

Interface energetics and cavity architecture underlying cage stability in six-subunit manganese catalase

Amrita Amrita

Indian Institute of Technology Jodhpur

Srijita Acharya

Indian Institute of Technology Jodhpur

Soumyananda Chakraborti

Birla Institute of Technology and Science - Hyderabad Campus

Sucharita Dey

sdey@iitj.ac.in

Indian Institute of Technology Jodhpur

Article

Keywords: MnCatalases, ferritin, cage protein, self-assembly, evolution, interface characteristics

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11149460/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: Competing interest reported. S.D. and S.C. are members of the Editorial Board of Scientific Reports. S.D. and S.C. had no involvement in the editorial handling or peer review of this manuscript. The authors declare no other competing interests.

Abstract

Manganese catalase (MnCatalase) is a ferritin-superfamily cage protein built from only six subunits, compared to 12 or 24 subunits in ferritin protein cages. It retains the conserved four-helical ferritin core and specializes in catalysis rather than iron storage. How a six-mer can sustain cage integrity, particularly under extreme conditions, remains unresolved. We quantified the structural basis of this economy through an integrated computational analysis – interface geometry and Rosetta-based energetics, all-atom molecular dynamics, and cavity surface characterization – across curated MnCatalase, 12-mer Dps and 24-mer classical ferritin structures spanning mesophilic to extremophilic sources. Long C-terminal extensions in MnCatalase generate one lateral interface with exceptionally high buried surface area, hydrogen-bond count and interaction energy – a feature conserved across all MnCatalases irrespective of source organism and absent in Dps and ferritin – indicating that increased per-interface load compensates for reduced subunit number. Thermophilic and extremophilic cages reinforce these interfaces further through non-polar enrichment and stronger electrostatic and van der Waals interactions, and line their cavities with non-polar residues at lower surface area-to-volume ratios. These observations are consistent with a trade-off between substrate accessibility and structural rigidity.

1. Introduction

Manganese catalase (MnCatalase) is an essential enzyme found in a wide range of prokaryotes that neutralizes hydrogen peroxide through catalysis, thereby preventing hydrogen peroxide-associated cytotoxicity. They are an example of the oxidoreductase category of enzymes, where MnCatalase, along with other ferritin family of proteins, forms a relatively rare cage-like architecture (dihydromethanopterin reductase as the only other example). These proteins are quite intriguing due to their unique structural features and potential applications in nanotechnology and biocatalysis¹. MnCatalase comprises Mn as a cofactor, facilitating the catalysis of H_2O_2 into H_2O and O_2 . This property has rendered the protein useful in the medical industry for the treatment of oxidative stress, in the pharmaceutical industry for the removal of peroxide from specific chemical solutions, and in environmental biotechnology for providing oxygen during aerobic bioremediation². MnCatalases are formed by six homomeric chains that self-assemble into a cage architecture with extensive intersubunit contacts and a structural Ca ion cross-linking neighbouring chains³. Within the MnCatalase family, there exist two distinct structural classes (mesophilic and thermophilic MnCatalases) based on the organisms in which they exist. Thermophilic MnCatalases from *T. thermophilus*⁴ and mesophilic MnCatalases from *L. plantarum*⁵ and cyanobacteria like *Anabaena*⁶ are the most studied MnCatalases due to the availability of their crystal structures. Recent literature explores these antioxidant enzymes from extremophiles like Archaeon *Pyrobaculum calidifontis*⁷ and those from *Geobacillus thermopakistanensis*⁸. These catalases show high activity and thermostability making them useful candidates for industrial applications. Thermophilic catalases may have functions other than peroxide decomposition (catalase and peroxidase); meanwhile, mesophilic catalase enzymes exerts a restricted access to H_2O_2 and serves explicitly as a catalase⁹. Apart from that, these structural classes differ in their monomeric design and the organization of their Mn active site, as well as the Ca ion crosslink, which is present in mesophilic catalases but absent in their thermophilic counterpart.

