

Supplementary Figure Legends

- Figure S1 Distribution of the RMSDs of the fits to the AlphaFold2 structures for all twelve ligand types that were successfully placed. Only cases in which all coordinating sidechains had pLDDT values of greater than 70 are recorded. The black, dashed line on each plot demarks the 0.5 Å RMSD threshold used to determine acceptable fits.
- Figure S2 Fractional abundance of (a) identified ligand-binding sites and (b) identified ligand-binding proteins for all organisms. Only cases in which all coordinating sidechains had pLDDT values of greater than 70 are recorded
- Figure S3 Examples of 4Fe-4S cluster binding sites identified with three coordinating cysteine residues. Protein chains are shown as cartoon colored using AlphaFold2's pLDDT score on a spectrum of red (low confidence) to blue (high confidence), 4Fe-4S clusters are shown as spheres with iron atoms colored pink and sulfur atoms colored yellow, and coordinating residues are shown as licorice colored by element with coordinating bonds labeled. (a) a binding site with a possible fourth coordinating cysteine residue (red arrow) that is too far away to be identified as part of the binding site using a 0.5 Å RMSD threshold. (b) a binding site with the fourth coordinating position exposed to bulk solvent. This image was prepared using VMD [50].
- Figure S4 Examples of Zn-binding sites identified with three coordinating histidine residues. Visualization scheme is the same as used in Figure S3. (a) a binding site with a nearby fourth histidine residue that is not quite in a coordinating geometry. (b) a binding site with a nearby aspartate residue as a candidate fourth coordinating residue. (c) a binding site with the fourth coordinating position exposed to bulk solvent.
- Figure S5 Venn diagram showing the extent of overlap between Fe-S cluster-binding proteins identified here in *E. coli* with those identified in previous prediction studies. Those identified here ("This Work") are represented by the red circle, those predicted by Estellon et al. [21] by the green circle, and those predicted by Valastava et al. [22] by the blue circle.

Figure S6 Distribution of the RMSDs of the fits to the AlphaFold2 structures for all twelve ligand types that were successfully placed for: (a) binding sites already annotated in UniProt, and (b) binding sites not annotated in UniProt. The blank panels in (a) are the result of two factors: (1) the rarity with which some ligand types are documented in UniProt (e.g. “2Fe-2S Cys₃ Asp₁” and “Zn His₄”), and (2) priority being given by our scheme to a different ligand type, e.g. UniProt-annotated “3Fe-4S Cys₃” binding sites are identified instead as “4Fe-4S Cys₃” binding sites).

Figure S7 Distribution of the RMSDs of the fits to the AlphaFold2 structures for all twelve ligand types that were successfully placed for: (a) binding sites already annotated in UniProt, or for which a structural homolog is identified with hhsearch, and (b) binding sites not annotated in UniProt and for which no structural homolog is identified with hhsearch. The blank panels that appear in (b) are the result of the strict criteria of using UniProt plus hhsearch to filter what we consider novel identified ligand binding sites.

Figure S8 Comparison of cysteines in the human proteome that are: (a) chemically reactive in experiment, (b) identified here in ligand binding sites, and/or (c) annotated in UniProt. All comparisons are shown as Euler diagrams with the outer circle representing the total number of cysteines in proteins that contained at least one reactive cysteine in the isoTOP-ABPP experiments [32], blue circles represent the subset of cysteines that are classed as “highly reactive” in the -ABPP experiments (“Kuljanin *et al.* Union”), red circles represent the subset of cysteines that are identified here as being parts of ligand binding sites, and grey circles represent the subset of cysteines that are annotated in UniProt as members of ligand binding sites (“UniProt”). (a) shows results obtained when all binding sites identified here are included (i.e. no pLDDT cutoff, (b) shows results obtained when only binding sites identified here for which all coordinating residues have pLDDT scores > 90 are included, (c) shows results obtained when only binding sites identified here for which all coordinating residues have pLDDT scores > 70 are included, and when all chemically reactive cysteines are included.

Figure S9 Comparison of cysteines in the *E. coli* proteome that are: (a) chemically reactive in experiment, (b) identified here in ligand binding sites, and/or (c) annotated in UniProt. All comparisons are shown as Euler diagrams with the outer circle representing the total

number of cysteines in proteins that contained at least one reactive cysteine in the isoTOP-ABPP experiments [23], blue circles represent the subset of cysteines that are classed as “highly reactive” in the experiments (“Wang *et al.*”), red circles represent the subset of cysteines that are identified here as being parts of ligand binding sites, and grey circles represent the subset of cysteines that are annotated in UniProt as members of ligand binding sites (“UniProt”). (a) shows results obtained when all binding sites identified here are included (i.e. no pLDDT cutoff, (b) shows results obtained when only binding sites identified here for which all coordinating residues have pLDDT scores > 90 are included, (c) shows results obtained when only binding sites identified here for which all coordinating residues have pLDDT scores > 70 are included, and when all chemically reactive cysteines are included.

