sodA(B)_screen_F(R) primers complementary to regions upstream and downstream from the deleted loci, and its phenotype confirmed using the in-gel SOD activity assay on soluble protein extracts from the WT and KO strains (FigS10C). The BL21(DE3) pLysSRARE2 ΔsodAΔsodB strain was constructed by transforming the BL21(λDE3) Δ*sodA*Δ*sodB* with pLysSRARE2 plasmid isolated from the commercially available ROSETTA2 (λDE3) pLysS strain (Novagen). All primers used for KO strain generation and verification were synthesised by Eurofins Genomics, with sequences provided in Source Data File S1.

SOD genes synthesis, cloning, and site-directed mutagenesis

Amino acid sequences of Fe/Mn SODs of interest identified in bioinformatic analyses were used to generate gene sequences codon-optimised for expression in *Escherichia coli* B using commercial Codon Optimization Tool (IDT). Predicted N-terminal targeting sequences identified with SignalP5[4], TargetP 2.0[5], ChloroP[6], and Phobius[7], and fusion domains identified with InterProScan 5[8], were manually verified in protein alignments as not being a part of the structural and functional conserved core of the SOD fold, and were excluded from the construct sequences. Sequence extensions complementary to the cloning site within pET-22b(+) (Novagen) were added to each codon optimised gene and resulting constructs (FileS2) were synthesised as gBlocks gene fragments (IDT). The amino acid sequences of the 82 wild type (WT) SOD constructs investigated in this study were listed in Source Data File S3. The synthetic constructs were used directly for enhanced Gibson assembly cloning[9,10] into the pET-22b expression vector backbone. Ratios of 4 fmol plasmid (15 ng) : 24 fmol insert (usually 10 ng) in DNA mix, and 2.5 µL of the DNA mix : 7.5 µL enzyme mix were used for the enhanced Gibson assembly. Site-directed mutagenesis of the SODs in pET-22b was performed using Q5® Hot Start DNA Polymerase (NEB) and KLD kit (NEB) following the manufacturer's protocol with the single exception of using 5 µL instead of 10 µL final reaction volume. Sequences of all primers used in site-directed mutagenesis are provided in FileS1. All plasmid propagations steps were performed in *E. coli* DH5α.

Expressing Fe/Mn SODs in the presence of Fe or Mn in 24-well culture plates

For screening of soluble expression and aCR verification experiments, BL21(λDE3) Δ*sodA*Δ*sodB* cells transformed with pET-22b SOD expression constructs were cultured in 4-5 mL M9 medium (0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 0.4 % glucose, 0.2 % casamino acids, 1 mM thiamine, 12.8 g/L $Na_2HPO_4x7H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl) in 24-well culture plates (Whatman Uniplate) at 37°C with 180 rpm orbital shaking. Pre-cultures were grown for 16 h in M9 medium containing 100 µg/mL ampicillin without metal supplementation. Aliquots (200 µL) of the pre-cultures were used to inoculate 4 mL M9 with ampicillin (100

μg/mL) medium supplemented with either 500 μM $MnCl_2$ or 100 μM $Fe(NH_4)_2SO_4$ and cultured until OD_{600} reached 0.4-0.6. At that point, additional Mn (1 mM final concentration) or Fe (200 μM final concentration) and isopropyl ß-D-1-thiogalactopyranoside (IPTG, 1 mM final concentration) were added, and the culture was continued for 4 h. Cultures in Fe- or Mn-supplemented media were always performed in separate 24-well plates to avoid metal cross-contamination. The final OD_{600} was measured, and the 24-well culture plates were centrifuged for 10 min at 4,000 *g*. Spent media were removed, and cell pellets were resuspended in cell lysis solution (20 mM Tris pH 7.5, 100 μg/mL lysozyme, 10 μg/mL DNAseI, 1x Roche cOmplete EDTA-free protease inhibitors), volume normalised based on their final OD_{600} measurements and transferred to 1.5 mL Eppendorf tubes. *E. coli* cells were lysed by 10 s sonication (MSE Soniprep 150 with exponential probe at 10-14 microns amplitude) in a cold room and centrifuged for 20 min at 21,100 *g*. The same protocol was followed for BL21(DE3) pLysSRARE2 Δ*sodA*Δ*sodB* except for 16 h expression at 16°C in M9 media containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol. The supernatants (soluble protein extracts) containing the expressed Fe/Mn SODs were used in subsequent bicinchoninic acid (BCA) protein concentration assay (Pierce), liquid and in-gel SOD activity assays, and SDS-PAGE electrophoresis. Metal purity of culture media were verified using inductively coupled plasma mass spectrometry (ICP-MS).