The quaternary architecture of the protein shows MnCatalase as a Homo-6-mer with a 3 – 2 point symmetry. Each protein subunit comprises an N-terminus extension, a 4-helix bundle catalytic core characteristic of a Ferritin-like superfamily¹⁰, and a C-terminus tail. The 2 terminal regions (residues 1–18 and 191–260)

distinguish MnCatalases from other ferritin family proteins in having a distinctly extended conformation involving very little secondary structure and wrapping the core like a belt, thereby protecting the core and stabilizing the cage architecture⁵. The MnCatalase catalytic domain comprises a binuclear manganese metal complex consisting of one glutamate carboxylate bridge ($\mu_{1,3}$) between the metal centers. EXXH metal binding motif of the glutamate carboxylate bridge is a common feature of the proteins containing a binuclear metal complex. A separate non-coordinated glutamic acid residue lies above the metal core, forming an H-bond with the solvent molecule attached to the Mn ion¹¹. A bulge formed by a pair of Glycine residues in the middle of helix A creates an access channel connecting the catalytic manganese cluster to the solvent-filled core of the protein. The access channel is lined by hydrophilic groups that may restrict the entry of ions while allowing access to neutral species like H_2O_2 , H_2O , and O_2 ⁵.

We aimed to study the structural and functional characteristics of MnCatalase at the monomer, subunit-subunit interface and cage level, with the aim to explore the similarities and distinctions between these classes of MnCatalases- Mesophilic MnCatalases (*Lactobacillus plantarum* (LactoB) and *Nostoc* sp. PCC7120 (Anabaena)), thermophilic MnCatalases and those from extremophiles like Archaea (*Pyrobaculum calidifontis*) and recently discovered *Geobacillus thermopakistanensis*. We would also like to explore the MnCatalases as well as between 6-mer MnCatalases and 24-mer Classical ferritins (C-ferritins) in general. The study gives an insight into mechanisms of protein cage assembly at the subunit, interface, and cage level and the factors that influence it, thus giving a comparative overview of the protein cages at two different length scales¹².

2. Results

2.1 Dataset curation

Crystal structures of four well-characterised¹¹, non-mutant and native MnCatalase proteins, PDB:1jku⁵ (LactoB: *Lactobacillus plantarum*), PDB:6j42³ (Anabaena: *Nostoc* sp.), PDB: 2v8t⁴ (Thermo: *Thermus thermophilus* HB27), and PDB: 2cwl (Thermo: *Thermus thermophilus* HB8), were taken as the initial reference in the study of MnCatalases. To further expand the analysis, MnCatalases from extremophiles (Archaea: *Pyrobaculum calidifontis* and GeoB: *Geobacillus thermopakistanensis*) were also examined, and their predicted AlphaFold structures were obtained. The modelled structure showed an overall quality score using ERRAT of 95.85 for GeoB and 94.98 for Archaea. The ipTM and pTM analyses are provided in Supplementary Table 1, Supplementary Figs. 1 and 2, and the quality of the modelled structure was validated. These proteins were categorised into the mesophilic group: Lactobacillus (PDB: 1jku), Anabaena (PDB: 6j42); the thermophilic group: (PDB: 2v8t, 2cwl); and the extremophile group: Archaea and GeoB. A sequence-based analysis was undertaken on the homologs of LactoB, Anabaena, thermophilic Geobacillus, and Archaea. The homologs were obtained from a BLAST search and selected for having at least 80% sequence identity and at least 80% query coverage. For a sequence-based comparative study of MnCatalases with ferritins, the analysis was expanded to native and non-redundant ferritins obtained from our previous dataset: 18 Classical ferritins (C-ferritins), 10 Bacterial ferritins (B-ferritins), 13 Bacterioferritin (Bo-ferritins), 30 Dps (DNA-binding protein from starved cell), along with 21 LactoB, 30 Anabaena, 4 Thermo, 4 GeoB and 6 Archaea homologs¹³ (Supplementary Table 2).

2.2 Sequence and structure-based characterization of the MnCatalase monomer

The MnCatalase monomer comprises three main segments: 1) an N-terminal polypeptide, 2) a 4-helix bundle catalytic domain, and 3) an extended C-terminal tail (Supplementary Fig. 3). These long terminal extensions ensure extensive contacts with every other subunit of the protein, affecting the interfacial interactions, as described later. The 4-helix domain is well conserved across the ferritin superfamily, as observed in C-ferritin, Dps, and MnCatalases from mesophiles, thermophiles, and extremophiles. Thus, to begin the comparative analysis of MnCatalases, both within and across ferritin subfamilies, a sequence-based study was initiated using the PROTPARAM tool¹⁴. Several physicochemical properties were obtained from the tool for the monomeric sequences of MnCatalase. These properties give us an insight into the sequence characteristics of these monomers and the physico-chemical variations between them (Table 1) as well as between other ferritin subfamilies (Supplementary Table 2).