Figure S10 A case highlighting AlphaFold2’s rare indecision when building metal binding sites in overlapping fragments. Visualization scheme is the same as used in Figure S3. (a) fragment O43149_F3 contains a single zinc coordinated by 4 cysteine residues (“Zn Cys₄”). (b) fragment O43149_F4 (and 5 subsequent fragments) contain a single zinc coordinated by 2 cysteine and two histidine residues (“Zn Cys₂ His₂”).

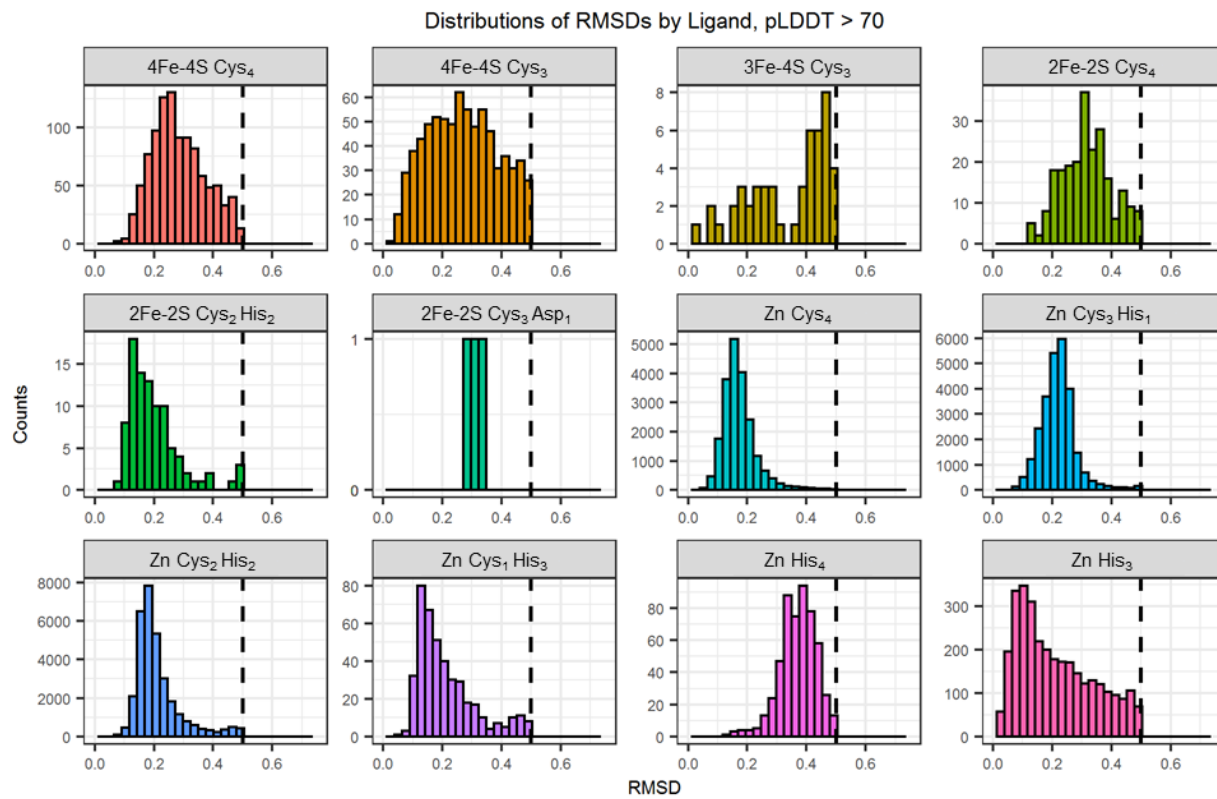


Figure S1

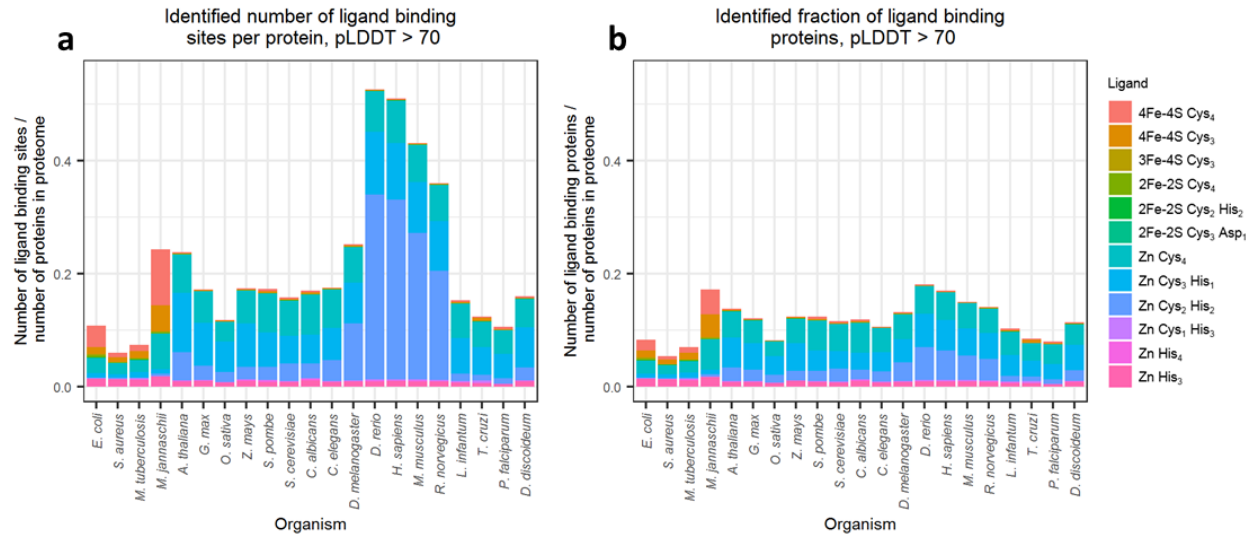


Figure S2

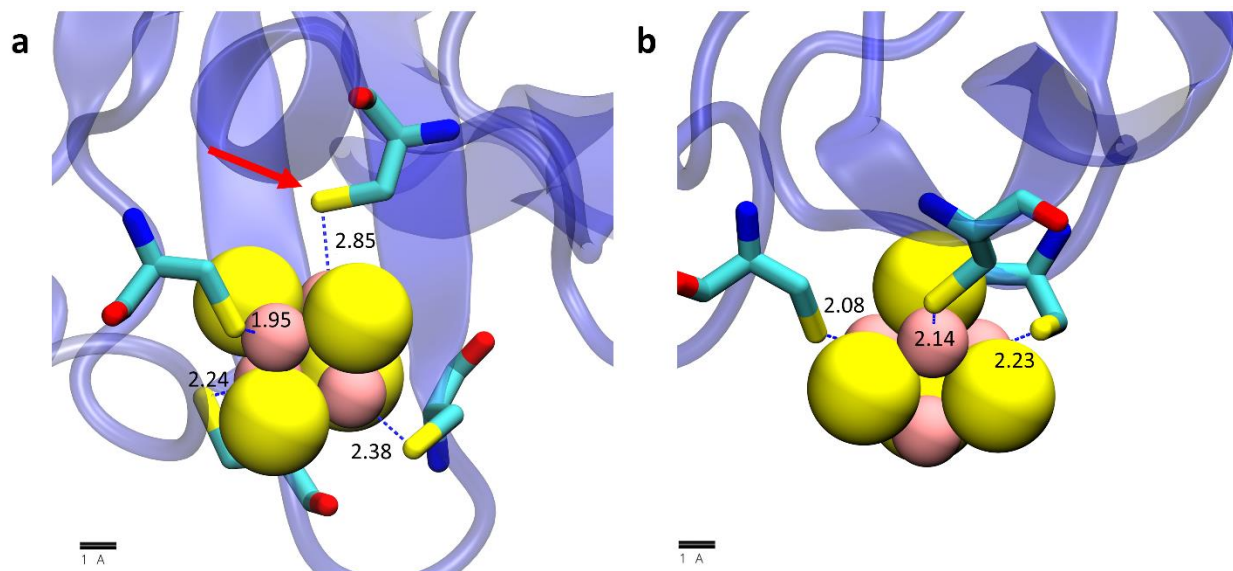