Protein expression and purification

Proteins were expressed in *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* in 1 L M9 medium supplemented with 200 μM $Fe(NH_4)_2SO_4$ or 1 mM $MnCl_2$ and induced with 1 mM IPTG for 4 h at 37°C with 220 rpm orbital shaking in 2 L baffled flasks. Bacterial cells were lysed by sonication of washed cell pellets in 20 mM Tris pH 7.5, 1x cOmplete EDTA-free protease inhibitor (Roche), 100 μg/mL lysozyme, 10 μg/mL DNAseI**,** followed by centrifugation at 19,000 *g* at 4°C. Cleared cell lysates were subjected to chromatographic separation using an ÄKTA purification system (Cytiva). Standard protocol for most characterised recombinant SOD proteins consisted of purification using anion exchange chromatography (AEC; HiTrap Q HP column, Cytiva) in 20 mM Tris pH 7.5 buffer with 0–1 M NaCl linear gradient elution, and subsequent size exclusion chromatography in 20 mM Tris pH 7.5, 150 mM NaCl buffer on a Superdex 200 16/600 (Cytiva) column. For purification of *B. fragilis* SodFM2 and its mutants, 20mM Tris pH 8.5 was used. The *E. coli* SodFM1 (SodA) was collected in the flowthrough during AEC and was subsequently subjected to additional cation exchange chromatography (5 mL SP FF column (Cytiva) in 20 mM MES pH 5.5, 0-1M NaCl linear gradient elution), where it was also collected in the flowthrough. Metal content of protein extraction buffers as well as metal-loading of the purified protein preparations were verified using ICP-MS.

Analytical SEC using Superdex 200 10/300 increase column (Cytiva) was performed to estimate the molecular weight of protein preparations in non-denaturating conditions based on elution volume. Column was calibrated using mixture of molecular weight standards ranging from 1,350 to 670,000 Da (Bio-Rad) and blue dextran (2,000 kDa, Sigma) in 20 mM Tris pH 7.5, 150 mM NaCl buffer.

SOD activity assays and (a)CR calculations

SOD activity was assessed quantitatively using the riboflavin-NBT liquid assay[11] with minor modifications. The assay was performed in 96-well plates where 20 μL of a sample was mixed in each well with 180 μL assay solution (10 mM methionine, 1.4 μM riboflavin, 66 μM NBT, and 10 μM EDTA in 50 mM potassium phosphate buffer, pH 7.8), immediately prior to 10 min incubation on a white light box followed by immediate measurement of absorbance at 560 nm. Serial 2-fold dilutions of tested samples were assayed to identify enzyme concentrations within the linear range of the assay that were close to 1 U in reference to a standard curve obtained using commercial bovine SOD standard (S5639, Sigma). To calculate specific SOD activities, the assay was performed in triplicate using samples diluted to the concentrations within the assay's linear range. Cambialism ratios (CRs) were calculated as a ratio of SOD activities of the verified Fe-loaded protein preparation divided by that of the verified Mn-loaded protein preparation. Approximate cambialism ratios were calculated the same way but using SOD activity values of soluble protein extract samples from the 24-well culture experiments after lysis and centrifugation. CR values > 0.5 were considered as evidence of higher Fe-preference, CR < 0.5 as evidence of higher Mn-preference, and enzymes with 0.5 < CRs < 2 were considered cambialistic.

For qualitative SOD activity assessment, approximately 15 µg (BCA assay) of soluble protein extracts were separated on 15 % native PAGE gels and assayed using an in-gel assay as described previously[12]. The in-gel peroxide inhibition assay was performed as described previously[12]. For protein staining, samples were resolved on 15 % (w/v) SDS-PAGE gels and stained with Quick Coomassie (Protein Ark). All gels were imaged using a ChemiDoc™ imaging system (Bio-Rad), using the same settings (including exposure time) for all gels compared within a single experiment. Band intensity was quantified with Image Lab 6.1 software (Bio-Rad) using a rectangle volume tool to outline activity bands with local background subtraction. Approximate cambialism ratios (aCR) from the in-gel assay were calculated using the measured band intensities following the formula: aCR=(intFe - intFeH_2O_2)/IntMnH_2O_2, where intFeH_2O_2 is band intensity of the Fe-loaded form following peroxide inhibition, intFe refers to the band intensity of Fe samples in non-peroxide-treated gels, and IntMnH_2O_2 is the band intensity of the Mn-loaded form following peroxide inhibition. Data available in Source Data Files S8, S9.

Metal concentration verification using ICP-MS

Aliquots of protein extraction buffers or purified protein samples (20 µM) in 20 mM Tris pH 7.5, 150 mM NaCl were each diluted to 5 mL with 2% HNO_3 for elemental analysis. Elemental composition of the resulting acid solutions were quantified using an iCAP RQ ICP-MS instrument (Thermo Fisher Scientific) according to the manufacturer's specifications through comparison with matrix-matched elemental standard solutions, using In and Ir (10 µg/L) as internal standards (University of Plymouth Enterprise Ltd).