2.2.1 Secondary structure architecture

Secondary structure composition was analysed for MnCatalases and ferritin monomers using the DSSP tool. All the MnCatalase subtype proteins showed the α -helix component between 50% to 60%, and the component of turns, bends and coils between 40% to 50% and the presence of some minor β -sheets. It was observed that extremophile MnCatalases have a higher content of total structured components, α -helices and β -sheets than mesophilic MnCatalases (Table 1). While the α -helix component was most prominent across all protein types and the β -sheet was negligible, a distinction was observed with regard to the presence of minor β -sheets in MnCatalases, whereas it was completely absent in ferritins (Supplementary Table 3). Another distinction was observed in the content of coils and turns in catalases, which accounted for more than 40% of the total secondary-structure content of the protein (the fraction of turns and coils is 0.43 in mesophilic MnCatalases and 0.42 in thermophilic MnCatalases). Comparatively, the fraction of coils and turns was much lower in ferritins and Dps and was in the range (0.22–0.28) (Supplementary Table 3).

2.2.2 Amino acid residue composition

We analyzed the amino acid composition across MnCatalase proteins (Table 1) and between MnCatalases and ferritins (Supplementary Table 2). We found that mesophilic MnCatalases had 48.5% (LactoB) and 49% (Anabaena) non-polar residues. In contrast, thermophilic MnCatalases and extremophiles were found to have a higher content of non-polar residues than other proteins: 59% for thermophilic MnCatalases, 55.4% for Archaea, and 52.4% for *the MnCatalase of GeoB*. The grand average of hydropathy (GRAVY)¹⁴ values for proteins were also found to be least negative for thermophilic (-0.16) and next for the extremophile MnCatalase (-0.28 for GeoB and -0.24 for Archaea), compared to mesophilic MnCatalases (-0.61 for LactoB and -0.42 for Anabaena).

The amino acid composition of terminal extensions was also explored. The MnCatalase extensions were rich in amino acid residues A, R/N, G, Q, S, P, E, and K, which are prominently found in Intrinsically Disordered Regions (IDRs)¹⁵. These eight residues accounted for more than 55% of the total residue content in MnCatalase extensions, highlighting the terminal extensions' flexibility and polarity.

2.2.3 Structural alignment study

Structural alignment tools show weak alignment between the MnCatalase of *Lactobacillus* and those of extremophiles such as Archaea and *Geobacillus* (RMSD values of 1.02 and 2.75, respectively). As expected, we observed a stronger structural alignment between the MnCatalases of thermophiles and extremophiles (RMSD = 0.33 and 0.34, respectively), suggesting a divergence of thermophilic and extremophile proteins from that of

mesophilic proteins. Overall, we observe an evolutionary conservation of the helical backbone region between the MnCatalases, ferritins and Dps and divergence in their terminal extensions as analyzed in the previous study¹⁰. A similar observation was made across ferritin subfamilies (C-ferritin, B-ferritin, Bo-ferritin, and Dps proteins), with terminal extensions conferring structural distinctions and unique functional attributes to the proteins while conserving the core catalytic properties in the helical backbone of the superfamily¹³.

2.2.4 Phylogenetic study

A circular phylogenetic tree was generated using the Maximum Likelihood method, and an analysis was made of the phylogenetic distance between subfamilies. The MnCatalases seem to be of early origin, with Archaea and thermophilic MnCatalases at close phylogenetic distance and Anabaena and Lactobacillus MnCatalases also at close phylogenetic distance. Archaea and thermophilic MnCatalase seem to have evolved early as observed in the early branch of the phylogenetic tree (Fig. 1A). This is validated by Klotz and Loewen who suggested that MnCatalases (Mn-catalases) have very ancient origins, possibly predating the rise of atmospheric oxygen¹⁶. Geobacillus MnCatalases, despite being an extremophile, are evolutionarily closer to mesophilic MnCatalases than Archaea and thermophilic MnCatalases.

