ARTICLES
https://doi.org/10.1038/s41589-021-00966-5
nature chemical biology

Check for updates

OPEN

Widespread occurrence of covalent lysine–cysteine redox switches in proteins

Fabian Rabe von Pappenheim[1,2], Marie Wensien[1,2], Jin Ye[3], Jon Uranga[3], Iker Irisarri[4,5], Jan de Vries[4,5], Lisa-Marie Funk[1,2], Ricardo A. Mata[3] and Kai Tittmann[1,2] ✉

We recently reported the discovery of a lysine–cysteine redox switch in proteins with a covalent nitrogen–oxygen–sulfur (NOS) bridge. Here, a systematic survey of the whole protein structure database discloses that NOS bridges are ubiquitous redox switches in proteins of all domains of life and are found in diverse structural motifs and chemical variants. In several instances, lysines are observed in simultaneous linkage with two cysteines, forming a sulfur–oxygen–nitrogen–oxygen–sulfur (SONOS) bridge with a trivalent nitrogen, which constitutes an unusual native branching cross-link. In many proteins, the NOS switch contains a functionally essential lysine with direct roles in enzyme catalysis or binding of substrates, DNA or effectors, linking lysine chemistry and redox biology as a regulatory principle. NOS/SONOS switches are frequently found in proteins from human and plant pathogens, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and also in many human proteins with established roles in gene expression, redox signaling and homeostasis in physiological and pathophysiological conditions.

Reactive oxygen species (ROS) are central to redox signaling in all domains of life and critically control cell growth, development, metabolism, aging and the response to stress conditions, such as, for example, infection by pathogens[1–4]. At elevated levels, ROS induce oxidative stress that has been implicated in a myriad of pathologies, including cancer, neurodegenerative diseases, inflammation and autoimmune conditions[5,6]. The underlying molecular mechanisms of redox signaling and oxidative stress have been mostly attributed to chemical modifications of cysteine residues in redox-sensitive proteins with key reactions being disulfide bridge formation between two cysteines or oxidation, glutathionylation and nitrosylation of individual cysteines[7,8]. We have recently reported the discovery of an allosteric lysine–cysteine redox switch with a covalent NOS bridge that regulates the enzymatic activity in response to changing redox conditions (Fig. 1)[9]. In the redox-sensitive enzyme transaldolase from *Neisseria gonorrhoeae*, the side chains of neighboring residues Lys 8 and Cys 38 at the protein surface form an NOS cross-link under oxidizing conditions that leads to a loss of enzymatic activity through 'structural cross-talk' with the active site. Under reducing conditions, the cross-link is disengaged, and enzymatic activity is restored. It had remained unclear whether this lysine–cysteine cross-link is a widespread regulatory redox modification in proteins or a specific feature of the studied transaldolase protein[9].

Here, we systematically mine the available protein structure database for proteins with undetected lysine–cysteine cross-links and find hitherto unidentified NOS bridges in proteins from all domains of life with critical roles in central cellular functions, including metabolism, gene expression, signaling, the ubiquitin pathway, DNA repair and redox homeostasis. These findings have wide-ranging biological implications in the context of redox signaling, oxidative stress and many human disease states.

Results

Geometric properties of NOS bridges. Before we analyzed the deposited experimental protein structures in the protein database (www.rcsb.org) to search for potentially undetected covalent NOS bridges, we sought to define the geometric properties of covalent NOS bridges versus non-covalent hydrogen-bond interactions between cysteine and lysine residues to identify reliable criteria to discern between these scenarios. A survey of hydrogen-bond interactions of cysteines in protein structures disclosed mean interatomic S–N distances of 3.44 Å with lysine residues (cysteine as acceptor) and 3.75 Å when cysteine acts as a hydrogen-bond donor[10]. However, the closest possible physical encounter between the two side chains remained to be defined. To close this knowledge gap, we conducted quantum chemical calculations, where we analyzed the geometric properties of (1) the NOS bridge, (2) a hydrogen bond between lysine as amine (-NH_2) and cysteine (-SH) and (3) a hydrogen bond between protonated lysine (-NH_3^+) and cysteine (-SH). An extensive conformational search was performed for the two residues in all three scenarios at different α-carbon atom distances (Extended Data Fig. 1). Several theoretical methods (also considering different dielectric constants; Supplementary Fig. 1) provided a consistent picture, whereby all minima featuring an N–S distance below 3.1 Å exhibited a covalent NOS bridge. The covalently bound NOS conformers displayed a bimodal distribution of the N–S 1,3 distance, with two maxima in the 2.6- to 2.7-Å range. This is a purely steric effect resulting from the relative positioning of the sulfur atom to the lysine chain (Supplementary Fig. 2). Hydrogen-bonded lysine–cysteine pairs are formed at much larger distances, with minimal values hovering at 3.2 Å for a protonated lysine–cysteine interaction, increasing to 3.5 Å once a proton is removed. The dipolar, charged Lys(NH_3^+)/Cys(S^-) configuration was only obtained when applying a dielectric continuum with a dielectric constant $\varepsilon > 10$.

[1]Department of Molecular Enzymology, Göttingen Center of Molecular Biosciences, Georg-August University Göttingen, Göttingen, Germany. [2]Department of Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany. [3]Institute of Physical Chemistry, Georg-August University Göttingen, Göttingen, Germany. [4]Institute for Microbiology and Genetics & Göttingen Center of Molecular Biosciences, Georg-August University Göttingen, Göttingen, Germany. [5]Campus Institute Data Science, Georg-August University Göttingen, Göttingen, Germany. ✉e-mail: ktittma@gwdg.de

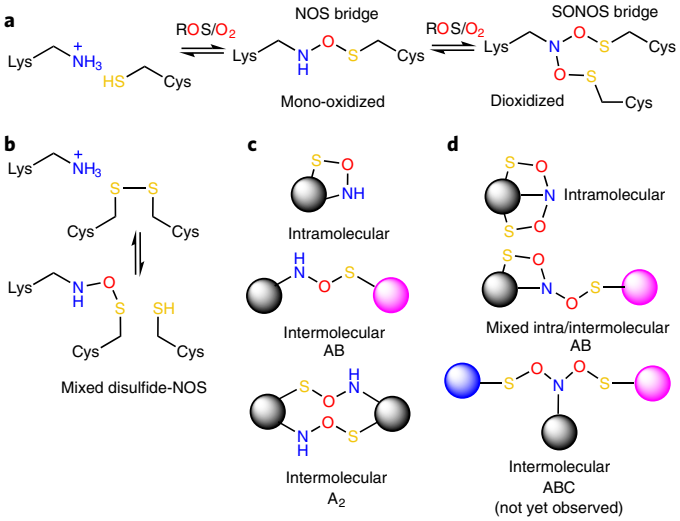


Fig. 1 | Chemical structures and topologies of NOS and SONOS redox switches in proteins. a, Structures and reaction scheme of NOS and SONOS redox bridge formation by ROS or oxygen in subsequent oxidation steps. **b**, Suggested structure of a 'mixed' NOS-disulfide redox switch, in which a disulfide is in equilibrium with an NOS bridge. **c**, Topologies of NOS bridges showing intramolecular and intermolecular cross-links as observed in experimentally determined protein structures. **d**, Topologies of SONOS bridges showing intramolecular and intermolecular cross-links as observed in experimentally determined protein structures. A SONOS bridge, where the lysine and two cysteines are contributed by three different proteins, has not been identified yet. The letters A, B and C indicate different proteins.

Hereby, the N–S distances are slightly reduced but only down to 3.05 Å (Supplementary Fig. 3). This state was, however, less stable than the charge-neutral Lys(NH_2)/Cys(SH) even up to dielectric constants mirroring water (within the limits of our model), leading to proton transfers. The computational results signal a clear difference between the various types of interactions, with the lowest energy distributions showing no overlap. The results are also consistent independent of the relative positioning of the two residues.

Detectability of NOS bridges in protein crystal structures. Next, we tested the impact of resolution and X-ray dose in X-ray crystallographic experiments on the traceability of covalent NOS bridges in protein structures (Extended Data Fig. 2). We had initially detected the NOS redox switch in protein crystals that diffracted beyond 1 Å, a resolution that is rarely accomplished in protein crystallography but is required to unambiguously define the chemical nature of atoms and their connectivities[11]. We used a subångström dataset obtained for one of these crystals and truncated the diffraction data at 1.0, 1.5, 2.0, 2.5 and 3.0 Å and further omitted the bridging oxygen atom of the NOS bridge for structure refinements to eliminate model bias (Extended Data Fig. 2a). Maps were calculated with resolution-specific B-factor blurring and simulated annealing of the NOS bridge lysine and cysteine residues. While a data truncation is physically not the same as obtaining datasets at different resolutions, this approach illustrates the general trends. As one would expect, the detectability of the covalent cross-link is critically dependent on the resolution: while the presence of the bridging atom is verifiable up to resolutions of 2 Å based on both the $2mF_o$–DF_c as well as mF_o–DF_c difference electron density maps (positive difference peaks), the interpretation of electron density maps at lower resolutions becomes ambiguous. In the transaldolase protein crystals we analyzed, the NOS bridge exhibits almost full occupancy, and our system may therefore be regarded as a best-case scenario (also considering the low B-factors). To test the impact of high-energy synchrotron radiation on the NOS bridge in protein crystals, which is known to promote photochemical reactions of redox-sensitive groups[12], we conducted dose-dependent experiments depositing doses between 0.27 and 5.4 MGy in 0.27-MGy increments (Extended Data Fig. 2b; X-ray crystallographic statistics in Supplementary Table 1). These experiments reveal that the NOS bridge in the tested protein crystals is relatively radiation hard and withstands doses up to 2.7 MGy, where radiation damage at neighboring acidic side chains becomes visible. When we deposited high doses of 5.4 MGy, we detected a slight decomposition of the NOS bridge. In sum, to reliably detect NOS bridges in protein crystal structures, a resolution of better than 2 Å is required as well as a high occupancy of this group and sufficient local order. Photochemistry during data collection is unlikely to cause NOS bridge formation artificially, as we have discussed before[9]; to the contrary, NOS bridges decompose while interacting with high-energy radiation but only at doses far higher than commonly used for data collection[12].

On the basis of the quantum chemical calculations and the X-ray crystallographic experiments on our model system, we downloaded all Protein Data Bank (PDB) entries with a resolution of ≤2 Å (65,000). We then searched for close contacts between lysine and cysteine residues, specifically pairs with interatomic N–S distances of less than 3 Å, which are in the regimen of covalent NOS bridges rather than hydrogen-bond interactions (>3.2 Å). We identified 266 deposited structures with ~400 potential NOS bridges (Supplementary Data 1) and manually inspected the putative lysine–cysteine cross-link site in all of the structures, for which structure factors and electron density maps were available. To eliminate a potential model bias, we also calculated mF_o–DF_c omit difference electron density maps, where the NOS bridge residues had been excluded from the structural models (representative examples are shown in Extended Data Fig. 3). The corresponding lysine and cysteine residues were modeled without a cross-link in almost all deposited structures, with a methylene bridge in very few structures and with an NOS link in one structure except our own structures. We remodeled both residues with a covalent cross-link either as an NOS bridge or, alternatively, as a direct N–S linkage (sulfenamide) using geometrically parametrized restraints obtained by our quantum chemical calculations. We consider formation of a methylene (N–CH_2–S) bridge as highly unlikely, as previously discussed by us and others[9,13]. We then re-refined the structures and compared the obtained models and electron density maps for the three alternative scenarios: (1) a covalent NOS linkage, (2) a covalent N–S linkage and (3) a non-covalent hydrogen-bond interaction (as deposited). In ~150 datasets, the existence of an NOS bridge is very likely (~100) or possible (~50), while a non-covalent hydrogen-bonding interaction or direct N–S linkage can be relegated in these datasets to a minor probability (Supplementary Data 1). In a few datasets, the electron density maps clearly indicate the presence of a cross-link, but the quality is insufficient to discriminate between an NOS bridge and a direct N–S bond (sulfenamide; Supplementary Data 1). Proteins likely or possibly containing NOS cross-links and corresponding homologs are found in all domains of life (Extended Data Fig. 4). Proteins from human or plant pathogens (Supplementary Table 2) and human proteins (Supplementary Table 3) are summarized individually.