Protein crystallography

Purified and concentrated protein samples were subjected to crystallisation with commercially available matrix screens: PACT, JCSG+, Structure, Morpheus (Molecular Dimensions) and Index (Hampton Research) in 96-well MRC crystallization plates (Molecular Dimensions) using a Mosquito liquid handling robot (TTP Labtech), with the sitting drop vapour-diffusion method at 20 °C. CRP-His Wolfebacteria SODFM1 crystallised in 100 mM sodium acetate trihydrate pH 4.5 and 25% w/v polyethylene glycol 3350; CRP-His Wolfebacteria SODFM1 mutant crystallised in 200 mM sodium chloride, 100 mM sodium acetate pH 5.0 and 20 % w/v polyethylene glycol 6000; B. fragilis SODFM2 crystallised in 200 mM sodium chloride, 100 mM Tris pH 8.5 and 25 % w/v polyethylene glycol 3350; and B. fragilis SODFM2 mutant crystallised in 10 mM zinc chloride, 100 mM sodium acetate pH 5.0 and 20% w/v polyethylene glycol 6000. All crystals were cryo-protected with the addition of 20% polyethylene glycol 400 to the crystalisation condition. Diffraction data were collected on the I03 and I24 beamlines at the Diamond Light Source Synchrotron (Didcot, UK) at 100 K. Data processing and refinement statistics are given in Source Data Table 6. Data were indexed and integrated with the pipeline Xia2[13] with either 3dii XDS[14] or DIALS[15] and scaled with Aimless[16]. Space group determination was confirmed with Pointless[17]. Five percent of observations were randomly selected for the Rfree set. The phase problem was solved by molecular replacement using Phaser[18]. The 1UES was used as a search model in molecular replacement for B. fragilis SODFM2 structures, whereas 1IX9 and 1GV3 were used as search models in molecular replacement for CPR-His SODFM1 and its mutant, respectively. Initial model building was done with CCP4build task on CCP4cloud[19]. The model were refined with Refmac[20] and manual model building was done with Coot[21]. Models were validated with Molprobity[22] and Coot. Structural representations were prepared with PyMOL (Schrödinger, LLC).

Structural comparison was performed for global secondary structure matching (SSM) in COOT and for least square comparison (LSQ-kab) of metal-binding residues in CCP4i. To visualise and compare electron density between structures the 2Fo-Fc map was displayed at 1σ level.

Bioinformatic protein family identification and characterisation

Our previously described dataset of genomes sampled across the tree of life[12] was expanded with additional Bacteroidales and Alphaproteobacteria genomes in order to improve sampling of the lineages identified as metal-preference switching hotspots. The final dataset contained 3058 genomes including 2613 Bacteria, 281 Archaea, and 146 Eukaryotes (Full list in Source Data File S4). Members of the SOD protein families were identified using PFAM[23] hidden Markov model (HMM) profiles (Sod_Fe_C, PF02777.21; Sod_Fe_N, PF00081.25; Sod_Cu, PF00080.23; SodNi, PF09055.14) as queries against a local database of 10,304,216 protein sequences encoded within all 3058 analysed genomes in hmmsearch (-E 1e-5) profile search implemented in

HMMER3.3[24]. SOR protein family members were identified with hmmersearch using PFAM[23] profiles (Ferric_reduct, PF01794; and Desulfoferrodox, PF01880), as well as BlastP (E-value threshold = 0.005)[3] searches using sequences of structurally characterised SORs (PDB: 1DO6, and 1DFX).

The identified 2820 Fe/Mn SOD homologues (including 116 sequences with solved crystal structures available in the PDB) were aligned in MAFFT[25] with a set gap opening threshold (--op 2), and trimmed with trimAL[26] with a set gap threshold (-gt 0.1) to a final alignment length of 244 positions (Source Data File S5). The Fe/Mn SOD phylogeny (Source Data File S6) was generated with 1000 ultrafast bootstrap[27] replicates in IQ-TREE[28], with the best fitting phylogenetic model WAG+R10 selected with ModelFinder[29] implemented in IQ-TREE. Available solved Fe/Mn SOD crystal structures (listed in Source Data Table 1-5) were downloaded from the PDB[30]; aligned, manually inspected, and displayed in PyMOL (Schrödinger, LLC.); protein 3D structural comparisons were performed using the 'All against all' function implemented in DaliLiteV5.1 on a local workstation[31]. Groups of correlated amino acid residues within Fe/Mn SOD alignments were identified using an amino acid correlation algorithm implemented in PFstats[32].

Generation of the tree of life

HMM profiles were generated from a curated set of 21 single-copy orthologues shared between bacteria, archaea, and eukaryotes[33]. Sequences of the orthologues were downloaded from the supplementary material of the previous study of the universally conserved orthologues[33], aligned with MAFFT[25], trimmed with trimAL[26] (gappyout mode), and used for the generation of HMM profiles in HMMER3.3[24]. The profiles were subsequently used in hmmsearch (-E 1e-5) profile search[24] against the local database of 10,304,216 protein sequences encoded within 3058 genomes sampled from across the three domains of life. Each of the identified orthologue groups were aligned with MAFFT[25], trimmed with trimAL[26] (gappyout mode), and used for the generation of phylogenies in IQ-TREE [28] under LG+G4 model with 1000 ultrafast bootstraps[27]. All alignments and phylogenies were manually inspected, Eukaryotic organellar sequences were filtered out based on the tree topologies and verified with BLASTP searches against the nr database, and seven widely distributed orthologous groups were selected for the final analysis. The final dataset contained the following Eukaryotic (Prokaryotic) orthologues: Ribosomal 40S S3 (Ribosomal 30S S3), Ribosomal 40S S15 (Ribosomal 30S S19), Ribosomal 60S L23 (Ribosomal 50S L14), Ribosomal 40S S11/14 (Ribosomal 30S S11), Ribosomal 40S S5/S7 (Ribosomal 30S S7), Ribosomal 40S S0/S2 (Ribosomal 30S S2), and RNA polymerase II subunit 3 (DNA-directed RNA polymerase subunit D). The trimmed alignments of the seven orthologues groups were concatenated with module Nexus of the Biopython package[34] to a final alignment of 1,097 amino acid sites, and was used to generate the tree of life (Source Data File S7) with 1000 ultrafast bootstrap[27] replicates in IQ-TREE[28] under LG+G4 model.