Figure S3

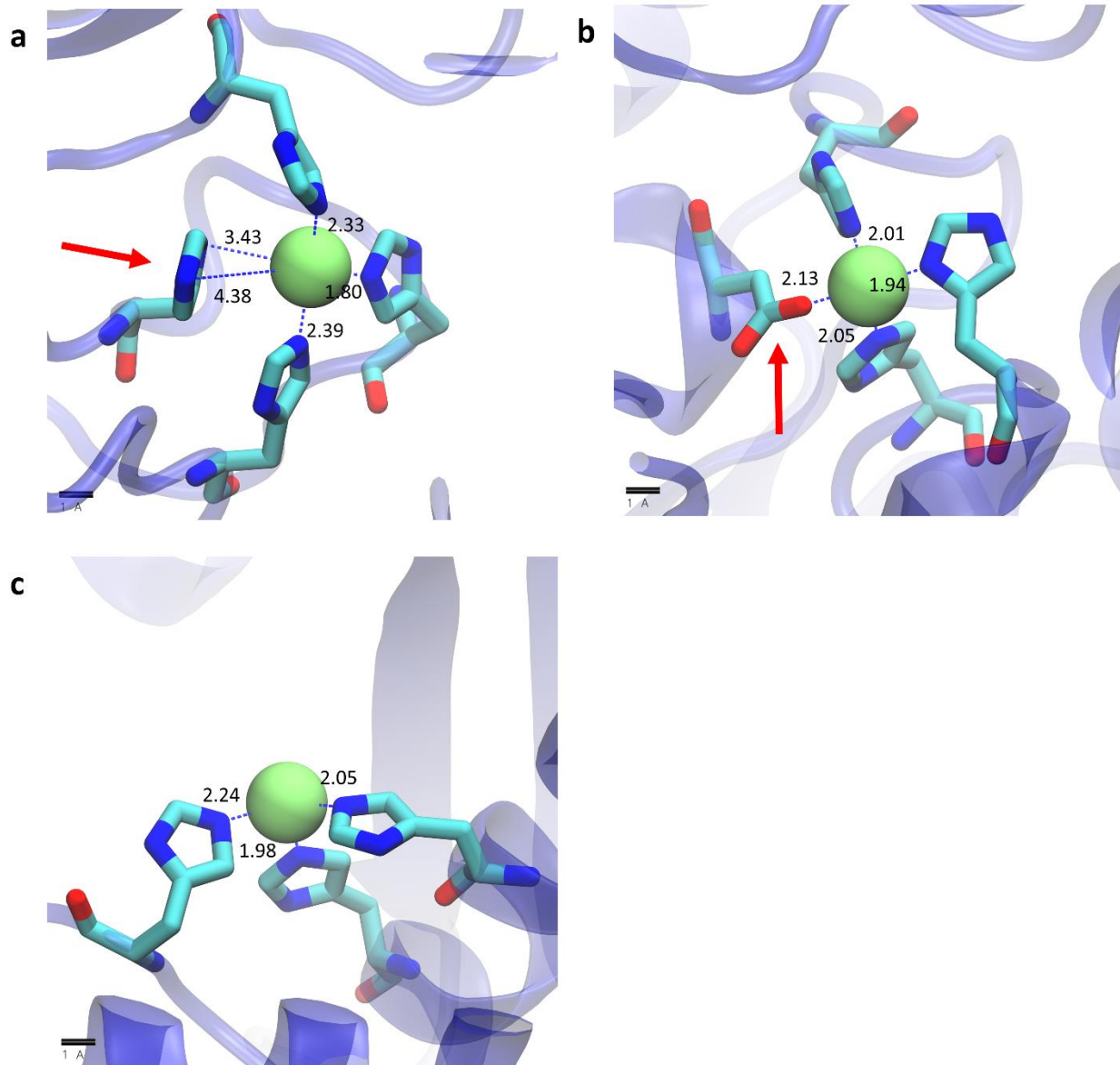


Figure S4

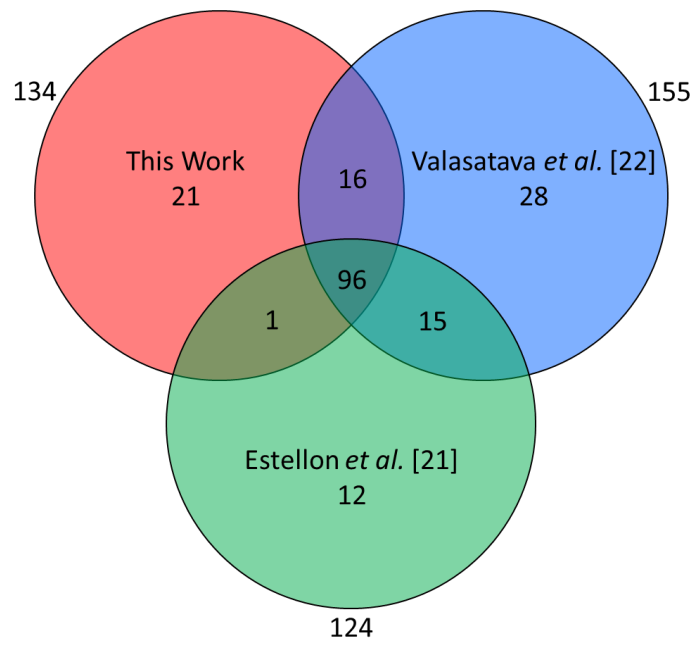


Figure S5

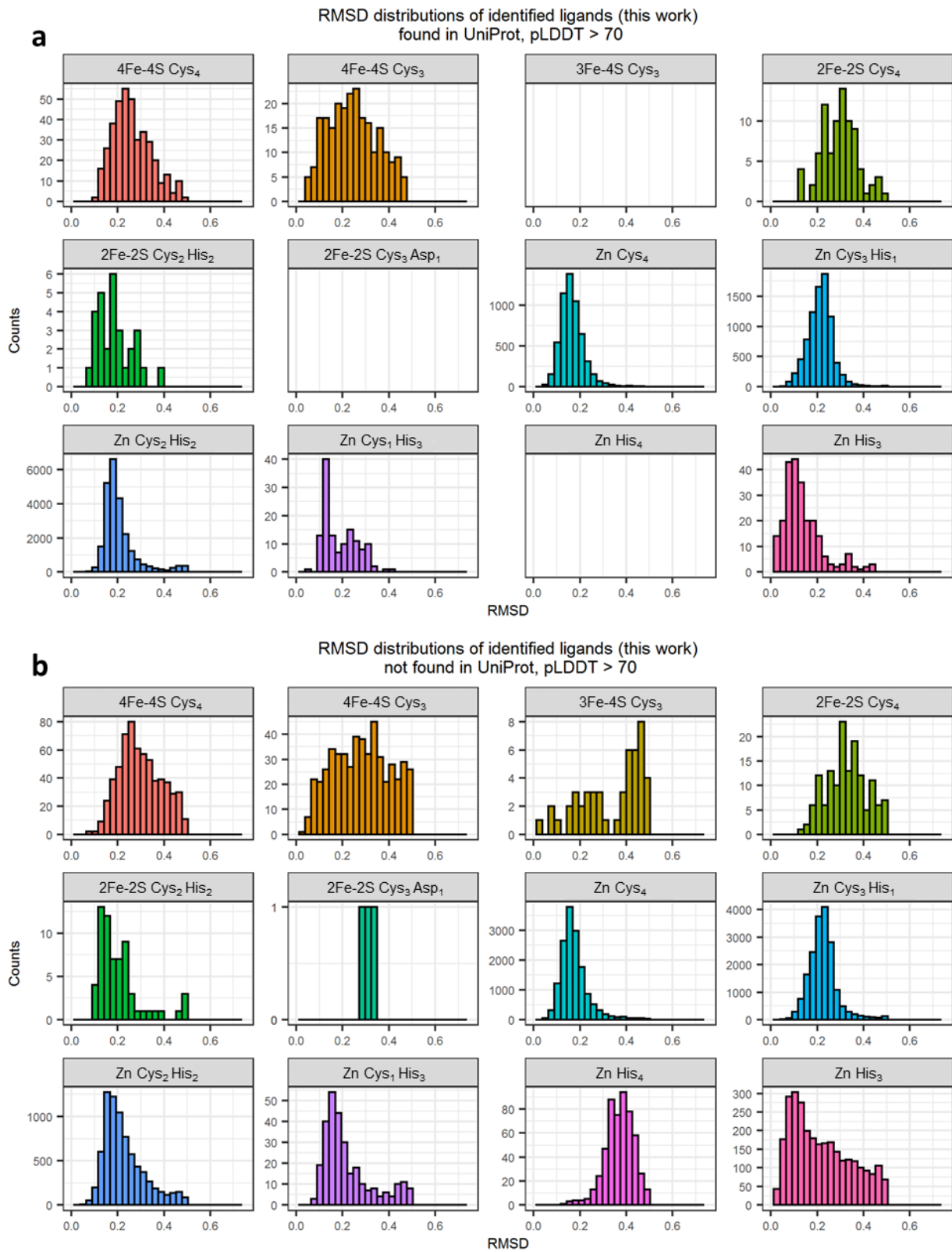