Structural and chemical motifs of NOS bridges. NOS bridges are observed in an amazing diversity of structural motifs and chemical variations (Figs. 1 and 2). First, we analyzed the distance distribution in sequence between the NOS bridge lysine and cysteine residues –Cys^a–$(X)_{n-1}$–Lys^{a+n}–, where a and $a+n$ denote the corresponding sequence positions, and n denotes the sequence distance (Supplementary Fig. 4). This analysis revealed a Gauss-like distribution, with the largest fraction of NOS bridges observed for

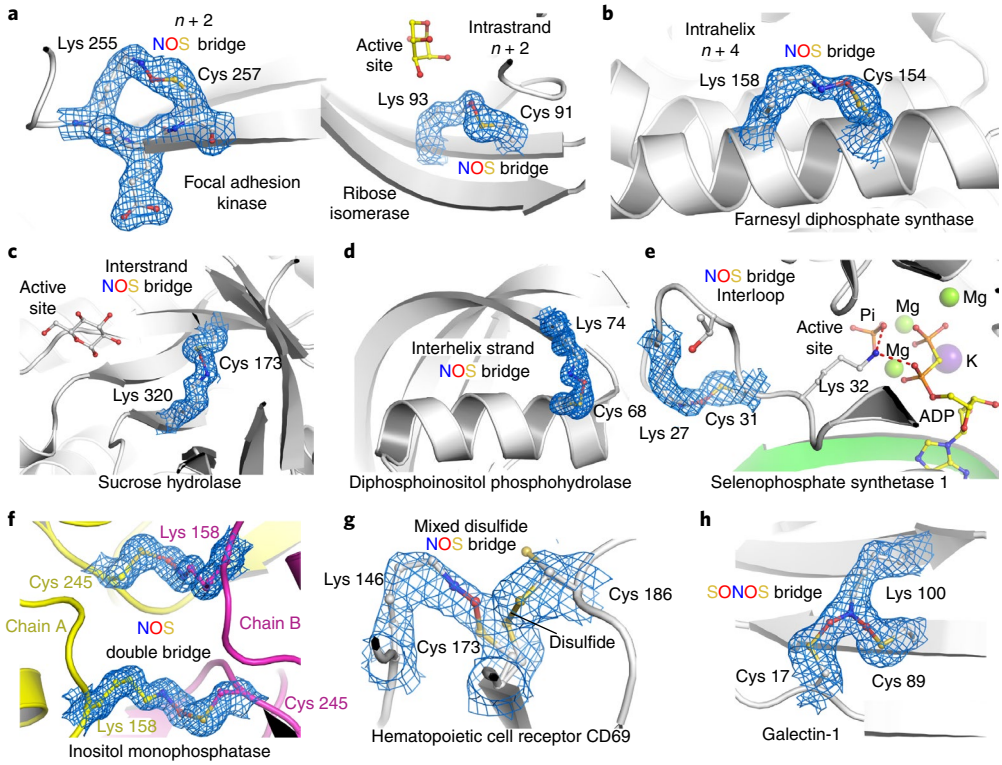


Fig. 2 | Structural and chemical motifs of NOS bridges in proteins. The corresponding $2mF_o$–DF_c electron density maps are shown in blue at a contour level of 1σ. **a**, Examples for NOS bridges in intrastrand or strand-like motifs with a sequence distance of $n+2$. Shown are focal adhesion kinase from *Gallus gallus* (PDB: 6CB0) and ribose isomerase from *Acinetobacter* sp. (PDB: 4Q0P). **b**, Example for an NOS bridge in intrahelix motifs with a sequence distance of $n+4$. Shown is the farnesyl diphosphate synthase from *Trypanosoma cruzi* (PDB: 6SDP). **c**, Example for an NOS bridge in interstrand (cross-strand) motifs. Shown is the sucrose hydrolase from *Xanthomonas axonopodis* (PDB: 3CZG). **d**, Example for an NOS bridge connecting a helix and a neighboring strand showing human diphosphoinositol phosphohydrolase (PDB: 6PCK). **e**, Example for an intraloop NOS bridge showing human selenophosphate synthetase 1 (PDB: 3FD5). **f**, Example for an intermolecular NOS double bridge between two chains in a homodimeric assembly. Shown is the inositol monophosphatase from *Medicago truncatula* (PDB: 5EQA). The two chains are colored individually in yellow and magenta, respectively. **g**, Example for a 'mixed' NOS-disulfide switch showing the human hematopoietic cell receptor CD69 (PDB: 1E8I, chain A) with an NOS bridge between Lys146 and Cys173 (30% occupancy) and a disulfide bridge between Cys173 and Cys186 (70% occupancy). **h**, Example for a SONOS bridge linking a lysine and two cysteines at the same time showing galectin-1 from *Rattus norvegicus* (PDB: 4GA9).

short sequence distances ($n<|10|$) and the smallest distance being two positions in sequence ($n+2$) in intrastrand/strand-like structures (Fig. 2a). Other motifs included intrahelical NOS bridges (mostly $n+4$; Fig. 2b), interstrand bridges in, for example, barrel structures (Fig. 2c), bridges connecting helices and strands (Fig. 2d) or intraloop bridges (Fig. 2e). Most NOS bridges are formed intramolecularly within one protein chain; in a few examples, we detected intermolecular cross-links, either as single bridges between two different proteins (A–B-type) or as an NOS 'double bridge' in homodimeric assemblies (A_2-type), where the corresponding lysine and cysteine residues of each chain form two reciprocal NOS bridges (Figs.1 and 2f). In one case, an NOS bridge was observed for a cysteine that is alternatively engaged in a disulfide bridge, suggesting a 'mixed' disulfide-NOS redox switch site (Fig. 2g). In terms of the oxidation state of the NOS cysteine, we exclusively found mono-oxidized species (sulfenic acid equivalent); there was not a single case where the cysteine sulfur was in a higher oxidation state (bonded to more than one oxygen atom). However, we discovered that some protein lysines may form NOS bridges with two cysteines at the same time, constituting a branching SONOS bridge in which the lysine nitrogen is in a dioxidized nitro state, as shown for, for example, galectin-1, a redox-sensitive protein with numerous important biological functions[14] (Figs. 1 and 2h and Supplementary Table 3). To the best of our knowledge, the SONOS group would be the first native branching cross-link between amino acids in proteins. Amino acid-derived redox cofactors, which are formed through cross-links between two and, in rare cases, three amino acids (for example, methionine–tyrosine–tryptophan) do not possess a single branching center as the lysine nitrogen atom of the SONOS group[15,16]. Other examples for proteins with SONOS bridges include the main protease (Mpro) from SARS-CoV-2 (Fig. 3), an important drug target for fighting the current coronavirus disease 2019 (COVID-19) pandemic[17,18], and the human NSF1–ISCU complex that is central to Fe–S cluster biogenesis in mitochondria[19] (Extended Data Fig. 5). In both cases, structures with either an NOS or a SONOS bridge of the same redox switch were detected, suggesting a stepwise oxidation (Fig. 1). While the initial oxidation (NOS) involved structurally proximal lysine and cysteine residues, the second oxidation (SONOS) involved, in both cases, a 'mobile' cysteine that structurally fluctuates between the active sites and the redox site (Fig. 3 and Extended Data Fig. 5).

Catalytic requirements for NOS/SONOS bridge formation. In our initial discovery of the allosteric NOS redox switch in the enzyme transaldolase from *N. gonorrhoeae*[9], we had speculated that a neighboring glutamate (Glu93) might be catalytically essential for formation of the NOS bridge, akin to the suggested role of a glutamate in isopeptide bond formation in some proteins[20] (Extended Data Fig. 6a). However, in many proteins with identified NOS or SONOS bridges, no amino acid capable of acid–base catalysis is found in

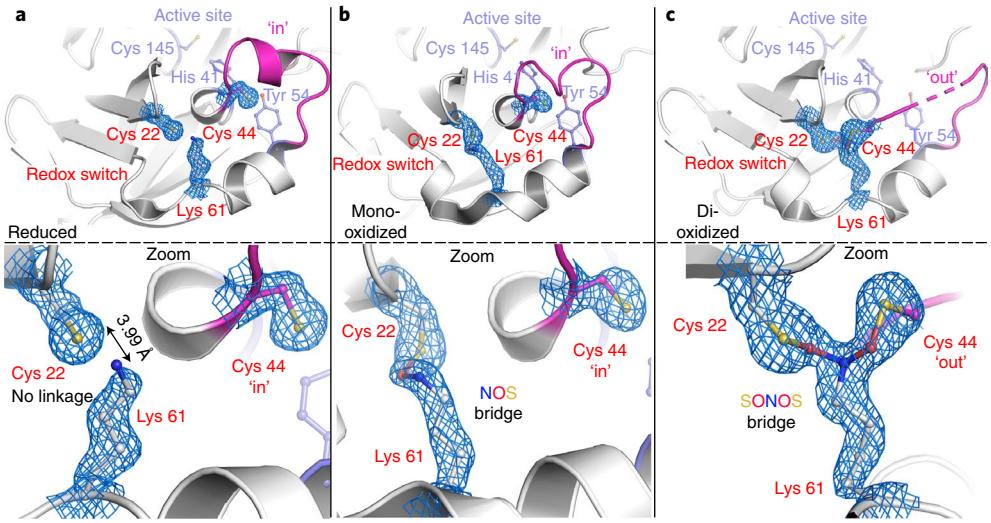


Fig. 3 | NOS and SONOS bridges in Mpro from SARS-CoV-2. The corresponding $2mF_o$–DF_c electron density maps are shown in blue at a contour level of 1σ. **a**, Structure of Mpro in the reduced state (PDB: 7JR3) showing the redox switch at the protein surface formed by residues Cys 22, Cys 44 and Lys 61 (highlighted in red) and the active site with residues Cys 145 (catalytic nucleophile), His 41 and Tyr 54 (highlighted in slate blue). A mobile loop bearing Cys 44 is indicated in magenta. Top, structural overview; bottom, close-up of the redox switch site. Note that residue Cys 44 is in the 'in conformation' and interacts with Tyr 54. **b**, Structure of Mpro in a mono-oxidized state with an NOS bridge formed between Cys 22 and Lys 61 (PDB: 6XMK). Top, structural overview; bottom, close-up of the redox switch site. Cys 44 is found in the 'in conformation'. **c**, Structure of Mpro in a dioxidized state with a SONOS bridge formed between Cys 22, Lys 61 and Cys 44 (PDB: 7JR4). Top, structural overview; bottom, close-up of the redox switch site. Cys 44 is found in the 'out conformation'. Competitive refinements (SONOS bridge only, two separate NOS bridges, mixture of SONOS and NOS) indicate full occupancy of the SONOS bridge.

the immediate vicinity of the redox switch, highlighting an intrinsic reactivity of structurally proximal lysine and cysteine residues, provided that ROS and/or oxygen is present. To test this hypothesis, we generated a transaldolase Glu93Gln variant in which a potential catalytic role of residue 93 is eliminated and analyzed the structure and enzymatic properties of the variant. The kinetic constants are very similar to the wild-type enzyme except for a slight 1.5-fold increase in the substrate Michaelis constant (Supplementary Table 4), and the variant remains redox activatable. Structural analysis of the Glu93Gln variant by X-ray crystallography (Supplementary Table 1) clearly indicates formation of the NOS bridge, arguing against a compulsory catalytic mode involving a neighboring amino acid (Extended Data Fig. 6b). The NOS cross-link is the only modification observable under non-reducing conditions, akin to our findings for the wild-type enzyme[9], strongly suggesting that this switch controls the enzymatic activity. The mechanistic proposal of a non-protein-catalyzed NOS bridge formation is also supported by the finding that a lysine–cysteine cross-link was unintentionally engineered into penicillin-binding protein 4 from *Staphylococcus aureus*, where the catalytic nucleophile Ser 95 at the active site had been replaced by a cysteine that forms a cross-link with the cocatalytic Lys 98 (Extended Data Fig. 6c–e); however, a sulfenamide species cannot be excluded in that case as, for example, reported for a related engineered protein[21]. Likewise, a positional inversion of the NOS bridge-forming lysine and cysteine residues in the human DNA repair enzyme OGG1 resulted in the formation of an NOS bridge akin to the wild-type configuration (Extended Data Fig. 6f–i). While amino acid or enzyme catalysts are not compulsory for NOS bridge formation in vitro, they could impact the kinetics of the reaction. Also, we cannot exclude enzyme or metal ion catalysis under in vivo conditions, as in the case of disulfide bonds[22].

Chemical functions of NOS lysine and cysteine residues. Next, we analyzed the putative chemical functions of the lysine and cysteine residues forming the NOS/SONOS redox switches in the context of the respective protein's structure and function. In the previously reported redox-sensitive transaldolase, the NOS bridge functions as an allosteric switch that changes the structure of the protein, including that of the active site, and thereby regulates enzymatic activity[9]. The operational modes of the NOS/SONOS redox switch residues in the now identified proteins are amazingly diverse and can be classified into several major classes, including lysines with direct catalytic roles, lysines with direct binding roles, cysteines with direct catalytic roles and allosteric switches (Fig. 4 and Extended Data Fig. 7). The manifold chemical tasks associated with the NOS lysine residues are particularly intriguing and include the full spectrum of lysine roles in protein function. NOS bridges were identified for proteins with catalytic lysines that form (1) Schiff bases with the pyridoxal phosphate (PLP) cofactor in PLP-dependent enzymes (for example, arginine/ornithine decarboxylase), (2) Schiff bases with enzymatic substrates (for example, KDPG aldolase), (3) carbamate intermediates in covalent CO_2 transfer to biotin (for example, oxaloacetate decarboxylase/Na^+ pump) and lysines that (4) act as acid–base catalysts in the active site of enzymes (for example, penicillin-binding protein; Fig. 4a–c and Extended Data Fig. 8a). Apart from these catalytic roles, numerous proteins were identified where the NOS lysine is directly involved in binding of enzymatic substrates (for example, DAH7P synthase) or of effectors, such as inositolphosphates (for example, rabphilin), often interacting with a negatively charged moiety of substrate or effector (Fig. 4d and Extended Data Fig. 8b). In other cases, the NOS bridge is located proximal to regulatory metal-binding sites (for example, Ca^{2+}; Extended Data Fig. 8c). An interesting observation concerns the involvement of the NOS bridge lysines in direct binding of DNA. In, for example, DNA polymerase, the lysine interacts with the base moiety of single-stranded DNA (Fig. 5a), while in homeobox protein Hox, it binds to the phosphate backbone of double-stranded DNA (Fig. 5b). Interestingly, NOS redox switches are also found for histone writers and erasers as well as for the transcription factor tubby, suggesting a critical and multilayered role of these switches