A. Phylogenetic Tree of ferritin superfamily in circular layout that includes C-ferritin (purple), B-ferritin (blue), Bo-ferritin (orange), Dps proteins (green), mesophilic (LactoB and Anabaena) MnCatalase (pink), Archaea MnCatalase (navy blue), *Geobacillus thermopakistanensis* MnCatalase (mustard) and thermophilic MnCatalase (grey). **B** Cage structure representation of. Mesophilic MnCatalase (PDB:1jku) in pink, Thermophilic MnCatalase (PDB: 2v8t) in grey, Anabaena MnCatalase (PDB:6j42) in greenish yellow, *Geobacillus thermopakistanensis* MnCatalase in yellow mustard and Archaea MnCatalase in blue. **C.** Structural alignment of a single subunit between MnCatalases of Lactobacillus and *Thermus thermophilus*, between Lactobacillus and Anabaena, between Lactobacillus and Geobacillus, and between Lactobacillus and Archaea. Structural alignment between MnCatalases of *Thermus thermophilus* and Anabaena, between *Thermus thermophilus* and Geobacillus, and between *Thermus thermophilus* and Archaea, with their Root Mean Square Deviation. **D.** Barplot depicting Grand Average of Hydropathicity (GRAVY), **E.** % Non polar residues and **F.** % (α -helix and β -sheet) for MnCatalase proteins.

*symbol denotes the significance level calculated using Welch's 2 sample t-test between Lactobacillus MnCatalase and Thermophilic, Geobacillus and Archaea MnCatalase. In all these cases $p < 0.001$; **The symbol denotes the significance level calculated using Welch's 2 sample t-test between Anabaena MnCatalase and Thermophilic, Geobacillus and Archaea MnCatalase. In the case of comparison with thermophilic MnCatalase, $p < 0.001$, and in other cases $p < 0.05$.

Table 1
Sequence-based physico-chemical properties of MnCatalase monomers

	Mesophilic MnCatalase		Thermophilic MnCatalase	Extremophile MnCatalase	
	Lactobacillus	Anabaena		Geobacillus thermopakistaniensis	Archaea
Assembly	6-mer	6-mer	6-mer	6-mer	6-mer
#homologs	21	30	4	4	6
MW (Da)	29779 ± 376	25894.59 ± 723	33215 ± 144	33370.45 ± 138	32331.3 ± 2480
Percent (Non-polar residues) (%)	48.5 ± 1.7	49 ± 0.5	59 ± 2.5	52.4 ± 1.2	55.4 ± 2.1
GRAVY	-0.61 ± 0.07	-0.42 ± 0.04	-0.16 ± 0.06	-0.28 ± 0.06	-0.24 ± 0.04
Isoelectric Point	4.95 ± 0.17	5.00 ± 0.19	5.3 ± 0.16	5.91 ± 0.17	6.18 ± 0.4
Top 3 AA	GLU, LEU,ALA	LEU,ALA, GLU	ALA, LEU, GLU	LEU, ALA, GLY	LEU, ALA, PRO
Instability Index	34.6 ± 2.9 (stable)	28.2 ± 4.8 (stable)	43.9 ± 4.3 (stable)	39.84 ± 3.3 (stable)	39.3 ± 3.2(stable)
Fraction (α helix + β sheet)	0.57	0.55	0.58	0.61	0.68
<MW>: Molecular weight in Daltons					

<GRAVY > ¹⁴: The Grand Average of Hydropathy calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence.

* Helix fraction, Turn fraction, and β-Sheet fraction data are obtained from the DSSP tool ¹⁷ using PDB structures of the proteins.

2.3 Characterization of the Subunit Interfaces of MnCatalases

2.3.1 Interface geometry

MnCatalase was found to have an interface geometry markedly different from that of the other ferritins (Supplementary Fig. 4). Ferritins are seen to have a 3-chain repetitive unit in their cage assembly comprising an anti-parallel interface (IntA) and a couple of perpendicular interfaces – IntB and IntC (Supplementary Fig. 4B). On the other hand, MnCatalase subunits are arranged laterally with two lateral interfaces (we name IntL and IntLX) (Fig. 2A). IntL and IntLX are distinguished based on the existence of an antiparallel β sheet contributed by both chains forming the interface. While IntL lacks the β sheet, these are part of IntLX and hence easily differentiated. The subunits are also angularly aligned such that a chain interacts with the third chain on either side, which we name IntN (Fig. 2A). Interface geometry analysis shows mesophilic MnCatalase to have a remarkably large