Figure S6

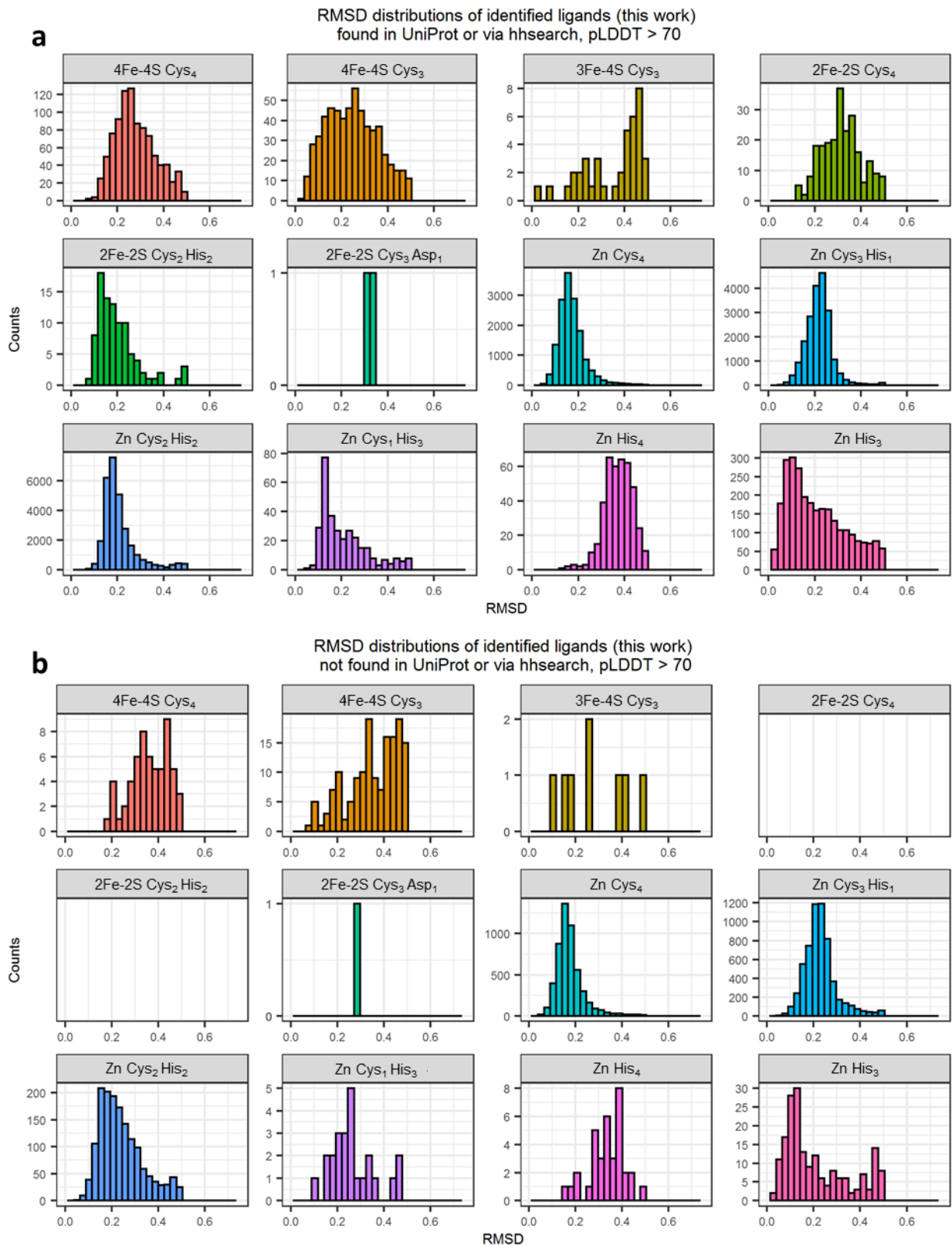


Figure S7

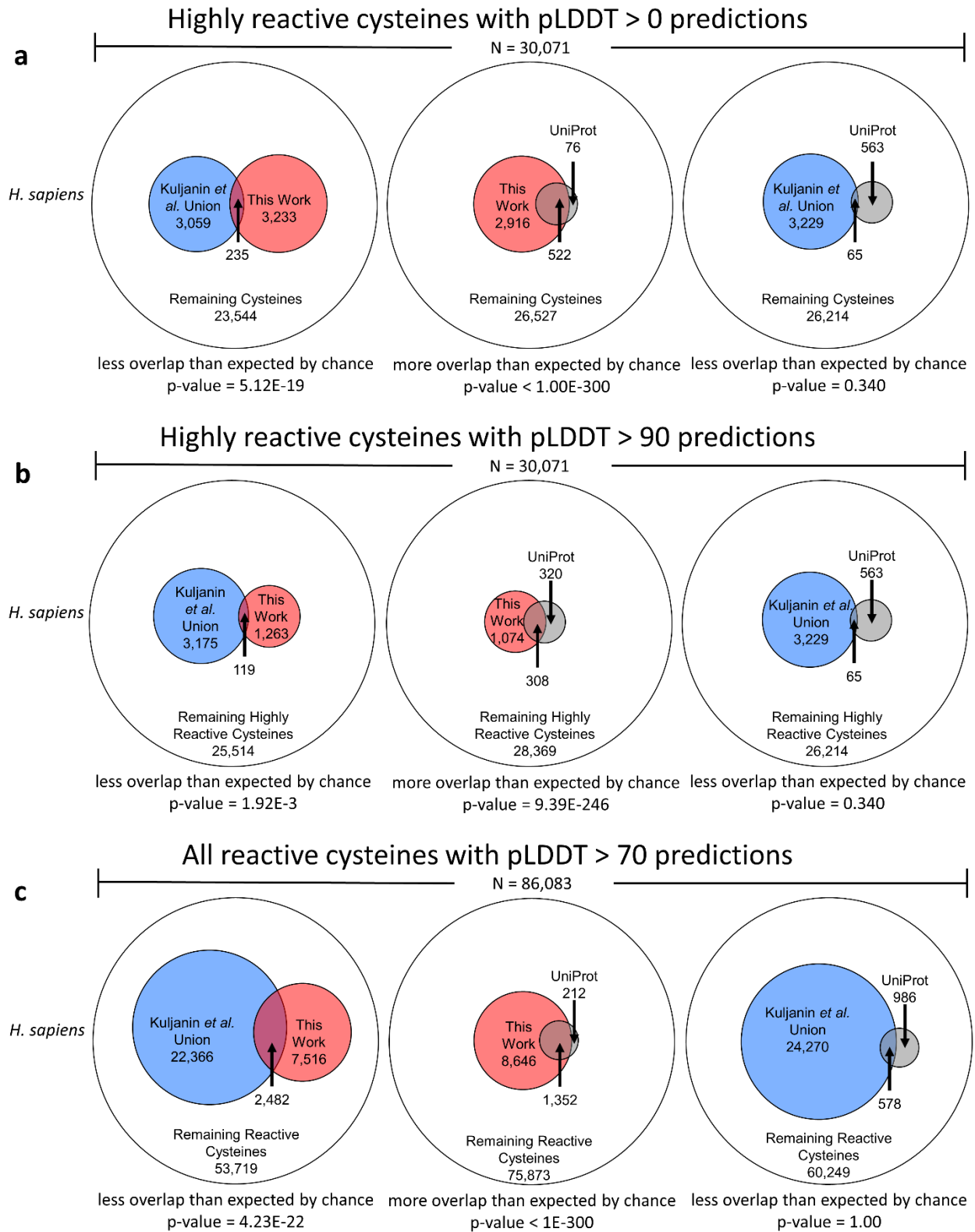


Figure S8