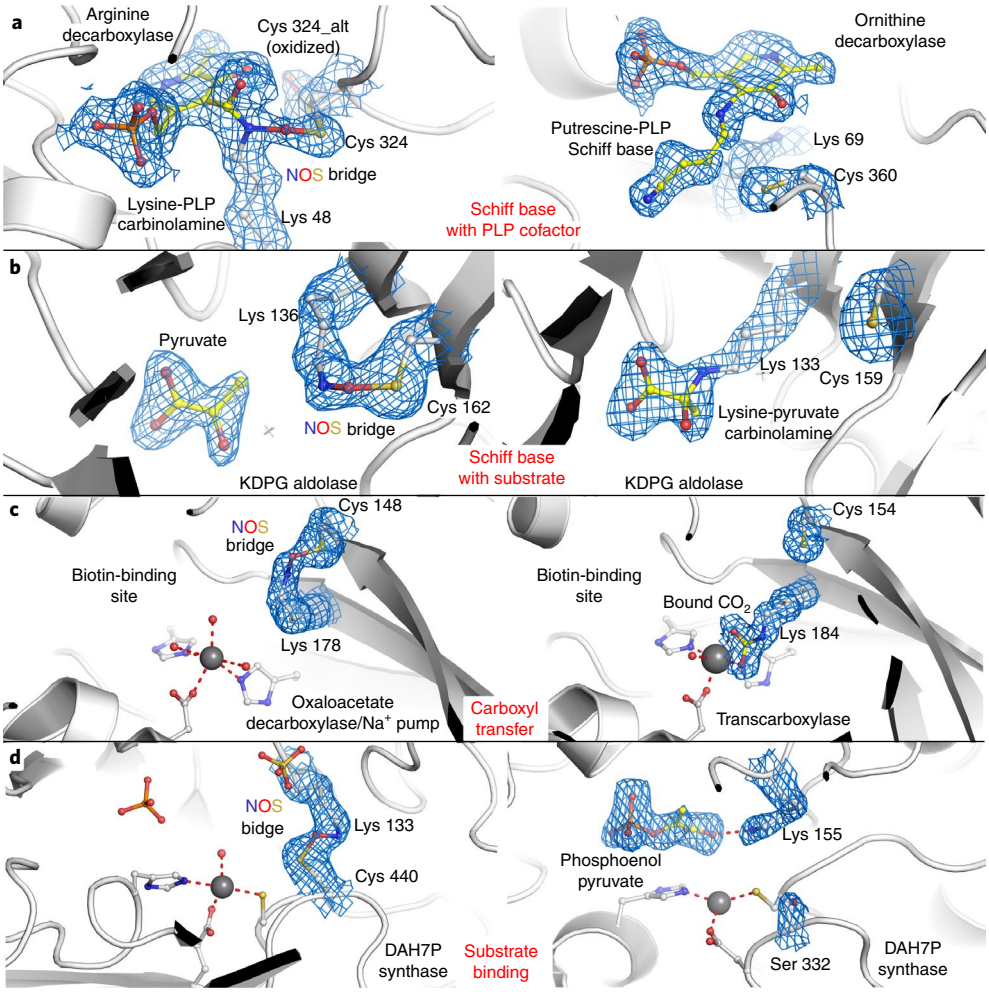


Fig. 4 | Functional roles of NOS bridge lysines in enzyme catalysis and in binding of enzymatic substrates or effectors. Left, proteins in the oxidized state with the NOS bridge present; right, same or closely related protein in the reduced state with the lysine exerting its function. The corresponding $2mF_o$–DF_c electron density maps are shown in blue at a contour level of 1σ. **a**, Catalytic lysines forming Schiff base intermediates in PLP-dependent enzymes; left, arginine decarboxylase from *Paramecium bursaria* chlorella virus (PDB: 2NV9), in which the catalytic Lys 48 forms an NOS bridge with Cys 324; right, ornithine decarboxylase from *T. brucei* in covalent complex with product putrescine (PDB: 1F3T). Note that in the presence of the NOS bridge, the reaction does not proceed beyond the carbinolamine. **b**, Catalytic lysines forming Schiff base intermediates with enzymatic substrates; left, KDPG aldolase from *Oleispira antarctica* in non-covalent complex with substrate pyruvate and with an NOS bridge between Lys 136 and Cys 162 (PDB: 3VCR); right, KDPG aldolase from *Escherichia coli* in covalent complex with substrate pyruvate (PDB: 1EUA). Note that in the presence of the NOS bridge, covalent catalysis is inhibited. **c**, Catalytic lysines in carboxyl transfer; left, oxaloacetate decarboxylase/Na^+ pump from *Vibrio cholerae* with an NOS bridge between Lys 178 and Cys 148 (PDB: 2NX9); right, transcarboxylase 5S subunit from *Propionibacterium freudenreichii* with carboxylated Lys 184 (PDB: 1RQB) thought to be an intermediate in CO_2 transfer to biotin. Note that in the presence of the NOS bridge, catalysis is inhibited. **d**, Lysines with roles in non-covalent binding of enzymatic substrates; left, DAH7P synthase from *Mycobacterium tuberculosis* (PDB: 3RZI) with an NOS bridge between Lys 133 and Cys 440; right, DAH7P synthase from *Listeria monocytogenes* in non-covalent complex with substrate phosphoenolpyruvate (PDB: 3TFC). This enzyme contains a serine (Ser 332) at the equivalent position of Cys 440 from *M. tuberculosis* DAH7P synthase and can therefore not form an NOS bridge. In the absence of the NOS bridge, the lysine forms a hydrogen bond with the carboxylate moiety of substrate phosphoenolpyruvate.

for the regulation of gene expression (Fig. 5c–e). Catalytic cysteines were detected for several ubiquitin E2 ligases (Extended Data Fig. 9) and the Fe–S cluster biogenesis complex ISCU–NSF1 (Extended Data Fig. 5), as discussed before.

Possible biological functions of NOS and SONOS switches. The putative NOS and SONOS bridges detected in the crystallographic protein structures were apparently formed due to the presence of oxygen. The responsible oxidants in vivo need be identified, but it seems reasonable to suggest that NOS/SONOS bridges form under oxidizing conditions as, for example, oxidative stress by ROS[1–6]. However, these switches might also function as sensors for oxygen. Some of the species with NOS bridge-containing proteins are facultative anaerobic organisms (*V. cholerae* and *N. gonorrhoeae*), and the adaption to anaerobic or aerobic conditions entails reprogramming on different cellular levels of metabolism, signaling, gene expression and so on. Also, interestingly, our analysis identified NOS bridges in histone demethylases, which were recently reported to be oxygen dependent in a HIF-independent manner[23].

Our previous studies had indicated that the NOS bridge is a reversible redox switch that prevents overoxidation of the involved cysteine residues, and a similar function can be envisaged for the here newly identified switches. Formation of NOS/SONOS bridges in proteins, where the lysine and cysteine residues have direct roles

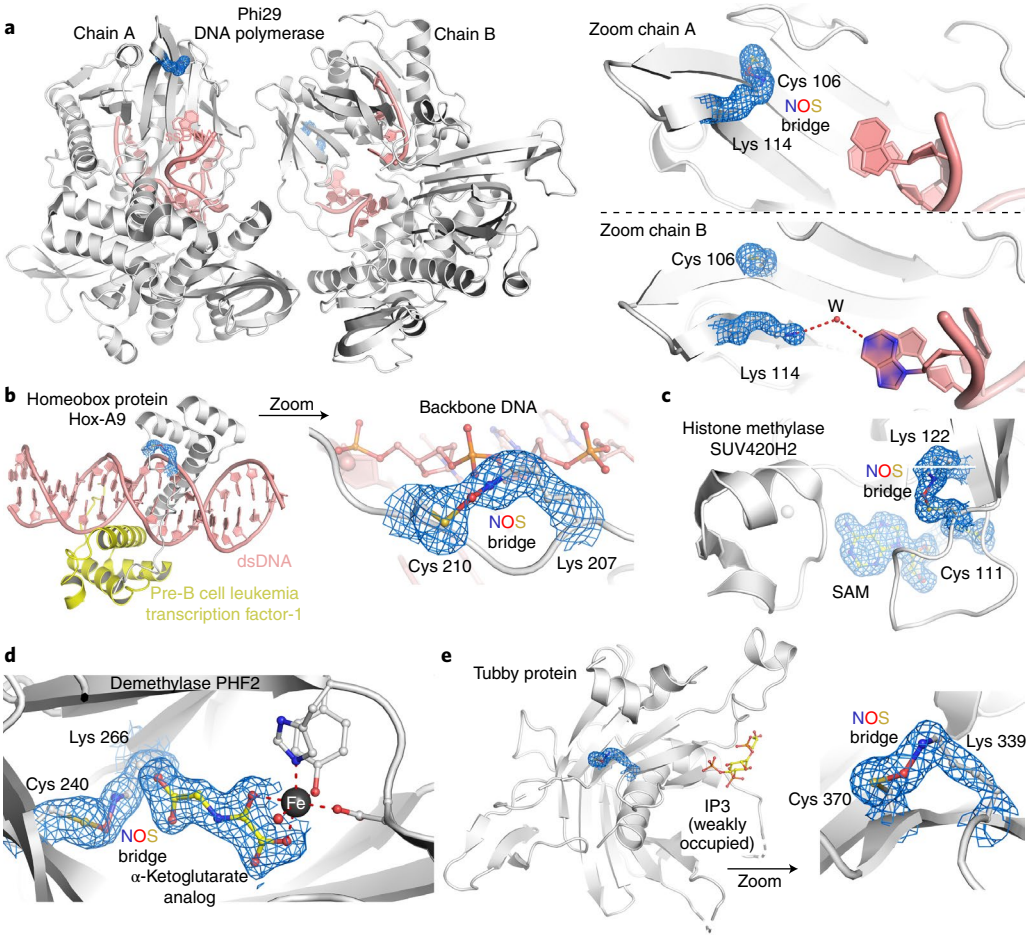


Fig. 5 | Proteins with NOS bridges inferred in DNA binding and transcription regulation. The corresponding $2mF_o$–DF_c electron density maps are shown in blue at a contour level of 1σ. **a**, DNA polymerase from *Bacillus* virus phi29 complexed with single-stranded DNA (PDB: 2PY5). Left, overall structure showing the two copies (chains A and B) in the asymmetric unit. The single-stranded DNA is highlighted in salmon red. Right, close-up of the NOS redox switch with residues Lys 114 and Cys 106. Note that in chain A, both residues form an NOS bridge (left), which is absent in chain B (right). In the absence of the NOS bridge, Lys 114 forms a hydrogen bond with a DNA base via a water molecule. **b**, Homeobox protein Hox-A9 from *Mus musculus* in complex with human pre-B cell leukemia transcription factor-1 and double-stranded DNA (dsDNA; PDB: 1PUF). Left, overall structure of the complex. The two proteins are colored individually, and the DNA is shown in salmon red. Right, close-up of the NOS bridge formed between Lys 207 and Cys 210 of Hox-A9. Note that the NOS bridge is at the binding interface with the backbone of the DNA, suggesting that Lys 207 in the reduced state interacts with the phosphate groups. **c**, Human histone-lysine *N*-methyltransferase SUV420H2 in complex with substrate *S*-adenosyl methionine (SAM) and with an NOS bridge formed between Lys 122 and Cys 111 (PDB: 3RQ4). Note the proximity of the NOS bridge with respect to the SAM-binding locale. **d**, Human demethylase PHF2 in complex with an analog of the α-ketoglutarate cofactor and with an NOS bridge formed between Lys 266 and Cys 240 (PDB: 3PU8). Note the proximity of the NOS bridge with respect to the cofactor-binding locale, suggesting a direct interaction of Lys 266 with the carboxylate moiety of the cofactor. **e**, Tubby protein from *Mus musculus* in complex with inositol trisphosphate (I3; PDB: 1I7E). Left, structural overview highlighting the ligand-binding site and the allosteric NOS switch. Right, close-up of the NOS bridge between Lys 339 and Cys 370.

in catalysis, will go along with a reversible loss of function. In case of direct roles in the binding of enzymatic substrates, effectors or DNA, NOS bridge formation would lead to a reduced affinity and with that a diminished biological activity, although the extent may vary depending on the structural context. For allosteric redox switches, both loss-of-function as well as gain-of-function scenarios seem possible with a variable modulation of biological function. In support of this suggested function, the available biochemical data for several key redox-sensitive proteins with identified NOS/SONOS redox switches bearing catalytic lysines or cysteines indeed indicate a reversible loss of function in vivo under oxidative stress conditions (reaction with oxidants hydrogen peroxide or diamide) that is related to cysteine oxidation, as in the case of, for example, DNA repair enzyme OGG1 (Schiff base lysine)[24], ornithine decarboxylase (catalytic lysine and cysteine)[25] or ubiquitin E2 ligases (catalytic cysteine)[26]. In the case of ornithine decarboxylase, a mutation of the NOS cysteine made the resultant protein variant redox insensitive, supporting a regulatory in vivo function. For numerous other cases, it has been established under in vitro conditions that either reducing conditions are required for full biological activity or that oxidizing conditions deteriorate protein activity[9,27–29]. In the case of galectin-1, which possesses an allosteric SONOS switch, it was established that the protein exerts different biological functions under oxidizing and reducing conditions in vivo and that binding of effectors requires the reduced state[30,31]. For the human protein, exactly these two cysteines engaged in forming the SONOS bridge in the *R. norvegicus* ortholog (Fig. 2h) were identified as the redox switch site and suggested to form a disulfide, although the latter could not be directly demonstrated[31]. The potential role of NOS redox switches in regulating DNA binding might provide a

mechanism for the well-established redox-dependent action of many transcription factors[8,32]. A putative regulatory role of the detected NOS/SONOS redox switches is further supported by the high degree of conservation of the lysine–cysteine pairs in the protein orthologs (either across different domains of life or in clades of related organisms) despite the fact that in many proteins, only one of the two residues has a direct functional role required for biological activity, as reported for, for example, OGG1 or rabphilin (Supplementary Figs. 5 and 6)[33,34].

NOS/SONOS bridges in proteins from pathogens and humans. We further classified the proteins with NOS/SONOS switches regarding their biological function. First, we analyzed proteins from human and plant pathogens, including viruses, bacteria and parasites (Supplementary Fig. 7 and Supplementary Table 2). In the case of pathogenic human viruses, NOS switch-containing proteins with essential roles for all critical phases of the virus life cycle, such as infection (adenovirus), replication (SARS-CoV-2) and maturation assembly (cytomegalovirus), were detected. In bacteria and parasites, we identified numerous essential biosynthetic enzymes (amino acid, isoprenoid and polyamine biosynthesis), Na^+ transporters required for ATP biosynthesis, transcription regulators, virulence factors and proteins in the context of pathogen–host interactions. Remarkably, numerous proteins originate from some of the world's most dangerous pathogens, including, among others, SARS-CoV-2 (COVID-19), *M. tuberculosis* (tuberculosis), *V. cholerae* (cholera), *Pseudomonas aerigunosa* (pneumonia), *S. aureus* (bacterial superinfections), *Legionella pneumophila* (Legionnaires' disease) or *Trypanosoma* species (Chagas disease, African sleeping sickness). NOS bridges were also discovered in major plant pathogens, including *X. campestris* and *X. axonopodis*, the causative agents of 'black rot' in cruciferous vegetables and bacterial pustule of soybean[35].