Other bioinformatics methods

Mean sequence identities were calculated as an average of trimAL[26] pairwise identity comparisons (-sident) on MAFFT[25] alignments trimmed with trimAL[26] (gappyout mode). Sequence logos were generated with WebLogo3[35] and displayed so that the total height of each residue corresponds to the information content of the residue in the given position, and the relative size of multiple letters within each position corresponds to their relative frequency at this position. The phylogenetic trees were visualized and annotated with the protein presence/absence in the analysed genomes or with coevolving amino acid residues using GraPhlAn[36]. Multiple sequence alignments were visualized and annotated using Jalview[37]. Phylogenetic trees were visualized and annotated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/), Archaeopteryx[38].

References

1 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).

2 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008, doi:10.1038/msb4100050 (2006).

3 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410, doi:10.1016/s0022-2836(05)80360-2 (1990).

4 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).

5 Almagro Armenteros, J. J. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* **2**, doi:10.26508/lsa.201900429 (2019).

6 Emanuelsson, O., Nielsen, H. & von Heijne, G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**, 978-984, doi:10.1110/ps.8.5.978 (1999).

7 Kall, L., Krogh, A. & Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* **35**, W429-432, doi:10.1093/nar/gkm256 (2007).

8 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).

9 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).

10 Rabe, B. A. & Cepko, C., doi:10.1101/2020.06.14.150979 (2020).

11 Beauchamp, C. & Fridovich, I. Superoxide dismutase: Improved assays and an assay applicable to acrylamide gels. *Analytical Biochemistry* **44**, 276-287, doi:10.1016/0003-2697(71)90370-8 (1971).

12 Barwinska-Sendra, A. *et al.* An evolutionary path to altered cofactor specificity in a metalloenzyme. *Nat Commun* **11**, 2738, doi:10.1038/s41467-020-16478-0 (2020).

13 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190, doi:10.1107/s0021889809045701 (2009).

14 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).

15 Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85-97, doi:10.1107/S2059798317017235 (2018).

16 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).

17 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282-292, doi:10.1107/S090744491003982X (2011).

18 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

19 Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. Distributed computing for macromolecular crystallography. *Acta Crystallogr D Struct Biol* **74**, 143-151, doi:10.1107/S2059798317014565 (2018).

20 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255, doi:10.1107/S0907444996012255 (1997).

21 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

22 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).

23 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419, doi:10.1093/nar/gkaa913 (2021).

24 Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195, doi:10.1371/journal.pcbi.1002195 (2011).

25 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
26 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).
27 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522, doi:10.1093/molbev/msx281 (2018).
28 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).
29 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).
30 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).
31 Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29-42, doi:10.1007/978-1-0716-0270-6_3 (2020).
32 Fonseca-Junior, N. J., Afonso, M. Q. L., Oliveira, L. C. & Bleicher, L. PFstats: A Network-Based Open Tool for Protein Family Analysis. *J Comput Biol* **25**, 480-486, doi:10.1089/cmb.2017.0181 (2018).
33 Williams, T. A., Cox, C. J., Foster, P. G., Szollosi, G. J. & Embley, T. M. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* **4**, 138-147, doi:10.1038/s41559-019-1040-x (2020).
34 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423, doi:10.1093/bioinformatics/btp163 (2009).
35 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
36 Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029, doi:10.7717/peerj.1029 (2015).
37 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).
38 Han, M. V. & Zmasek, C. M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**, 356, doi:10.1186/1471-2105-10-356 (2009).

Extended data legends

Extended Data Fig. 1 Fe/Mn SODs constitute the largest and most widely distributed family of superoxide-detoxifying enzymes