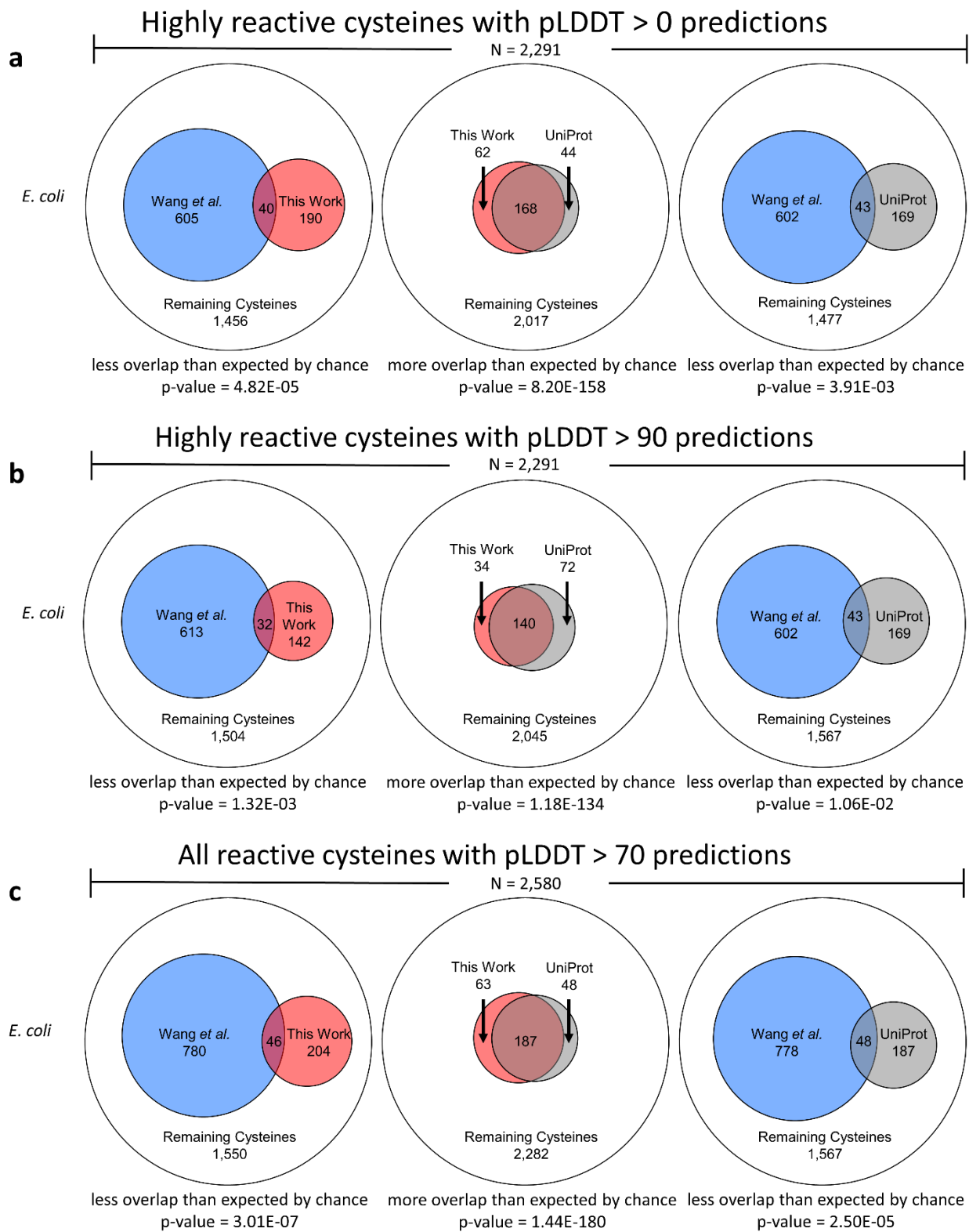


Figure S9

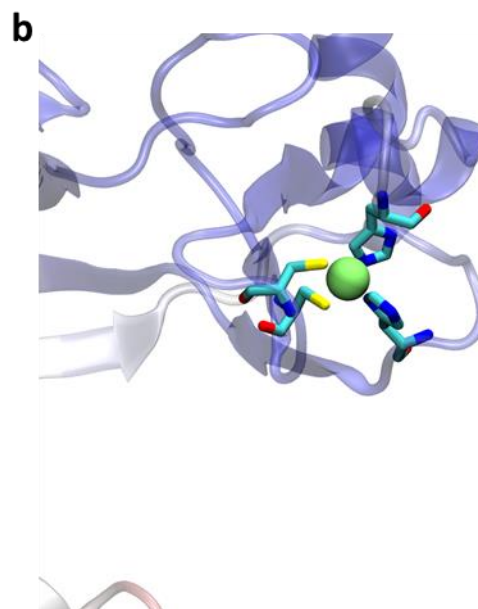
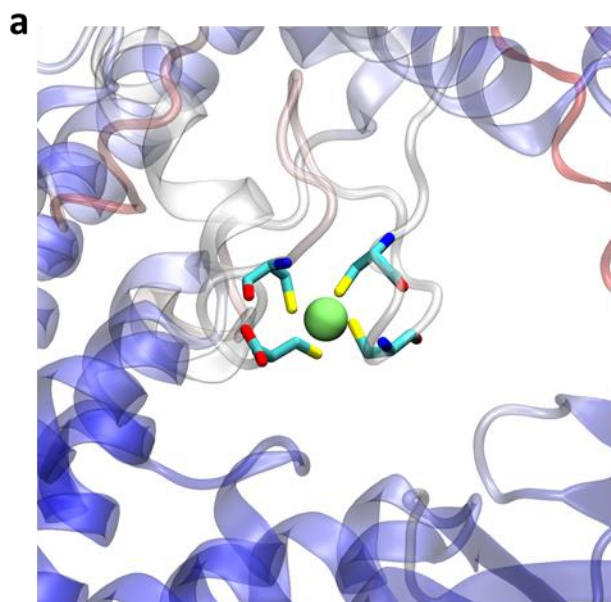


Figure S10