In humans, NOS redox switches are found in different protein families with established roles in the oxidative stress response, redox signaling and homeostasis (Supplementary Fig. 7 and Supplementary Table 3). Interestingly, many proteins are localized in the nucleus or membrane (Supplementary Data 1). Major families comprise proteins in DNA repair (excision of oxidized bases, cleavage of protein–DNA cross-links), in transcription regulation (sensing and binding of DNA, histone writers and erasers), in ribosomal protein translation (translation factors), in protein degradation (ubiquitin E2 ligases, E3 ring ligase), in signaling (various kinases, neurotransmission, inositolphosphate signaling) and in the biosynthesis of redox-sensitive cofactors and rare amino acids (Fe–S cluster, SAM and selenocysteine; see examples in Figs. 2 and 5 and Extended Data Figs. 5 and 9). Also, the enzyme selenophosphate synthetase 1 was identified (Fig. 2e), which is known to be an essential factor of cellular redox homeostasis by regulating the expression levels of, for example, glutaredoxin and glutathione transferase, which protect cells from oxidative stress[36]. Many of the identified human proteins have been inferred in severe diseases, including various cancers, Alzheimer's disease, Parkinson's disease, obesity, autoimmune diseases and others (Supplementary Table 3). The discovery of the naturally evolved NOS redox switches may therefore unlock new therapeutic directions in plenty of disease states by addressing or manipulating the switch site. Prominent examples are the Mpro from SARS-CoV-2 (refs. [17,18]) or human focal adhesion kinase, which are validated drug targets for treatment of COVID-19 and invasive cancers[37], respectively.

Discussion

NOS and SONOS redox switches are ubiquitous regulatory elements in proteins that reversibly alter protein function in response to changing redox conditions as, for example, under oxidative stress. The involvement of lysines with direct roles in enzyme catalysis and/or in binding of enzymatic substrates, nucleic acids and effectors expands the chemical repertoire of organisms to deal with changing redox conditions and constitutes a new general regulatory principle in biology. Apart from new directions in medical applications, the identified design principles of naturally evolved NOS/SONOS switches are likely to inspire peptide and protein design[38]; in particular, the newly found branching SONOS cross-link bears great potential. The potential existence of sulfenamide species as another form of lysine–cysteine redox switches needs to be further explored[21,39].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-021-00966-5.

Received: 17 June 2021; Accepted: 20 December 2021;
Published online: 14 February 2022

References

1. Schieber, M. & Chandel, N. S. ROS function in redox signaling and oxidative stress. *Curr. Biol.* **24**, R453–R462 (2014).
2. Sies, H. & Jones, D. P. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nat. Rev. Mol. Cell Biol.* **21**, 363–383 (2020).
3. Bazopoulou, D. et al. Developmental ROS individualizes organismal stress resistance and lifespan. *Nature* **576**, 301–305 (2019).
4. Dickinson, B. C. & Chang, C. J. Chemistry and biology of reactive oxygen species in signaling or stress responses. *Nat. Chem. Biol.* **7**, 504–511 (2011).
5. Sies, H., Berndt, C. & Jones, D. P. Oxidative stress. *Annu. Rev. Biochem.* **86**, 715–748 (2017).
6. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* **49**, 1603–1616 (2010).
7. Paulsen, C. E. & Carroll, K. S. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chem. Rev.* **113**, 4633–4679 (2013).
8. Choi, H. J. et al. Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105**, 103–113 (2001).
9. Wensien et al. A lysine–cysteine redox switch with an NOS bridge regulates enzyme function. *Nature* **593**, 460–464 (2021).
10. Zhou, P., Tian, F., Lv, F. & Shang, Z. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **76**, 151–163 (2009).
11. Neumann, P. & Tittmann, K. Marvels of enzyme catalysis at true atomic resolution: distortions, bond elongations, hidden flips, protonation states and atom identities. *Curr. Opin. Struct. Biol.* **29**, 122–133 (2014).
12. Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl Acad. Sci. USA* **103**, 4912–4917 (2006).
13. Wang, J. Crystallographic identification of spontaneous oxidation intermediates and products of protein sulfhydryl groups. *Protein Sci.* **28**, 472–477 (2019).
14. Camby, I., Le Mercier, M., Lefranc, F. & Kiss, R. Galectin-1: a small protein with major functions. *Glycobiology* **16**, 137–157 (2006).
15. Davidson, V. L. Protein-derived cofactors revisited: empowering amino acid residues with new functions. *Biochemistry* **57**, 3115–3125 (2018).
16. Ghiladi, R. A., Knudsen, G. M., Medzihradszky, K. F. & de Montellano, P. R. O. The Met–Tyr–Trp cross-link in *Mycobacterium tuberculosis* catalase-peroxidase (KatG): autocatalytic formation and effect on enzyme catalysis and spectroscopic properties. *J. Biol. Chem.* **280**, 22651–22663 (2005).
17. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).
18. Jin, Z. et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
19. Boniecki, M. T., Freibert, S. A., Mühlenhoff, U., Lill, R. & Cygler, M. Structure and functional dynamics of the mitochondrial Fe/S cluster synthesis complex. *Nat. Commun.* **8**, 1287 (2017).
20. Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**, 1625–1628 (2007).
21. Rodkey, E. A. et al. Crystal structure of a preacylation complex of the β-lactamase inhibitor sulbactam bound to a sulfenamide bond-containing thiol-β-lactamase. *J. Am. Chem. Soc.* **134**, 16798–16804 (2012).

22. Gross, E., Kastner, D. B., Kaiser, C. A. & Fass, D. Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell. *Cell* **117**, 601–610 (2004).
23. Chakraborty, A. A. et al. Histone demethylase KDM6A directly senses oxygen to control chromatin and cell fate. *Science* **363**, 1217–1222 (2019).
24. Bravard, A. et al. Redox regulation of human OGG1 activity in response to cellular oxidative stress. *Mol. Cell. Biol.* **26**, 7430–7436 (2006).
25. Coleman, C. S., Stanley, B. A. & Pegg, A. E. Effect of mutations at active site residues on the activity of ornithine decarboxylase and its inhibition by active site-directed irreversible inhibitors. *J. Biol. Chem.* **268**, 24572–24579 (1993).
26. Obin, M. et al. Redox regulation of ubiquitin-conjugating enzymes: mechanistic insights using the thiol-specific oxidant diamide. *FASEB J.* **12**, 561–569 (1998).
27. Veres, Z., Kim, I. Y., Scholz, T. D. & Stadtman, T. C. Selenophosphate synthetase. Enzyme properties and catalytic reaction. *J. Biol. Chem.* **269**, 10597–10603 (1994).
28. Webby, C. J., Baker, H. M., Lott, J. S., Baker, E. N. & Parker, E. J. The structure of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Mycobacterium tuberculosis* reveals a common catalytic scaffold and ancestry for type I and type II enzymes. *J. Mol. Biol.* **354**, 927–939 (2005).
29. Shah, R., Akella, R., Goldsmith, E. J. & Phillips, M. A. X-ray structure of *Paramecium bursaria* chlorella virus arginine decarboxylase: insight into the structural basis for substrate specificity. *Biochemistry* **46**, 2831–2841 (2007).
30. Than, N. G. et al. Emergence of hormonal and redox regulation of galectin-1 in placental mammals: implication in maternal–fetal immune tolerance. *Proc. Natl Acad. Sci. USA* **105**, 15819–15824 (2008).
31. Guardia, C. M. et al. Structural basis of redox-dependent modulation of galectin-1 dynamics and function. *Glycobiology* **24**, 428–441 (2014).
32. Abate, C., Patel, L., Rauscher, F. J. & Curran, T. Redox regulation of Fos and Jun DNA-binding activity in vitro. *Science* **249**, 1157–1161 (1990).
33. Wang, K., Maayah, M., Sweasy, J. B. & Alnajjar, K. S. The role of cysteines in the structure and function of OGG1. *J. Biol. Chem.* **296**, 100093 (2021).
34. Guillén, J. et al. Structural insights into the Ca^{2+} and PI(4, 5)P_2 binding modes of the C2 domains of rabphilin 3A and synaptotagmin 1. *Proc. Natl Acad. Sci. USA* **110**, 20503–20508 (2013).
35. da Silva, A. R. et al. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463 (2002).
36. Tobe, R. et al. Selenophosphate synthetase 1 is an essential protein with roles in regulation of redox homoeostasis in mammals. *Biochem. J.* **473**, 2141–2154 (2016).
37. McLean, G. W. et al. The role of focal-adhesion kinase in cancer—a new therapeutic opportunity. *Nat. Rev. Cancer* **5**, 505–515 (2005).
38. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
39. Gray, M. J., Li, Y., Leichert, L. I. O., Xu, Z. & Jakob, U. Does the transcription factor NemR use a regulatory sulfenamide bond to sense bleach? *Antioxid. Redox Signal.* **23**, 747–754 (2015).

Methods

Experimental procedures. *General information.* The protein concentration of *N. gonorrhoeae* transaldolase (*Ng*TAL) was determined by UV/Vis spectroscopy by measuring the absorption at a wavelength of 280 nm and using the molar extinction coefficient (ε_{NgTAL} = 28,420 M^{-1} cm^{-1}) determined according to Gill and Hippel[40].

Mutagenesis, protein expression and purification. For expression of *Ng*TAL (UniProtID: Q5F6E9), we used a pET SUMO vector, as recently described[9]. Mutant strains were generated by site-directed mutagenesis PCR using the QuikChange site-directed mutagenesis protocol (Stratagene). We used the following primer pair: *Ng*TAL variant Glu93Gln forward, 5′-GTCTGGCACAACATGAAAGCAC-3′, and reverse, 5′-GTGCTTTCATGTTGTGCCAGAC-3′. Expression and purification were performed as recently reported[9].

Steady-state kinetic analysis. Steady-state kinetic analysis of *Ng*TAL wild type was performed using a coupled enzymatic assay that monitors the conversion of ketose donor D-fructose-6-phosphate and aldose acceptor D-erythrose-4-phosphate into products sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate at a wavelength of 340 nm (refs. [41,42]). Measurements were conducted for both reducing and non-reducing conditions. For measurements under reducing conditions, the protein stock was supplemented with 20 mM dithiothreitol (DTT), resulting in a final concentration of 1 mM DTT in the assay mix. Initial rates were estimated by either linear regression of the absorbance signal over the first 5 s of the measurements or, if substrate activation was observed, by using equation (1):

$$A_{340}(t) = A_0 - \Delta ss \cdot t + \frac{\Delta ss - \Delta_0}{k_{obs}} \cdot \left[1 - exp^{(-k_{obs} \cdot t)}\right] \quad (1)$$

Here, A_0 denotes the starting absorbance at 340 nm, Δss denotes the absorbance change at established steady state (steady-state rate), Δ_0 denotes the absorbance change at $t = 0$ (initial rate) and k_{obs} denotes the first-order rate constant of activation for the transition into the fully active enzyme form.

Thus, obtained steady-state activities were analyzed using the Michaelis–Menten equation and Hill equation.

*Crystallization, X-ray data collection, processing and model building. Ng*TAL wild type and variant Glu93Gln were crystallized and cryoprotected, as detailed in ref. [9]. Diffraction data of single protein crystals were collected using synchrotron radiation at beamline P14 of DESY EMBL, Hamburg, Germany, at a wavelength of either 0.9763 Å (*Ng*TAL wild type, dose series) or 0.689 Å (*Ng*TAL variant Glu93Gln) at 100 K using an EIGER 16M detector. For crystals of *Ng*TAL wild type, we conducted a dose-series experiment, depositing-defined doses (0.27–5.4 MGy in 0.27-MGy increments over 360° using identical start positions; the diffraction-weighted doses were calculated using RADDOSE-3D[43]) per dataset. For data processing, the XDS package was used[44]. Subsequent refinement and model building was performed using PHENIX.REFINE[45] and COOT[46]. Phasing was performed by rigid body refinement using our previously determined *Ng*TAL structure (PDB: 6ZX4) as an initial model. For initial coordinate disturbance, truncated datasets were subjected to simulated annealing in PHENIX.REFINE using a value of 7,000 K as start_temperature. The resulting maps were blurred to simulate more realistic B-factors for the selected resolutions[47] (1.0: 15; 1.5: 25; 2.0: 30; 2.5: 40; 3.0: 65). The geometry of the structure was validated using MolProbity[48]. Representations of structures were prepared using PyMOL (Schrödinger, The PyMOL Molecular Graphics System, version 1.8.). The Ramachandran statistics are 98.85% in the favored and 1.15% in the allowed region for 7ODO, 98.85% in the favored and 1.15% in the allowed region for 7ODP, 98.85% in the favored and 1.15% in the allowed region for 7ODQ and 98.55% in the favored and 1.45% in the allowed region for 7OEY.

Computational methods. We performed a series of electronic structure calculations to sample the conformational space and the relative energy of a model system consisting of a single lysine and cysteine residues. The residues were truncated at the α-carbon, with the latter represented as a methyl group to reduce the influence of electrostatics in the terminal moieties, closer to what is to be expected from a protein backbone. Three different bonding situations were sampled. The first corresponds to a neutral lysine interacting with a cysteine (NHS). The second system considered was a protonated lysine with a cysteine (NHS^+). Both cases will build close contacts through hydrogen bonds, either with the lysine as donor (NHS^+) or the cysteine (NHS). In the last system, we considered the covalently bound residues, with the lysine nitrogen and the cysteine sulfur bonding with an intercalated oxygen atom (NOS).