a The coloured semi-circles represent the distribution of Fe/Mn SODs (magenta), Cu/Zn SODs (green), and Ni SODs (red), mapped onto a phylogeny of 146 Eukaryotes, 2577 Bacteria, and 288 Archaea. SORs (superoxide reductases, orange) were often found in the genomes that lack any detectable SOD sequences (SOD-negative, grey; and Source Data Tab.7): in Archaea, including representatives from Euryarchaeota, Proteoarchaeota, TACK, and DPANN Archaea; Bacteria, with notable examples from classes Erysipelotrichia and Negativicutes within phylum Firmicutes; and Eukaryotic unicellular microbes *Spironucelus salmonicida* and *Giardia intestinalis*, a well-studied example of a SOR-only eukaryotes (DOI: 10.1016/j.bbapap.2009.10.011, 10.1016/j.freeradbiomed.2011.07.017, 10.1371/journal.pgen.1004053). Neither of the canonical superoxide-scavenging enzymes were detected in 389 out of 3011 analysed genomes (SOD/SOR-negative, black). This group included members of the phylum Tenericutes (highlighted in grey within Firmicutes), such as the human pathogen *Mycoplasma pneumoniae*, which provides a known example of an aerobic organism without any identifiable SOD homologues encoded in its genome and without any measurable SOD activity in protein extracts (DOI: 10.1128/jb.146.1.222-232.19, 10.1016/0006-291X(80)91186-9). **B** Table displaying the number of the identified protein sequences, their mean sequence identity, and the number of genomes encoding at least one representative of each of the protein families. Fe/Mn SODs constitute the most highly conserved (35.7% mean sequence identity), and the most common (found in 67.4% of all analysed genomes) family of superoxide-detoxifying enzymes found in all three domains of life (Eukaryotes, Bacteria, Archaea). **C** Each of the analysed protein families (Fe/Mn SOD, CuSOD, NiSOD, and SOR) utilises different metal cofactors and is structurally distinct, consistent with independent evolution of the capacity for superoxide detoxification within evolutionarily distinct protein folds.

Extended Data Fig. 2 Fe/Mn SOD protein family can be subdivided into 5 distinct subfamilies based on phylogenetics, protein sequence, and 3D structure comparative analyses

a. The rings, coloured to represent the presence of either of the three alternative water-coordinating residues, $H_{Cterm}/Q_{Cterm}/Q_{Nterm}$, within SOD catalytic centres (three innermost rings) and the oligomerisation state identified in available crystal structures (two outermost rings), were mapped onto the maximum likelihood protein tree of all Fe/Mn SODs identified in our sampling of genomes from across the tree of life (Extended Data Fig. 1). N-terminal Gln (Q_{Nterm}: blue) is exclusively found in dimeric (green) SodFM2s, C-terminal water-coordinating Gln (Q_{Cterm}: orange) is found mainly in dimeric SodFM1s and tetrameric (brown) SodFM5s, and C-terminal His (H_{Cterm}: red) is most commonly (Fig. 2b) found in tetrameric (purple) SodFM3s and SodFM4s. **b.** An unrooted version of the SodFM protein tree presented in (a) **c.** Each cartoon represents a structural alignment of monomers of all SodFM1-5s with an available crystal structure in the protein data bank (PDB). The architecture of the two N-terminal α-helices (blue/turquoise/green) within monomers distinguishes the dimeric SodFM1s and SodFM2s from the tetrameric SodFM3s, SodFM4s, and SodFM5s. **d.** The tree represents a structural comparison of SodFMs based on crystal structures available in the PDB, generated using a Distance matrix alignment (DALI). The DALI analysis reveals that SodFM subfamilies identified through phylogenetics (b) and amino acid sequence analysis (a) also form distinct groups when only their 3D structures are compared. **e.** Maximum likelihood phylogenetic protein tree generated using amino acid sequences of the Fe/Mn SODs with available crystal structures used in the DALI analysis (d).

Extended Data Fig. 3 Position and identity of the conserved water-coordinating residue is not the sole determinant of SodFM1-5 phylogenetic grouping

a. A key fragment of the multiple sequence alignment of representative amino acid sequences from all five SodFM subfamilies. The rectangles outline sequence regions that were removed in order to test the robustness of the SodFM1-5 groupings. This enabled confirmation that the distinct features of SodFM1-5s that are responsible for their grouping in protein phylogenies are encoded in sequence regions outside of the water coordinating Gln

or His (Q/H_{H20}) and their immediate surroundings (black), and of the sequence region corresponding to the variable structural feature within the first two N-terminal α-helices (green) that distinguishes dimeric and tetrameric SodFMs (Extended Data Fig. 2a). Arrowheads indicate metal-coordinating residues (turquoise), water-coordinating N-terminal Gln, Q_{Nterm} (blue); C-terminal Gln, Q_{Cterm} (orange); C-terminal His, H_{Cterm} (red). **b.** The protein tree of SodFMs generated using a protein sequence alignment containing all regions of interest is the same as that used in previous figures (Extended Data Fig. 2a and b). **c-e.** Protein phylogenies generated after the removal of the regions containing the water-coordinating residue (c), the region encompassing the tetramerisation interface within two N-terminal helices (d), or all these regions at the same time (e) from the amino acid alignments (a) used to generate the trees. The overall SodFM protein tree topology has been retained in all trees (b-e), indicating that SodFM1-5 subfamily sequence-determinants are encoded within regions outside of the most apparent group-specific features.