In a first set of calculations, we studied the dependence of the N–S distances on the dielectric constant. To this purpose, a representative conformer for each of the systems was optimized in different solvents (vacuum: $\varepsilon = 0.0$; diethylether: $\varepsilon = 4.2$; acetone: $\varepsilon = 20.5$; methanol: $\varepsilon = 32.6$; 1,2-ethanediol: $\varepsilon = 40.2$; dimethylsulfoxide: $\varepsilon = 46.8$; water: $\varepsilon = 78.4$) using the SMD[49] solvation model. The software package Gaussian16-A.03 (Gaussian 16, revision C.01) was used for the aforementioned calculations, with the B3LYP-D3(BJ)/def2-SVP[50–53] level of theory. A constrained α-carbon distance of 10 Å was applied for all the calculations with different solvents. The results are provided in Supplementary Fig. 1. The only difference observed was a smaller difference between the distances observed in the NHS and NHS^+ systems. Considering the stable behavior of the N–S distances in different environments, the following calculations were performed in vacuum for simplicity and for a more straightforward generalization.

Three different α-carbon distances (6, 8 and 10 Å) between lysine and cysteine were applied for the extended sampling runs (the results are provided in Extended Data Fig. 1). Initial conformer sampling was performed using CREST-2.10.2 (ref. [54]), whereby the conformer optimization was conducted with xTB-6.4.0 (ref. [55]) using the semiempirical tight binding-based quantum chemistry method GFN2-xTB[56] with the 'extreme' optimization criteria ($E_{conv} = 5 \times 10^{-8}$ Hartree) and the 'NCI' option.

After initial global sampling at the semiempirical level, the structures were refined in two stages at the density functional theory level. For NOS, the collected GFN2-xTB conformer structures were further optimized using B3LYP-D3(BJ)/def2-SVP[50–53], again with the use of the Gaussian16-A.03 program package. Conformers with imaginary frequencies were directly abandoned and not used further for optimization or data statistics. For neutral and protonated hydrogen-bonded systems (NHS and NHS^+), the conformers were additionally filtered to guarantee the existence of hydrogen bonds and, thereby, afford the closest contacts possible. Criteria of a distance shorter than 2.5 Å between the hydrogen (H) and the hydrogen acceptor (A) and an angle ADH (D: hydrogen donor) smaller than 30° were taken. Determination of the hydrogen atom for the building of the hydrogen bond follows the shortest distance among the potential hydrogen bonds between the amine and thiol residues.

The conformations obtained from B3LYP-D3(BJ)/def2-SVP were further sorted according to their electronic energies from low to high. If the energy difference between two conformations was larger than 0.1 kcal mol^{-1} and the root mean square difference (r.m.s.d.) for the superimposed structures was larger than 0.125 Å (ref. [57]), the two conformations were considered to be unique and subject to further calculations. From this pool of structures, we considered the lowest-lying conformers (up to 2 kcal mol^{-1} in difference to the global minimum) and performed further optimizations at the B3LYP-D3(BJ)/def2-TZVPP level of theory. For both NHS and NHS^+ systems, the additional filter process to verify the existence of hydrogen bonds was again applied. The results of the first stage (B3LYP-D3(BJ)/def2-SVP) and the second stage (B3LYP-D3(BJ)/def2-TZVPP) of optimizations were used and displayed in Extended Data Fig. 1. An overview of different geometric values and their correlation for the NOS bond are provided in Supplementary Fig. 2.

In a further set of calculations, we investigated what could be the lowest energy conformer for both NHS and NHS^+ bearing an N–S distance of 2.7 Å. By comparing the results of this sampling, we were able to assess the energy penalty associated with bringing the two residues in such a close vicinity without any covalent bond formation. This involved two further sampling rounds with the α-carbon distances set to 8 Å. The latter choice was considered as non-critical, given that the relative energetic and geometric information for the different bonding situations did not change dramatically following variation of the said parameter. Each sampling run was performed following the same steps as before but with an added constraint to the N–S distance of 2.7 Å. No additional filter process for hydrogen bonds was applied. The r.m.s.d. criteria were also not applied because we were not interested in a distribution but simply the most energetically stable geometry. The energies of the lowest conformers ('global' minima) were compared for the runs with and without the distance constraint. This resulted in an energy gap of 4.3 kcal mol^{-1} (Supplementary Fig. 8). The thermochemical calculations for formation of the SONOS bridge are shown for the SARS-CoV-2 Mpro (Supplementary Fig. 9), which showcase a putative pathway for two subsequent oxidation steps.

Search of the PDB for potential NOS or SONOS bridges. On 2 December 2020, we searched all deposited structures at the PDB (www.rcsb.org) with a resolution of 2 Å or better (65,327 of ~170,000 structures) for potential NOS or SONOS bridges by using the program NCONT of the CCP4 suite[58]. In view of the quantum chemically calculated N–S interatomic distance of 2.6–2.7 Å for NOS bridges, we defined thresholds of 3.00 Å as an upper limit (exclusion of non-covalent interactions) and 2.45 Å as a lower limit (exclusion of sulfenamides with N–S bonds) for the search. As non-covalent hydrogen-bond interactions between lysines and cysteines exhibit interatomic distances of >3.2 Å, we considered a cutoff at 3 Å to be a robust discriminator between covalent and non-covalent cysteine–lysine interactions. A total of 285 PDB entries with ~400 potential NOS/SONOS bridges were identified; for 266 of these, the corresponding structure factors were deposited. We examined all entries manually and inspected the potential NOS/SONOS redox switch sites. We also calculated mF_o–DF_c electron density OMIT maps to eliminate model bias using PHENIX.POLDER[59], omitting all entities constituting the suggested bridge. In case the electron density maps indicated the presence of a covalent cross-link between the lysine N atom and the cysteine S atom, we rebuilt the structural model with an NOS/SONOS bridge and compared the obtained models and electron density maps with those calculated for structures with a non-covalent interaction between the lysine and cysteine side

chains. Model building was performed using COOT[46], and refinements were done with PHENIX.REFINE[45].

Analysis of sequence conservation. The analysis of sequence conservation of lysine and cysteine residues forming NOS or SONOS bridges was performed for the 1,000 closest related protein sequences obtained using BLASTp[60] for the respective protein of interest. Alignment was performed using MAFFT[61].

Analysis of phylogenetic distribution of NOS bridge-containing proteins. Sequences of proteins identified to likely or possibly contain NOS/SONOS bridges (Supplementary Data 1) were downloaded from the PDB database. For each protein, homologs were searched with BLASTp[62] against NCBI's non-redundant database, using a stringent E value cutoff ($<10^{-25}$). A custom Python script powered by ETE3 (ref. [63]) was used to identify the taxonomic lineage of each hit by querying NCBI's taxonomy database with each hit's taxid. Results were summarized at the third hierarchical level of NCBI's taxonomy and plotted, along with the taxonomic affiliation of proteins identified in PDB, onto the Tree of Life, according to recent studies[64–68].

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

Data availability

The refined structural protein models and corresponding structure–factor amplitudes are deposited under PDB accession codes 7OEY (*Ng*TAL variant Glu93Gln oxidized), 7ODO (*Ng*TAL wild type oxidized, 0.27-MGy dose), 7ODP (*Ng*TAL wild type oxidized, 2.7-MGy dose) and 7ODQ (*Ng*TAL wild type oxidized, 5.4-MGy dose). All structures cited in this publication are available under their respective PDB accession codes. All other data are available on request.

References

40. Gill, S. C. & Von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).
41. Tsolas, O. & Joris, L. Transaldolase. *Methods Enzymol.* **42**, 290–297 (1975).
42. Sautner, V. et al. Converting transaldolase into aldolase through swapping of the multifunctional acid–base catalyst: common and divergent catalytic principles in F6P aldolase and transaldolase. *Biochemistry* **54**, 4475–4486 (2015).
43. Zeldin, O. B., Gerstel, M. & Garman, E. F. RADDOSE-3D: time-and space-resolved modelling of dose in macromolecular crystallography. *J. Appl. Crystallogr.* **46**, 1225–1230 (2013).
44. Kabsch, W. Xds. *Acta Cryst. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
45. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
46. Emsley, P. & Lohkamp, B. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
47. Carugo, O. How large B-factors can be in protein crystal structures. *BMC Bioinformatics* **19**, 61 (2018).
48. Williams et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
49. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
50. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
51. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
52. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
53. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
54. Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **15**, 2847–2862 (2019).
55. Bannwarth, C. et al. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1493 (2021).
56. Bannwarth, C., Ehlert, S. & Grimme, S. Gfn2xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and densitydependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
57. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169–7192 (2020).
58. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
59. Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Struct. Biol.* **73**, 148–157 (2017).
60. Myers, W. G. W. M. E., Altschul, S. F. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
61. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
62. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
63. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
64. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
65. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
66. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
67. Schulz, F. et al. Towards a balanced view of the bacterial tree of life. *Microbiome* **5**, 140 (2017).
68. Waite, D. W. et al. Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* **70**, 5972–6016 (2020).
69. Puvar, K. et al. *Legionella* effector MavC targets the Ube2N~Ub conjugate for noncanonical ubiquitination. *Nat. Commun.* **11**, 2365 (2020).

Acknowledgements

This study was supported by the Max-Planck Society and the DFG-funded Göttingen Graduate Center for Neurosciences, Biophysics and Molecular Biosciences GGNB. We acknowledge access to beamline P14 at DESY/EMBL Hamburg, Germany, and thank G. Bourenkov and T. Schneider for local support. We thank T. Curran, B. Schulman, S. Lorenz, B. Matthews, R. Hilgenfeld, R. Lill, R. Jahn, N. Brose, A. Pearson, D. Fass, C. Scholz, C. Berndt, H. Stark, A. Chari, E. Paknia, M. Alcarazo, R. Golbik, H. Urlaub, U. Curth and J. Rutter for discussion.

Author contributions

K.T. designed and coordinated research. F.R.v.P. and K.T. analyzed the protein database and interpreted crystallographic data. F.R.v.P. performed model building and crystallographic refinements. F.R.v.P. analyzed sequence conservation of NOS/SONOS residues. M.W. expressed, purified, biochemically characterized and crystallized *Ng*TAL proteins with support from L.-M.F. J.U. and R.A.M. planned the quantum chemical calculations and the model system for analysis of NOS bonding. J.Y. performed the density functional theory calculations and analyzed the data. I.I. and J.d.V. conducted bioinformatic analyses and inferred taxonomic distribution of protein homologs. F.R.v.P., R.A.M. and K.T. wrote the manuscript with input from all authors.

Funding

Open access funding provided by Max Planck Society.

Competing interests

F.R.v.P., M.W., L.-M.F., J.U., R.A.M. and K.T. have filed a European patent application (application number EP21164101.4) for regulating protein activities by targeting the NOS or SONOS redox switches. The remaining authors declare no competing interests.

Additional information

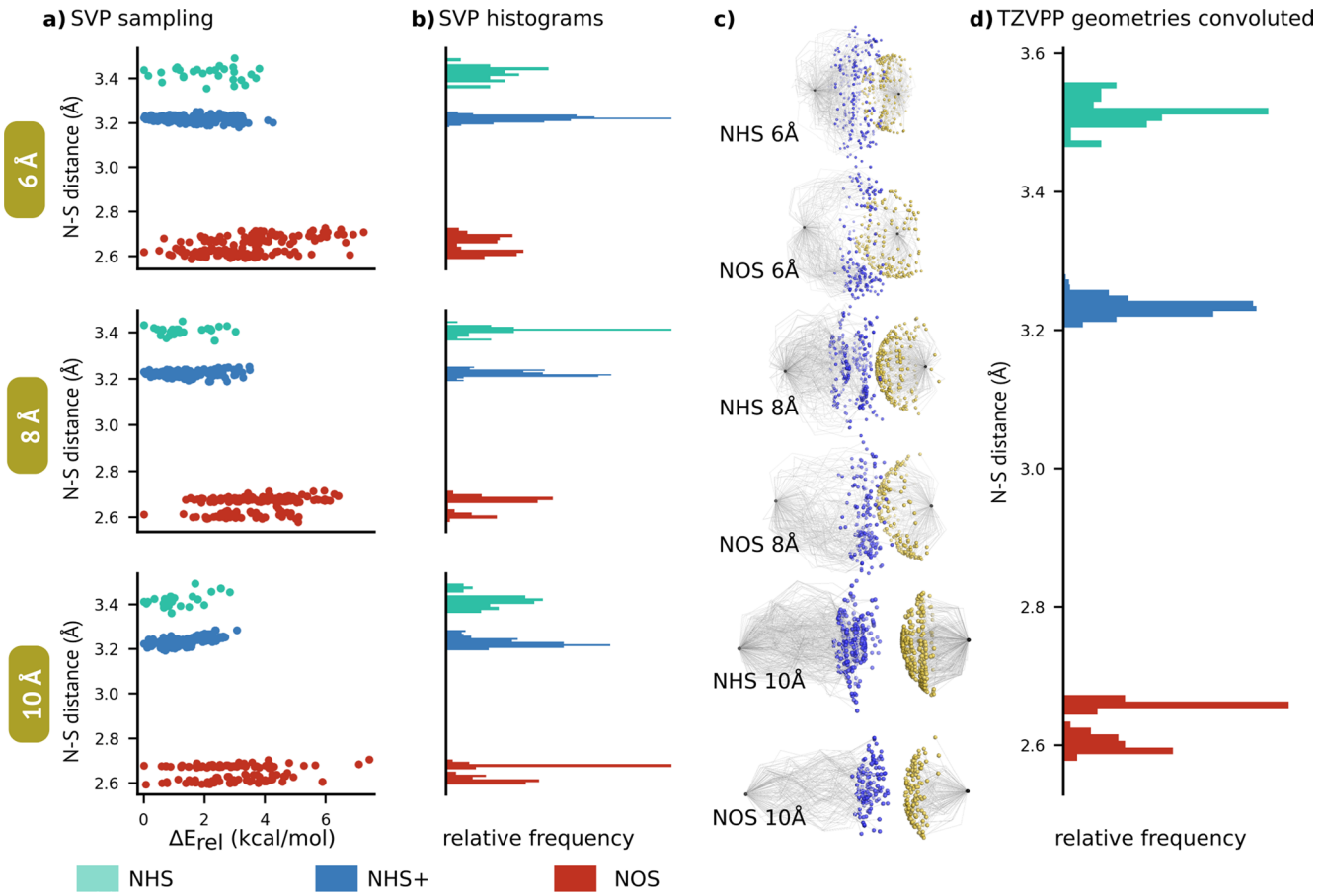
Extended data is available for this paper at https://doi.org/10.1038/s41589-021-00966-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-021-00966-5.

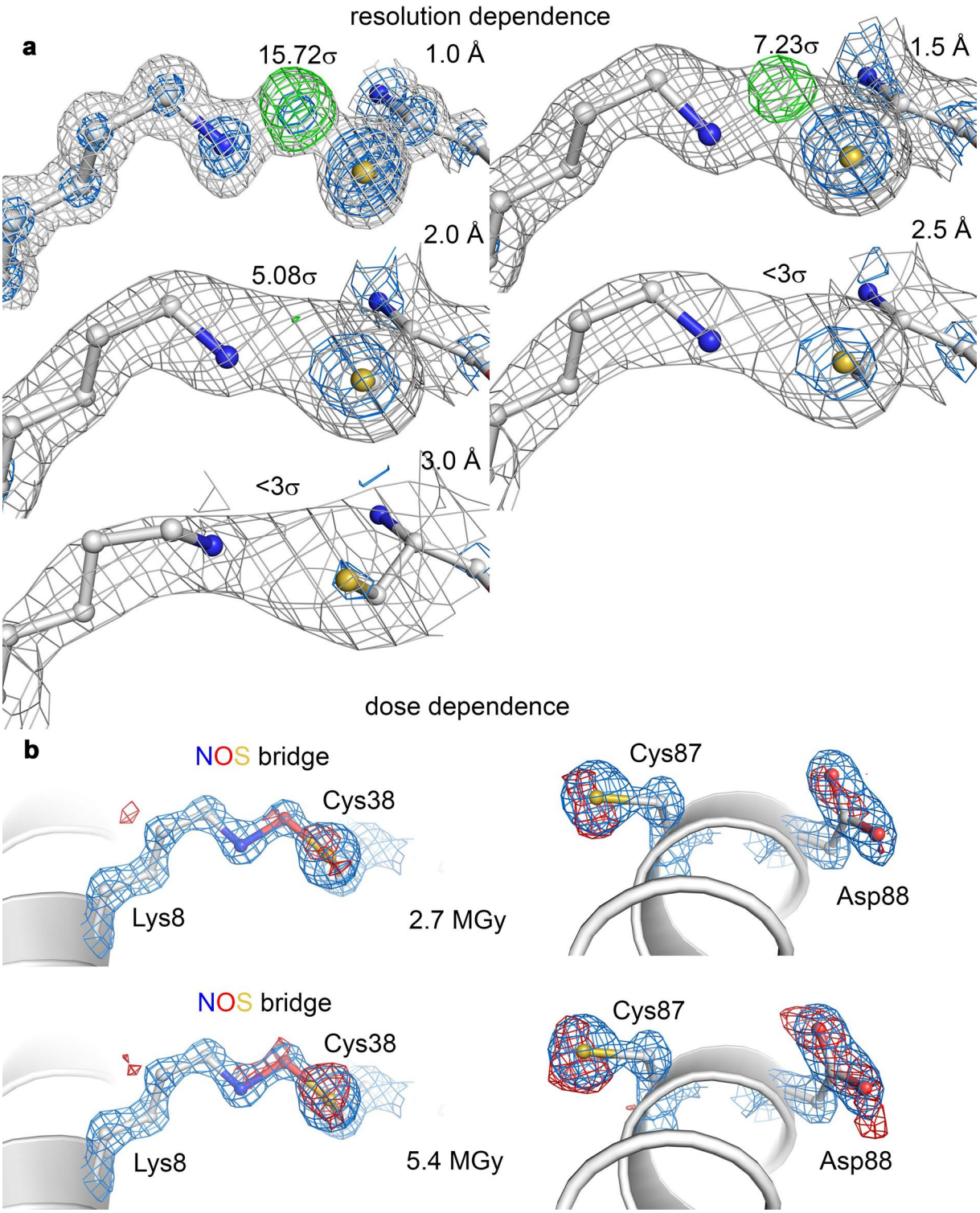
Correspondence and requests for materials should be addressed to Kai Tittmann.

Peer review information *Nature Chemical Biology* thanks Ivo Tews and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

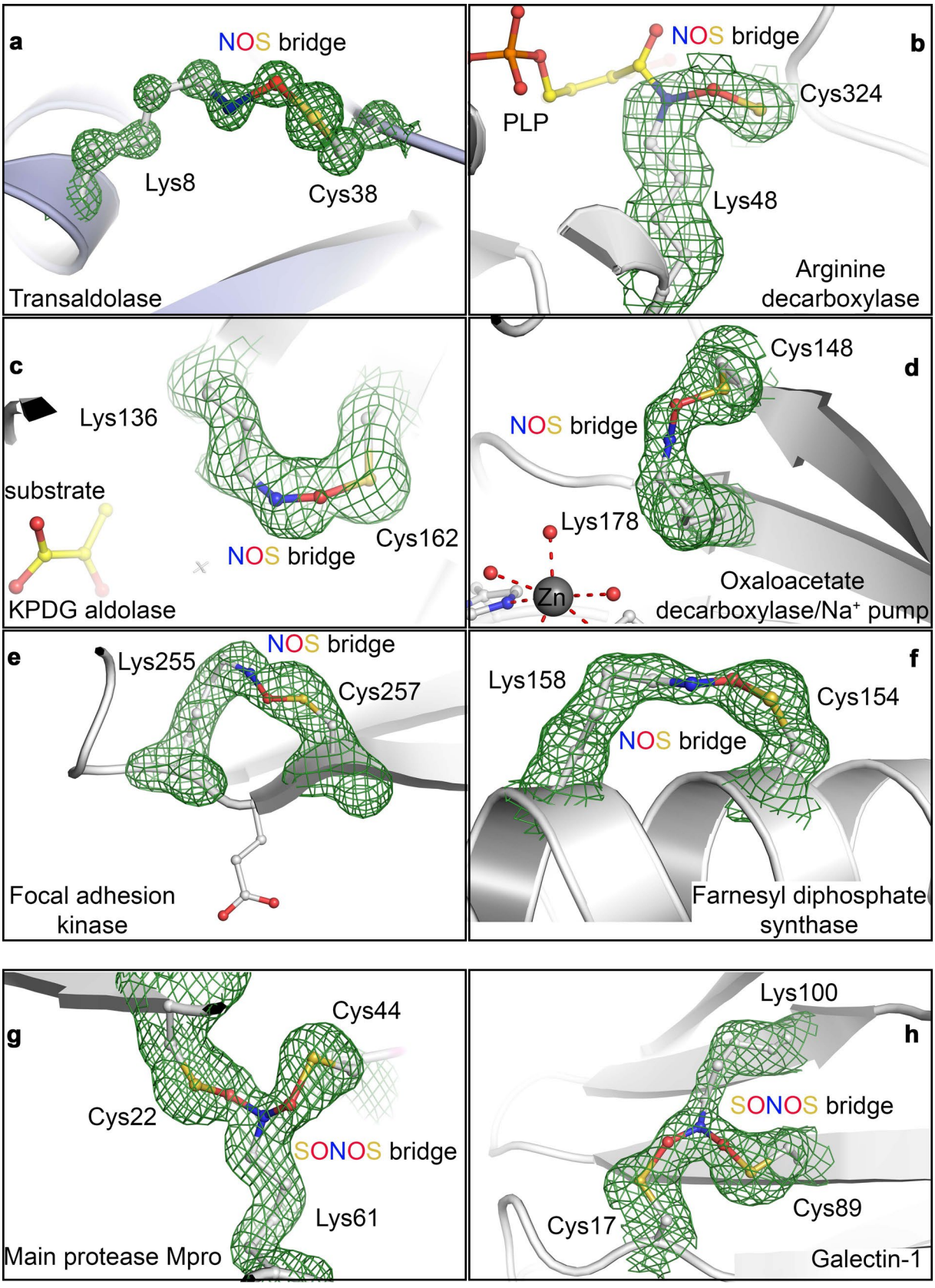


Extended Data Fig. 1 | Computed structural information for conformers of Lys and Cys residues with and without the NOS bond. Computed structural information for conformers of Lys and Cys residues with and without the NOS bond, using different alpha-carbon distances (6 Å, 8 Å and 10 Å). The residues are computed as models, truncated at the alpha-carbons. The sampled structures at the semi-empirical level were refined at the B3LYP-D3(BJ)/def2-SVP level of theory (more details in the Supplementary Information), with the lowest conformers being reoptimised with a larger basis (def2-TZVPP). We have considered three different conformers: NOS - covalent bonding between the residues, NHS - hydrogen bond with the Lys deprotonated and NHS+ - hydrogen bond with the Lys protonated. (**a**) N-S distance (in Å) as a function of the relative conformer stability within each group computed with the SVP basis; (**b**) histogram for the SVP basis results; (**c**) overlap of SVP optimised structures for NHS and NOS; (**d**) histogram for all structures with the TZVPP basis. The images in **c** are obtained by an arbitrary alignment of the structures (given that only two points are fixed - the alpha-carbon positions). This can lead to somewhat different representations, or apparent 'banding' of the structures. This effect will depend on the alignment chosen and should not be misinterpreted. The figures are there to help illustrate the conformational freedom of these interactions..



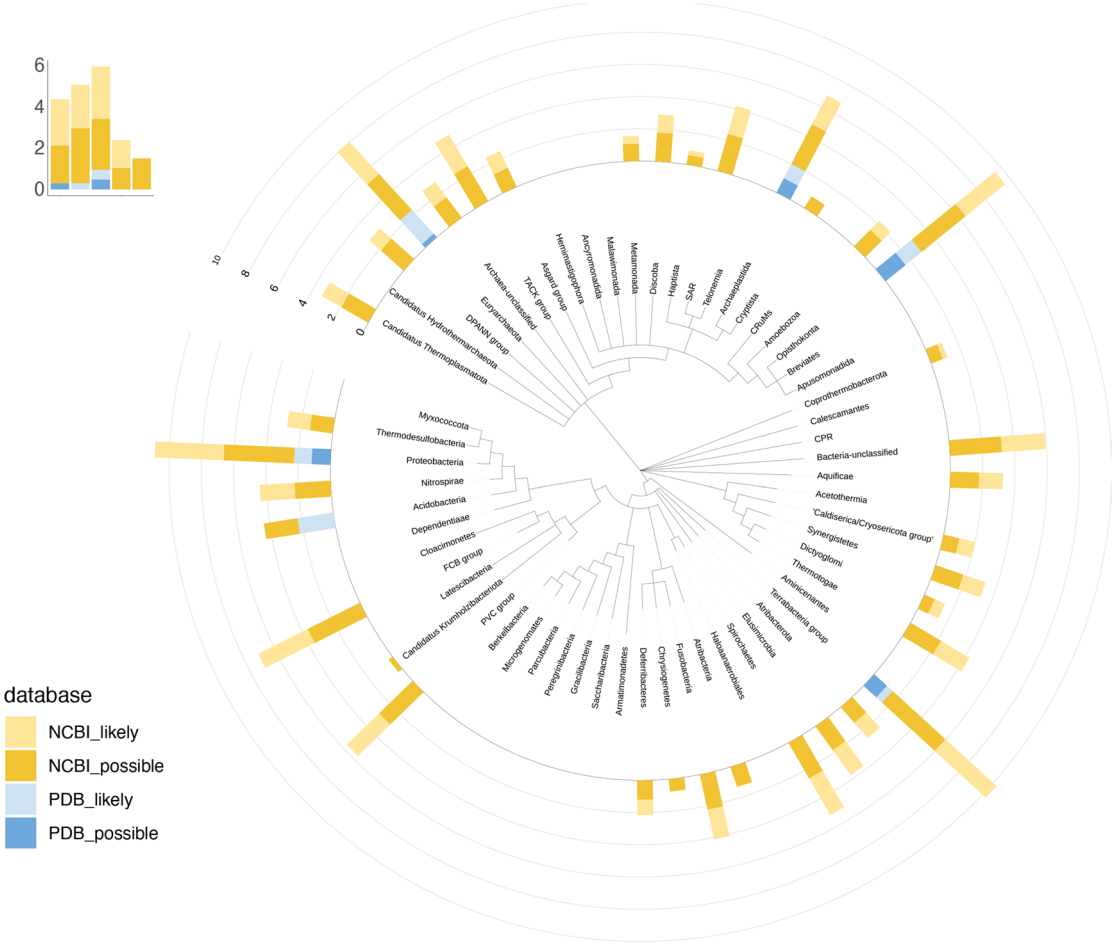
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Detectability of NOS bridges in protein crystal structures in dependence from resolution and deposited dose. Detectability of NOS bridges in protein crystal structures in dependence from resolution and deposited dose. (**a**) Data truncation at different resolutions. A previously determined sub-ångstrøm resolution crystallographic dataset obtained for *Neisseria gonorrheae* transaldolase (pdb code 6XZ4)[9], which forms an allosteric NOS bridge between Lys8 and Cys38, was truncated at different resolutions from 1.0-3.0 Å as indicated. For refinement, which included simulated annealing and B-factor blurring, the lysine and cysteine residues were modeled without the bridging oxygen atom. The calculated structural models and 2mFo-DFc electron density (blue: 3σ, grey: 1σ) and mFo-DFc difference electron density (green: 5σ) maps are shown. The respective intensities of the positive peaks in the difference electron density maps are indicated. Note the loss of structural information in both maps regarding the bridging oxygen atom with decreasing resolution. (**b**) Dose dependence. Crystallographic datasets for a single crystal of *Neisseria gonorrheae* transaldolase were collected at different doses 0.27, 2.7 MGy and 5.4 MGy at cryogenic temperature. The refined structural models of the NOS bridge formed between Lys8 and Cys38 and of a neighboring helix bearing residues Cys87 and Asp88 are shown for datasets obtained at doses of 2.7 MGy and 5.4 MGy. The respective 2mFo-DFc electron density maps shown in blue are contoured at 2σ. The calculated difference in intensity in the electron density maps relative to that obtained for a dose of 0.27 MGy is shown in red at a contour level of 3σ. Note the progressive loss in intensity for Cys87 and Asp88 with increasing dose, which indicates 'radiation damage'. For the NOS bridge, radiation damage sets in at higher doses of 5.4 MGy.