Extended Data Fig. 4 Signatures of metal-preference switching can be detected in amino acid correlation analysis of SodFM2 subfamily members

a. Residues identified in amino acid correlation analysis of SodFM2s were colour coded and mapped onto the SodFM2 protein phylogeny as separate semi-circles. Multiple amino acid correlation groups with distinct distribution patterns across the SodFM2 tree were identified. The alphaproteobacterial SodFM2.4s constituted the most distinct subgroup of SodFM2s, which grouped closely with eukaryotic chloroplast SodFM2.3s and Cyanobacterial SodFM2.2s. Alongside Bacteroidales SodFM2.6, SodFM2.4s also represented one of the two metal-preference switching evolutionary hotspots (black elliptical outlines) identified in our dataset. Characteristically in these two evolutionary hotspots, residues from the first correlation group (innermost magenta semi-circles) were often replaced with those from the second group (the innermost green semi-circle), one of which was the key metal-preference determinant X_{D-2} (Fig. 1c and 2c). The top ring (black) indicates SODs with available crystal structures. Species names and the outermost semi-circle (orange) indicated the SODs with aCR values experimentally verified in this study (Fig. 2a). **b.** Sequence logos of each SodFM2 sub-group, with amino acid correlation groups mapped onto the logos as coloured semi-circles. Within the analysed SodFM2 alignment residue X_{D-2} is at position 160. Black arrows indicated two highly distinct His residues identified in SodFM2.5, which contained many sequences from diverse Eukaryotic microbial pathogens including *Plasmodium falciparum, Toxoplasma gondii, Trypanosoma brucei,* and *Cryptosporidium parvum.* Arrowheads under each logo signify metal coordinating residues (magenta), catalytic water coordinating Q_{Nterm} (orange), and the residue X_{D-2} (green).

Extended Data Fig. 5 Cambialism has evolved multiple independent times from two distinct SodFM subfamilies in Bacteroidales

a. An enlarged version of the species tree of Bacteroidales (Fig. 3a) extracted from the tree of life (Fig. 1a). Colour-coded circles represent the presence of SodFM1 (pink), SodFM2 (blue), CuSOD (yellow), NiSOD (green), or SOR (orange). Coloured rectangles display experimentally verified aCRs (Fig. 2a, and Extended Data Fig. 8) for SodFM1 and SodFM2s. Red dashed line (a) outlines the lineages with SodFM2s containing a hallmark of higher Mn-preference, residue G_{D-2}. Coloured triangles indicate inferred ancestral Fe/Camb- (red), or Mn/Camb-switching (blue), and likely lateral gene transfer (LGT, green) events. The likely source (dashed black line) and direction (arrowhead) of the LGTs were inferred based on the SodFM2 protein tree topology (b). Selected species names were colour-coded to enable comparisons between the species tree (a) and the protein trees (b and d). **b-d.** Enlarged fragments of the tree of all Fe/Mn SODs (Fig. 2a) containing Bacteroidia SodFM2s (b) and SodFM1s (c and d). Annotations correspond to those used in the species tree (A). Branches corresponding to SodFM1 sequences from classes of Bacteroidota and those from closely related Ignavibacteria were colour coded (c and d), whereas the black branches correspond to SodFM1s from other taxa. The regions of the SodFM1 tree containing Bacteroidia sequences of interest were outlined with green dashed lines (c), and their magnified versions were presented in (d).

Extended Data Fig. 6 Alphaproteobacteria-specific SodFM2.4 was acquired in their last common ancestor

a. An enlarged fragment of the species tree of Alphaproteobacteria (Fig. 3b) extracted from the tree of life (Fig. 1a). Colour-coded circles represent the presence of SodFM1 (pink), SodFM2.4 (light blue), a SodFM2 other than SodFM2.4 (dark blue), SodFM4 (red), CuSOD (yellow), NiSOD (green), or SOR (orange). Coloured rectangles

display experimentally verified aCRs for SodFM1 and SodFM2s (Fig. 2a, and Extended Data Fig. 8). Coloured triangles indicate inferred acquisitions of higher Mn-preference (red), or reversions back to the most likely ancestral Fe-preference (blue). The likely acquisition of the ancestral SodFM2.4 in the last common ancestor of the Alphaproteobacteria following the split from Ricketsiales was indicated (large pink triangle). Selected species names were colour-coded to enable comparisons between the species tree (a) and the protein tree (b). **b.** Zoomed in fragments of the tree of all Fe/Mn SODs (Fig. 2a) containing all SodFM2.4s. Annotations correspond to those used in the species tree (a). The letters mapped onto the protein tree (positions 147, 160, and 167; numbering as described in Extended Data Fig. 4) correspond to the SodFM2.4-specific residues identified in the amino acid correlation analysis (Extended Data Fig. 4a). The overall topology of the protein tree (b) resembles that of the species tree (a), consistent with ancient acquisition of a SodFM2.4 (large pink triangle, and Fig. 3F) with an increased Mn-preference followed by vertical inheritance. Grouping of the Fe-preferring SodFM2.4s (blue triangles) alongside the closely related canonical SodFM2.4s provides further support for the hypothesis that the Fe-preferring SodFM2.4s represent evolutionary reversion to their ancestral Fe-preference.

Extended Data Fig. 7 Combination of a standardised SOD activity assay with SodFM-free protein expression strain provides a reliable method for metal-preference determination
a. Examples of in-gel activity assay results (in-gel) for Fe-, Mn-, and cambialistic-SodFMs with common indicators of each metal-preference and resistance to peroxide (H_2O_2). **b.** In-gel SOD activity assay on total protein extracts from wild-type (WT) *E. coli* BL21(DE3) overexpressing SodFMs(*Ec*SodA and *Ec*SodB, and *Sa*SodA and *Sa*SodM) in the presence of Fe (red) or Mn (orange). Samples contain endogenous *Ec*SodA (turquoise arrowheads) and *Ec*SodB (purple) expressed from the WT genome, as well as numerous heterodimeric forms (magenta) assembled from monomers of *Ec*SodA, *Ec*SodB, and the overexpressed enzymes. **c,d.** To address the issue of the contaminating endogenous *E. coli* SodFMs, we generated *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* clean double knock out strain (ΔΔ*sodAB*)**.** No endogenous SodFM activity bands were observed in total protein extracts from ΔΔ*sodAB* (c, right lane) and in those from ΔΔ*sodAB* overexpressing SodFM isozymes (**d,** *E. coli* SodA and *S. aureus* SodM) in the presence of Fe (red) or Mn (orange), and a negative control transformed with empty plasmid (pET22a). **e.** In-gel activity assay on soluble protein extracts from SodFM (SaSodA, and SaSodM) expressing ΔΔsodAB cultured in the presence of varying metal concentrations added to the M9 medium indicates that metal supplementation is required to metal-load overexpressed SodFMs. **f.** Comparison of aCR values estimated from the liquid SOD assay on soluble protein extract (liquid aCR), aCR from in-gel band intensities using the same samples (in-gel aCR), and 'true' CR samples calculated using metal-verified protein preparations generated here (this study) or from the previously published data (literature data).

Extended Data Fig. 8 Investigating metal-preference and peroxide-sensitivity of natural SodFMs
a-h, In-gel (a, c, e, g, i) and liquid (b, d, f, h, j) activity assay on soluble protein extracts from *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* over-expressing SodFM1 and 5s (a, b), SodFM2s (c,d), SodFM3s (e,f), SodFM4s (g,h). i-l. Experiments performed on expanded sampling of Bacteroidia SodFM1s and SodFM2s (i,j), and alphaproteobacterial SodFM2s (k,l), of interest identified in the bioinformatics analyses of metal-preference evolution. Proteins were expressed following 4 h IPTG induction in M9 medium supplemented with either 200 µM Fe (red) or 1 mM Mn (orange). Band patterns can be interpreted following the examples presented in the Extended Data Fig. 7. Negative controls consisted of the expression strain transformed with an empty vector (pET22). Some SodFMs (grey) were expressed at higher levels in *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* transformed with pRARELysS plasmid over-night (O/N) at 16ºC. All in-gel activity assays (in-gel, and in-gel + H_2O_2) are native polyacrylamide gels, and all Coomassie blue-stained protein gels (SDS) are denaturing gels. The identity of the water coordinating residue (H_2O_{coord}) and that of the residue X_{D-2} for each of the tested enzymes were annotated under the gels. Liquid SOD activity assay (b, d, f, h, j) was performed on the samples presented in **(**a, c, e, g, i **)**. Bars represent average enzyme activities from assay performed in triplicate and are presented in a stacked format, therefore the total bar hight represents the sum of Fe- and Mn-activities, and the line separating the top bars (Mn activity, orange) from the bottom bars (Fe activity, red) can be interpreted as representing a relative proportion of the activity with each metal (cambialism ratio). Error bars represent standard deviations of three replicates. The presented values are expressed in U of SOD activity per mg of total protein measured using BCA assay.

Extended Data Fig. 9 Testing the effects of mutagenesis of key secondary coordination sphere residues on metal-preference of SodFMs

a-l, In-gel (a, c, e, g, i, k) and liquid (b, d, f, h, j, l) activity assay on soluble protein extracts from *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* over-expressing SodFM mutants and wild type (WT) controls. **a,b** *S. aureus* cambialistic SodFM1 (SodM), and its X_{D-2} mutants, and canonical *S. aureus* Mn-SodFM1 (SodA) representing likely ancestral state of the SodA/SodM last common ancestor. **c,d** *B. fragilis* cambialistic SodFM2, and its X_{D-2} mutants. The X_{D-2} residues substituted in *B. fragilis* and *S. aureus* SodFM mutants (a,d) were selected based on residues found in other natural enzymes at this position (Fig. 2b).

e-l, Mutants of: SodFM1s (e,f), SodFM2s (g,h), SodFM4s (i,j), SodFM3s (k,l).

m. Mutagenesis of SodFM2.4-specific residues (Extended Data Fig. 4) in *A. tumefaciens* Mn-SodFM2.4. The results were analysed and presented as described in Extended Data Fig. 8. Identity of the mutated residues were indicated above the gels.