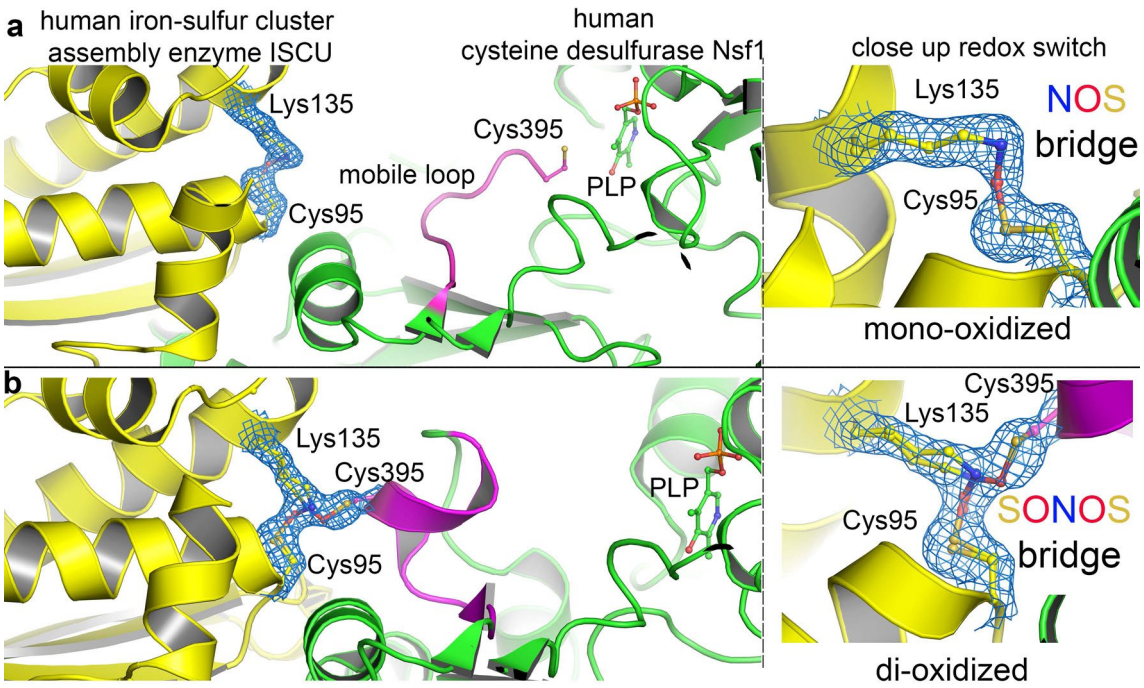


Extended Data Fig. 3 | See next page for caption.

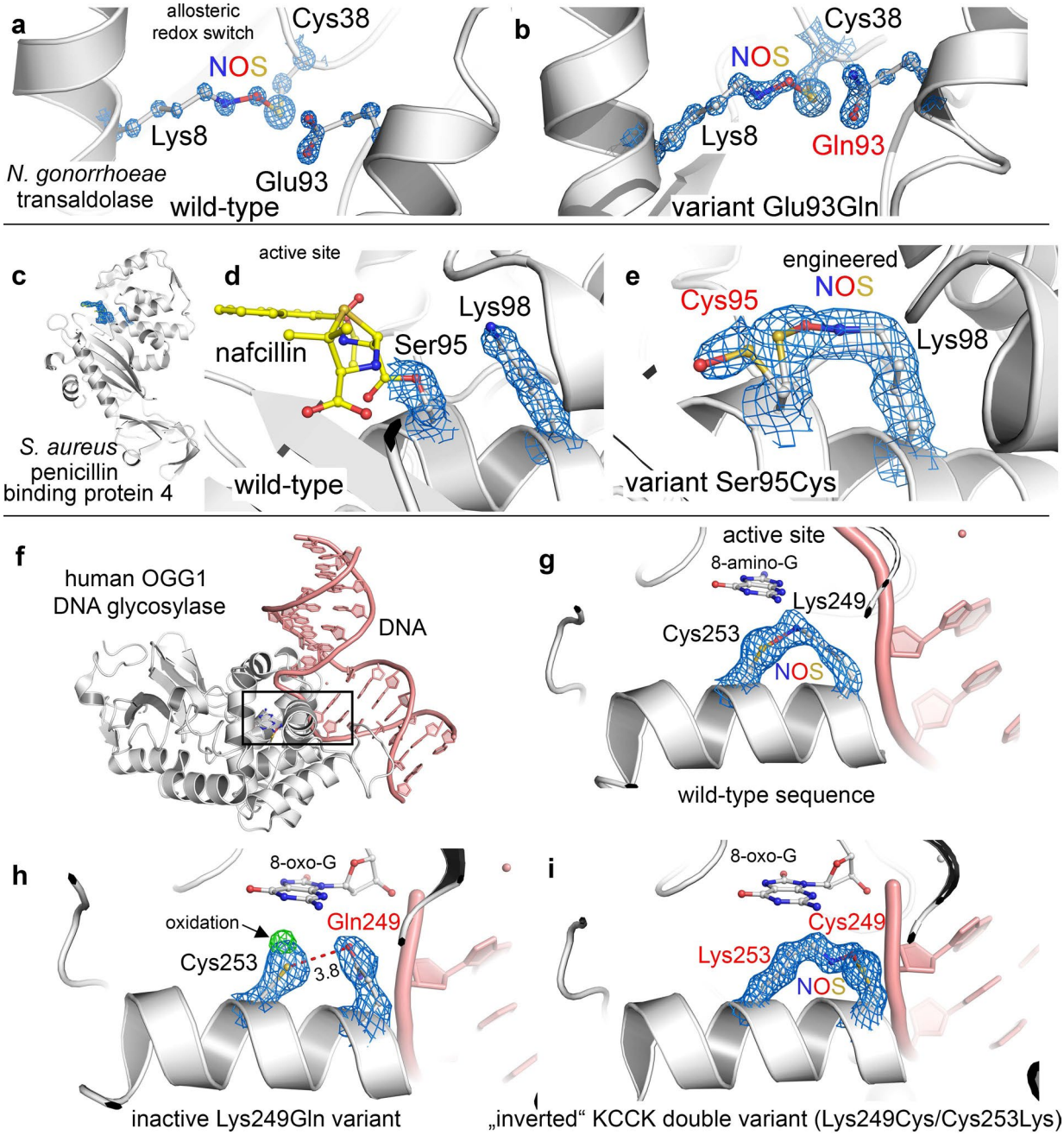
Extended Data Fig. 3 | Unbiased mFo-DFc omit maps for NOS and SONOS bridges in representative examples calculated by Phenix.polder[59]. Unbiased mFo-DFc omit maps for NOS and SONOS bridges in representative examples calculated by PHENIX.POLDER[59]. The lysine and cysteine residues incl. the NOS or SONOS bridges were excluded from the structural models prior to map calculation to eliminate model bias. The corresponding pdb codes and contour levels are indicated. (**a**) transaldolase from *Neisseria gonorrhoeae* (pdb code 6ZX4, 5σ), (**b**) arginine decarboxylase from *Paramecium bursaria* Chlorella virus (pdb code 2NV9, 3.5σ), (**c**) KPDG aldolase from *Oleispira antarctica* (pdb code 3VCR, 4σ), (**d**) oxaloacetate decarboxylase/Na^+ pump from *Vibrio cholerae* (pdb code 2NX9, 3σ), (**e**) focal adhesion kinase from *Gallus gallus* (pdb code 6CB0, 5σ), (**f**) farnesyl diphoshate synthase from *Trypanosoma cruzi* (pdb code 6SDP, 3σ), (**g**) main protease from SARS-CoV-2 (pdb code 7JR4, 5σ) and (**h**) galectin-1 from *Rattus norvegicus* (pdb code 4GA9, 3σ).



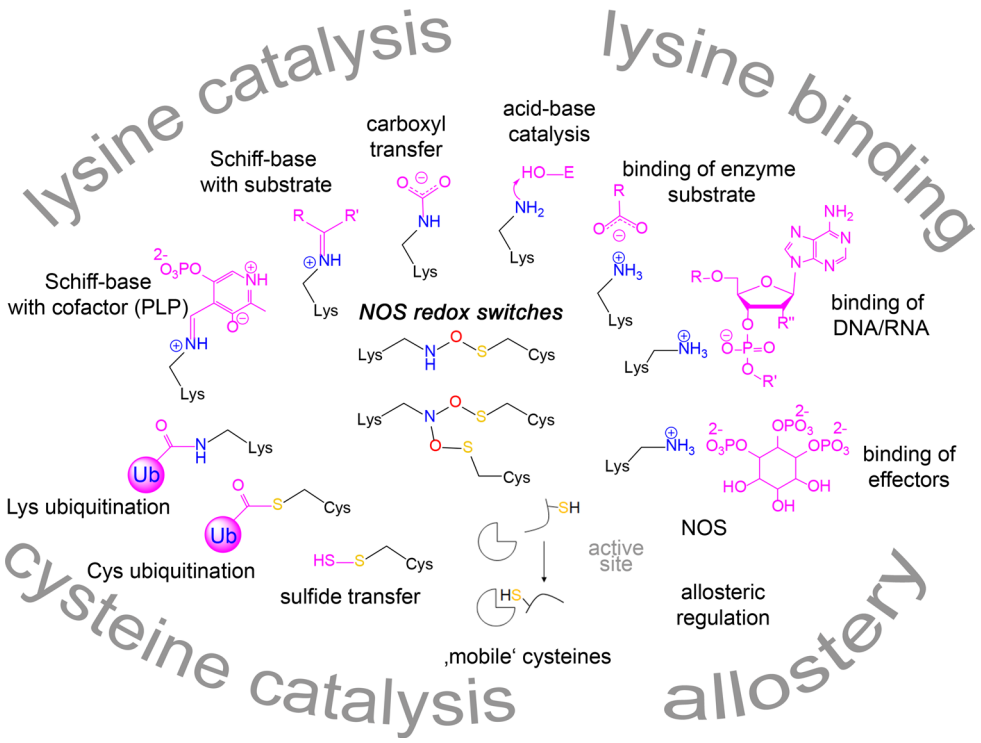
Extended Data Fig. 4 | Distribution of proteins containing NOS crosslinks across the Tree of Life. Distribution of proteins containing NOS crosslinks across the Tree of Life. Shown are proteins likely/ probably containing NOS bridges (blue) as identified from PDB structures (Supplementary Data 1), as well as homologs of these proteins as identified in NCBI's non-redundant (NR) database (yellow). Values are log10-transformed.



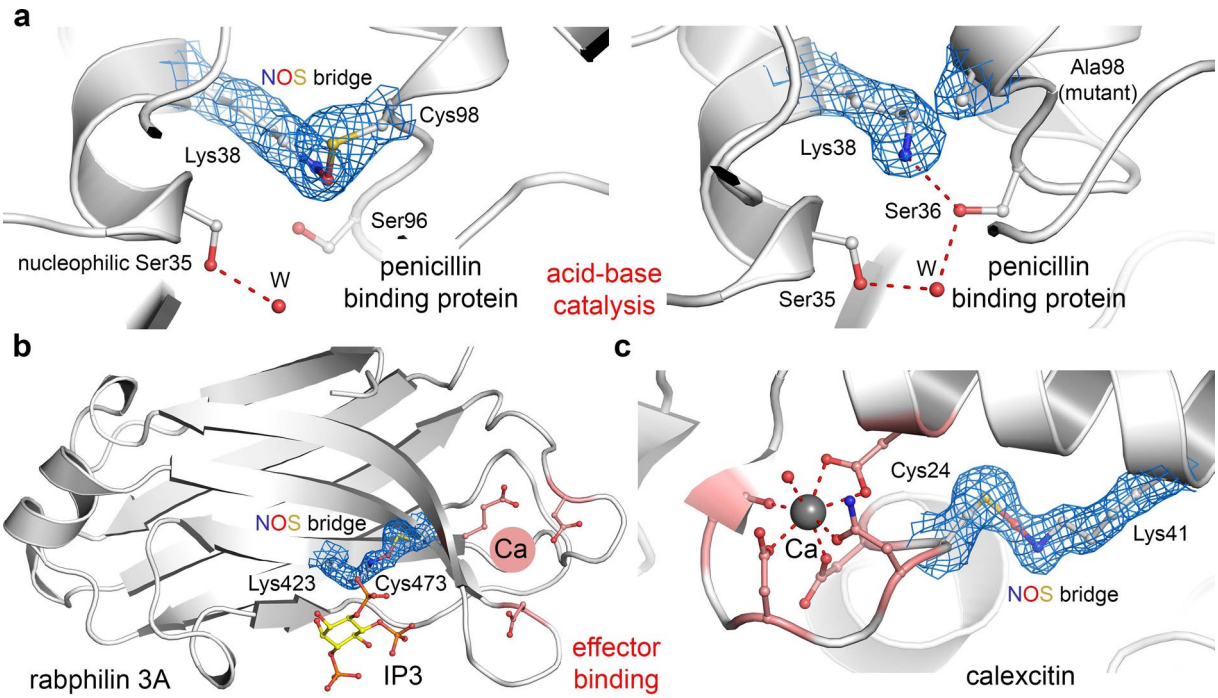
Extended Data Fig. 5 | NOS and SONOS bridges in the human Fe/S cluster biosynthesis complex. NOS and SONOS bridges in the human Fe/S cluster biosynthesis complex. The corresponding 2mFo-DFc electron density maps are shown in blue at a contour level of 1σ. (**a**) Structure of the complex consisting of iron-sulfur cluster assembly enzyme ISCU (highlighted in yellow) and cysteine desulfurase Nsf1 (highlighted in green) in the mono-oxidized state with an intramolecular NOS bridge between Cys95 and Lys131 of ISCU (pdb code 6UXE). A mobile loop from Nsf1 with catalytic residue Cys395 catalyzing sulfide transfer from the PLP cofactor at the active site of Nsf1 to ISCU is indicated in magenta. Left panel: structural overview. Right panel: close-up of the redox switch site. Note that Cys395 visits the active site of Nsf1. (**b**) Structure of the complex in the di-oxidized state with a SONOS bridge formed between Cys95 and Lys131 of ISCU and Cys395 of Nsf1 (pdb code 6WI2, also observed in 6WIH). Left panel: structural overview. Right panel: close-up of the redox switch site. Competitive refinements (SONOS bridge only, 2 separate NOS bridges, mixture of SONOS and NOS) indicate a mixture of the SONOS bridge (refined occupancy 60%) and the NOS bridge between Lys135 and Cys95 (refined occupancy 40%) for this dataset.