Extended Data Fig. 10 Structural investigation of SodFM mutants with changed metal-preference

a-h. Overall architecture (a, c, e, g) and enlarged view of the catalytic centre (b, d, f, h) of *B. fragilis* wild type (WT) camb-SodFM2 (a, b), its Fe-G_{D-2}T,F_{D-1}C mutant (c, d), and CPR Wolfebacteria Fe-SodFM1 (e, f) and its camb-H/Q,V_{D-2}G mutant (g, h). The grey mesh surrounding stick representations of residues of interest correspond to 2Fo-Fc electron density map in the solved structures. Spheres represent iron (orange), manganese (blue), and water (red) molecules. *B. fragilis* structure was solved for double G_{D-2}T,F_{D-1}C mutant for consistency with previously solved structures of *S. aureus* SodFM mutants (PDB 6qv8, and 6qv9). The mutation F_{D-1}C had no apparent effect on metal-preference compared to that of the *B. fragilis* G_{D-2}T single mutant (Extended Data Fig. 9c,d). **i.** General secondary structure topology diagram of SodFMs displaying the position of the key residues involved in: metal coordination (H/D_{FM}, black), metal-preference determination X_{D-2} (yellow), and two alternative locations of catalytic water coordinating H/Q within either C-terminal (H_{Ctrem}/Q_{Cterm}, magenta) or N-terminal (Q_{Nterm}, blue) domains. Cylinders represent α-helices and arrows represent β-strands. **j.** Analytical size-exclusion chromatography of WT CPR Wolfebacteria Fe-SodFM1. The protein elution pattern (j) is consistent with the homodimeric architecture observed in its crystal structure (e) despite containing water-coordinating His residue common in homotetrameric SodFM3s/4s.

Source data information

Source Data File S1. Primer sequences used in this study.

Spreadsheet containing sequences of oligonucleotides used for generation and verification of BL21 Δ*sodA*Δ*sodB* mutant (sodAB KO BL21), and for site directed mutagenesis of SodFMs (SodFM mutagenesis primers).

Source Data File S2. SOD gBlocks synthetic constructs.

Spreadsheet containing codon-optimised nucleotide sequences of synthesised *sod* genes with overhangs for Gibson Assembly cloning into pET-22b(+) vector.

Source Data File S3. SODs gBlocks expressed amino acid sequences.

List of amino acid sequences of 82 wild type SodFMs investigated in this study. The list includes sequences of SodFMs which did not express in the tested conditions.

Source Data File S4. Species sampling across tree of life

Spreadsheet containing list of Bacteria, Archaea and Eukaryota genomes sampled across the tree of life.

Source Data File S5. All SodFMs MAFFT alignment trimal gt01 trimmed

Multiple sequence alignment of SodFM protein sequences used to generate the SodFM phylogeny (Source Data File S6).

Source Data File S6. SodFM WAGR10 IQ-TREE

The phylogeny of the SodFM protein family.

Source Data File S7. Tree of life LG gamma IQ-TREE 1000uF bootstrap

The tree of life.

Source Data File S8. SOD activity mutants

Spreadsheet containing results of the enzymatic activity assay of soluble cell extracts of *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* over-expressing mutants and wild type (WT) controls of SodFMs. The table includes approximate cambialism ratios (aCR) (raw data and data corrected for changes in expression level based on coomassie-stained band intensity analysis), quantitative analysis of metal preference changes, and statistical significance of change in activity between WTs and the mutants. Data presented on Extended Data Fig.9.

Source Data File S9. Wild type SOD activity table

Spreadsheet containing average and standard deviation of a triplicate of liquid activity assay of soluble cell extracts of *E. coli* BL21(DE3) Δ*sodA*Δ*sodB* over-expressing wild type enzymes, and resulting approximate cambialism ratio (aCR) across studied groups of SodFMs. Data presented on Extended Data Fig.8.

Source Data Table 1. SodFM1 PDB

Spreadsheet containing PDB ID list and details of crystal structures available in the PDB for SodFM1. Based on phylogenetics and protein sequence analyses including amino acid correlation analysis, the published structures (PDB IDs) and all SodFM1 (species IDs) studied in this work are categorised into subgroups.

Source Data Table 2. SodFM2 PDB

Spreadsheet containing PDB ID list and details of crystal structures available in the PDB for SodFM2. Based on phylogenetics and protein sequence analyses including amino acid correlation analysis, the published structures (PDB IDs) and all SodFM2 (species IDs) studied in this work are categorised into subgroups.

Source Data Table 3. SodFM3 PDB

Spreadsheet containing PDB ID list and details of crystal structures available in the PDB for SodFM3. Based on phylogenetics and protein sequence analyses including amino acid correlation analysis, the published structures (PDB IDs) and all SodFM3 (species IDs) studied in this work are categorised into subgroups.

Source Data Table 4. SodFM4 PDB

Spreadsheet containing PDB ID list and details of crystal structures available in the PDB for SodFM4. Based on phylogenetics and protein sequence analyses including amino acid correlation analysis, the published structures (PDB IDs) and all SodFM4 (species IDs) studied in this work are categorised into subgroups.

Source Data Table 5. SodFM5 PDB

Spreadsheet containing PDB ID list and details of crystal structures available in the PDB for SodFM5 and list of all SodFM5 (species IDs) studied in this work.

Source Data Table 6. X-ray data collection and refinement statistics

Crystallographic table with details of data processing and refinement, for all presented structural models.

Source Data Table 7. SodCu, Ni, FM, SOR

Spreadsheet containing number of SodFM, CuSOD, NiSOD and SOR homologues identified in the analysed Eykaryota, Archaea, and Bacteria genomes; as well as lists of genomes lacking any SODs, and those lacking SORs and SODs.

Source Data Table 8. SodFM expansion

Spreadsheet containing list of genomes with numbers of identified SodFM homologues, categorised by SodFM subfamilies 1-4.