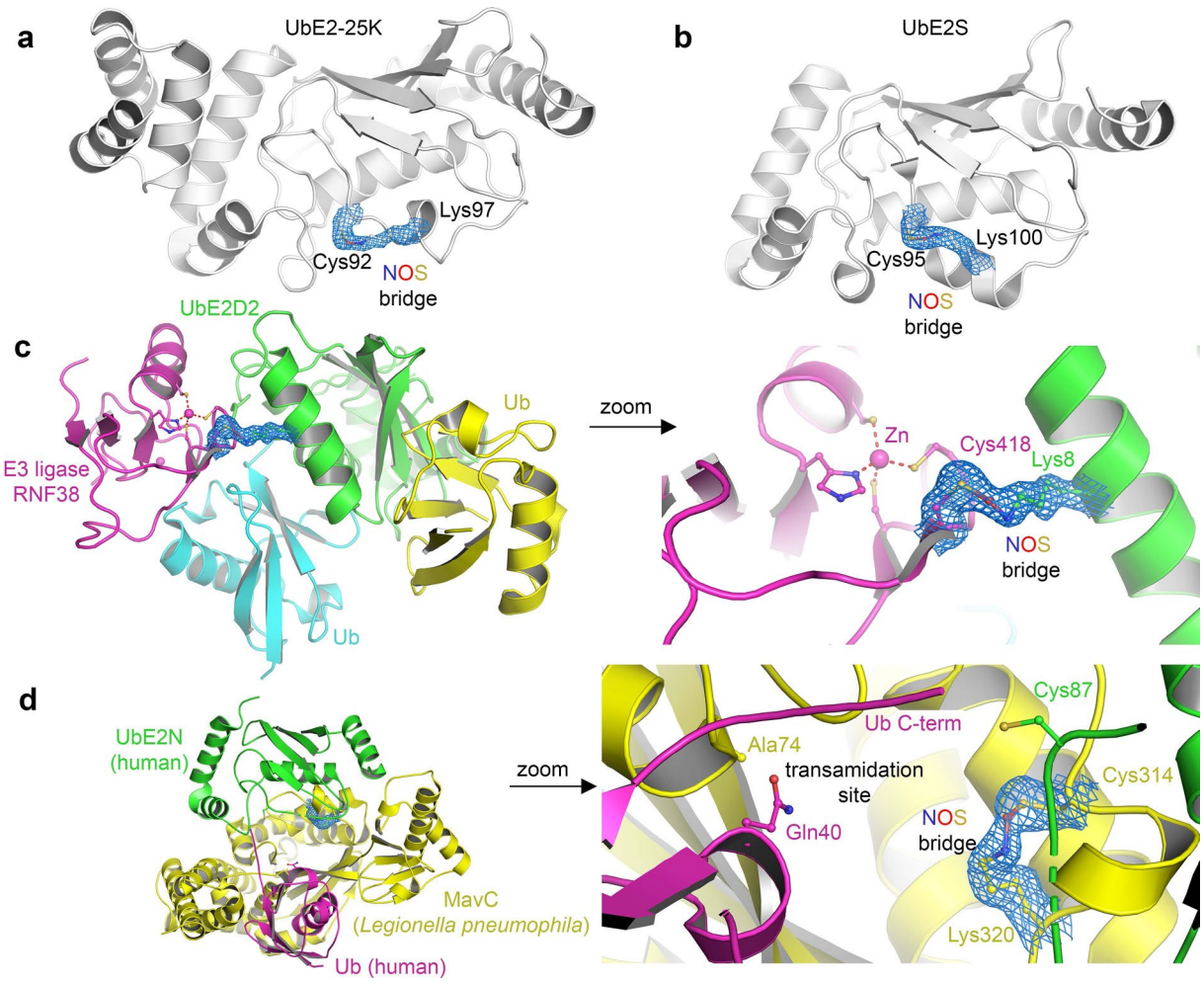
Extended Data Fig. 6 | Structural requirements for formation and engineering of NOS bridges. Structural requirements for formation and engineering of NOS bridges. (**a**, **b**) Structure of the allosteric NOS bridge in *Neisseria gonorrhoeae* transaldolase wild-type (a, pdb code 6XZ4)[9] and variant Glu93Gln (b, this study) showing NOS residues Lys8 and Cys38 as well as neighboring residue 93. The structural models of Lys8, Cys38 and Glu93 or Gln93 are superposed with the corresponding 2mFo-DFc electron density maps at 3σ. Note that NOS bridge formation does not require the presence of a catalytic residue at position 93. (**c**, **d**, **e**) An unintentionally engineered lysine-cysteine crosslink in *Staphylococcus aureus* penicillin binding protein 4 showing the overall protein structure (c) and a close-up of the active site with catalytic residues Ser95 (nucleophile) in linkage with nafcillin and Lys98 (d, pdb code 5TY2). In variant Ser95Cys (e), where the catalytic Ser95 has been replaced by cysteine, an NOS bridge was formed between introduced Cys95 and Lys98. The 2mFo-DFc electron density map is contoured at 1σ. In variant Ser95Cys, residue 95 was modeled in two alternative conformations. Note that a sulfenamide linkage with an N-S bond could also explain the density. (**f**-**i**) NOS bridges in the human DNA glycosylase OGG1. (**f**) Overall structure of OGG1 in complex with DNA (pdb code 1M3Q). (**g**) Close-up of the active site showing catalytic residue Lys249 (Schiff-base with DNA) and neighboring residue Cys253 forming an NOS bridge. The 2mFo-DFc electron density map is contoured at 1.5σ⊡ (**h**) Active site structure of inactive OGG1 variant Lys249Gln (pdb code 1EBM), in which the catalytic lysine had been replaced by glutamine. The 2mFo-DFc electron density map is shown in blue at 1.5σ and the mFo-DFc difference electron density map in green at 3σ⊡ Note that Gln249 and Cys253 form a H-bond rather than an NOS bridge and that Cys253 is likely to be oxidized as indicated by the positve peak in the difference electron density map. (**i**) Active site structure of OGG1 KCCK double variant Lys249Cys/Cys253Lys (pdb code 2XHI), in which the sequence positions of the NOS-bridge-forming Lys and Cys residues are inverted. The 2mFo-DFc electron density map is shown at 1.5σ. Note the presence of the NOS bridge in this 'inverted' variant akin to the wild-type sequence argueing against a mechanism, where catalysis is provided by neighboring residues for the NOS bridge to form.



Extended Data Fig. 7 | Chemical functions of lysine and cysteine residues forming NOS and SONOS bridges in proteins. Chemical functions of lysine and cysteine residues forming NOS and SONOS bridges in proteins. Four major functional categories could be identified including i) lysines with catalytic roles in enzyme mechanisms, ii) lysines involved in binding of enzyme substrates, nucleic acids and effectors, iii) cysteines with catalytic roles in enzyme mechanisms and iv) allosteric bridges, which are located remotely relative to the active/functional site. Structures of key reaction intermediates and interaction partners are highlighted. Note that these functions are exerted under reducing conditions that is in the absence of NOS/SONOS bridges. Formation of the NOS or SONOS bridge under oxidizing conditions leads to either a loss-of-function (catalytic lysines, catalytic cysteines), diminished biologial activity (lysine with binding roles) or modulated function (allosteric switches). Specific information about all proteins regarding origin, biological function, type of NOS/SONOS redox switch, suggested mechanism of the redox switch and potential relevance in disease states is compiled in Supplementary Tables 2 and 3.



Extended Data Fig. 8 | Functional roles of NOS bridge lysines in enzyme catalysis and in binding of enzymatic substrates or effectors (see also Fig. 4). Functional roles of NOS bridge lysines in enzyme catalysis and in binding of enzymatic substrates or effectors (see also Fig. 4). The corresponding 2mFo-DFc electron density maps are shown in blue at a contour level of 1σ. (**a**) Catalytic lysines acting as acid-base catalysts. Left: Penicillin binding protein (transpeptidase) from *Streptomyces sp.* K15 with an NOS bridge between Lys38 and Cys98 (pdb code 1SKF). Right: Penicillin binding protein variant Cys98Ala from *Streptomyces sp.* K15 (pdb code 1ES3). As Lys38 is thought to activate the nucleopilic Ser35 by acid-base catalysis, formation of an NOS bridge under oxidizing conditions is likely to inhibit catalysis. (**b,c**) Lysines with roles in binding of effectors. (**b**) Rabphilin 3a from *Rattus norvegicus* in complex with effector inositol trisphosphate (IP3) and with an NOS bridge formed between Lys423 and Cys473 (pdb code 4NP9). Ligand IP3 is weakly occupied with traceable density for the phosphate portions only. Lys423 is directly located at the binding site of IP3. The proximal Ca^{2+} binding site is indicated. (**c**) Calexcitin from *Loligo pealeii* with an NOS bridge formed between Lys41 and Cys24 proximal to the Ca^{2+} binding site (pdb code 2CCM).



Extended Data Fig. 9 | NOS bridges in proteins from the ubiquitin system. NOS bridges in proteins from the ubiquitin system. The corresponding 2mFo-DFc electron density maps are shown in blue at a contour level of 1σ. (**a**) Human ubiquitin-conjugating enzyme E2-25K (UbE2-25K) with an NOS bridge formed between catalytic residue Cys92 and Lys97, a potential auto-ubiquitination site (pdb code 3E46). (**b**) Human ubiquitin-conjugating enzyme E2-S (UbE2S) with an NOS bridge formed between catalytic residue Cys95 and Lys100, a potential auto-ubiquitination site (pdb code 6QHK). (**c**) Complex between ubiquitin-conjugating enzyme E2 D2 (UbE2D2), E3 ring ligase RNF38 and two ubiquitin (Ub) molecules (pdb code 4V3L). Left: Overall structure of the complex showing the proteins in individual colors. Right: Close-up of the intermolecular NOS bridge formed between Lys8 of E2 and Cys418 of E3. Note the proximity of the NOS bridge with respect to the Zn^{2+}-binding site of E3. (**d**) Structure of a complex between MavC from *Legionella pneumophila* in complex with human ubiquitin-conjugating enzyme E2 N (UbE2N) and human ubiquitin (Ub) (pdb code 6ULH). MavC is a bacterial effector that inactivates the human ubiquitin system by catalyzing a transamidation reaction between Lys92 of UbE2N and Gln40 of Ub forming a dead-end complex[69]. Left: Overall structure of the complex showing the proteins in individual colors. Right: Close-up of the intramolecular NOS bridge in MavC formed between Lys320 and Cys314. Note the proximity of the NOS bridge with respect to the transamidation site.

nature research

Corresponding author(s): Tittmann, Kai

Last updated by author(s): Dec 4, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection

Kinetics and spectroscopic analysis
Spectra Manager version 2.07.02 (Build 4) (UV-Vis-based Kinetics)

X-ray crystallography
mxCuBE 2 (X-ray data collection at P14 DESY/EMBL, Hamburg, Germany)

Electronic structure calculations
Gaussian16-A.03
CREST-2.10.2
xTB-6.4.0

Data analysis

Kinetics and spectroscopic analysis
Spectra Manager version 2.07.02 (Build 4) (UV-Vis-based Kinetics)
SigmaPlot version 11.0 (Circular Dichroism, UV-Vis-based Kinetics)

Crystallography
XDS VERSION Mar 15, 2019 BUILT=20190315 (X-ray, data processing)
XSCALE VERSION Jan 31, 2020 BUILT=20200417 (X-ray, data scaling)
CCP4 version 7.0.78 (X-ray, processing and refinement)
phenix.refine version 1.13_2998 (structure, refinement)
COOT version 0.8.9.2 (structure, model building)
MolProbity-Server Version 4 (structure validation)
The PyMOL Molecular Graphics System version 1.3 (structure representation)

Electronic structure calculations
Plotting and data analysis of electronic structure results
Python3.1 (libraries: numpy and matplotlib)

Phylogenetic analysis
MAFFT 7
iTOL 5
Jalview 2.11

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The refined structural protein models and corresponding structure-factor amplitudes are deposited under PDB accession codes 7OEY (NgTAL variant Glu93Gln oxidized), 7ODO (NgTAL wild-type oxidized, 0.27 MGy dose), 7ODP (NgTAL wild-type oxidized, 2.7 MGy dose), 7ODQ (NgTAL wild-type oxidized, 5.4 MGy dose). All structures cited in this publication are available under their respective PDB accession codes. All other data are available on request.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	n.a.
Data exclusions	n.a.
Replication	Measurements were repeated at least two times.
Randomization	n.a.
Blinding	n.a.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*

Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) — *State the source of each cell line used.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals — *For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Field-collected samples — *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) *Specify: functional, structural, diffusion, perfusion.*

Field strength *Specify in Tesla*

Sequence & imaging parameters *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*