buried surface area (BSA) in lateral interface IntL and thermophilic MnCatalase to have a larger BSA in IntLX. In both cases, MnCatalase shows one of the lateral interfaces to be more prominent than either the C-ferritin or Dps interfaces (Supplementary Fig. 4C). The large lateral interface is attributed to the long C-terminus extension that wraps around the sideward subunit like a belt, increasing the BSA, Local atomic density (Ld), and thereby enhancing subunit-subunit interactions. The interfacial characteristics of Archaea and GeoB are similar to those of thermophilic MnCatalase, with Int LX being the more prominent interface, while Anabaena MnCatalase interface properties are similar to those of LactoB MnCatalase, with IntL as the more dominant interface (Table 2). The divergence across MnCatalase extends to variations in interfacial interactions as well.

2.3.2 Interface electrostatics

The number of H-bond interactions calculated for MnCatalases correlates with the interface BSA and is significantly high for LactoB and similar for thermo and extremophile MnCatalase. Interestingly, the numbers are lower for Anabaena MnCatalase (Table 2). Other interfacial electrostatics, including van der Waals attraction, Electrostatic energy and ΔG_{int} (Rosetta Energy Unit), are markedly higher for thermophilic MnCatalase and extremophile MnCatalase than those of mesophilic MnCatalase (Lactobacillus and Anabaena) (Table 2). Another clear distinction was observed between MnCatalases and ferritins. MnCatalases were seen to have a higher count of stable H-bond interactions for one of the lateral interfaces as compared to any of the interfaces of C-ferritin and Dps (Supplementary Table 4). This property, along with a larger BSA, could have contributed to stabilising the interface with lower ΔG_{int} (Rosetta Energy Function) for the MnCatalase lateral interface as compared to the C-ferritin and Dps interfaces (Supplementary Fig. 4E). However, the total contribution of BSA and ΔG for the cage, i.e., ΔG_{ass} , is lower in MnCatalase than in the C-ferritin (Fig. 3D), highlighting the contribution of having a higher number of chains and interfaces (more than the individual interfaces) in stabilizing the quaternary structure.

A. Cylindrical cartoon representation of LactoB MnCatalase (PDB:1jku) showing three of the six chains comprising lateral interfaces IntL and IntLX and angular interface IntN. The distinction between IntL and IntLX is based on the presence of an antiparallel β -sheet contributed by each of the chains forming the interface (IntLX) and absence in IntL. Barplot distribution of **B.** Delta G of interface (ΔG_{int}) (Rosetta Energy Unit), **C.** Interface Electrostatic Energy in kcal/mol and **D.** Interface Van der Waals attraction in kcal/mol for the three interfaces (IntL, IntLX and IntN) of Mesophilic (Lactobacillus and Anabaena), Thermophilic (*Thermus thermophilus*), and extremophile (*Geobacillus thermopakistaniensis* and Archaea). In all the barplots, Lactobacillus MnCatalase interfaces data is shown in the shades of pink, Anabaena data is shown in the shades of greenish yellow, Thermophilic MnCatalase interfaces data is shown in the shades of grey, Geobacillus is shown in the shades of mustard yellow and Archaea MnCatalase interfaces data is shown in the shades of blue. AF denotes the AlphaFold structures of *Geobacillus thermopakistaniensis* (GeoB) and Archaea (*Pyrobaculum calidifontis*)

*, ** and *** symbols denote the significance level ($p < 0.05$) calculated using Welch's 2 sample t-test between Thermophilic and mesophilic MnCatalases, GeoB and mesophilic MnCatalases and Archaea and mesophilic MnCatalases.

Table 2
Physico-chemical and geometrical characterization of MnCatalase subunit interfaces

	Interface type	Mesophilic MnCatalase		Thermophilic MnCatalase (2)	Extremophile MnCatalase	
		Lactobacillus MnCatalase	Anabaena MnCatalase		Geobacillus thermopakistanensis (AF)	Archaea (AF)
	IntL	9518 ± 58	6017 ± 68	2450 ± 15	2387 ± 2	2355 ± 3
<Buried surface Area (A ²)>, BSA	IntLX	3344 ± 46	3807 ± 77	7764 ± 107	7826 ± 4	8499 ± 1
	IntN	3061 ± 57	1236.7 ± 89	2061 ± 19	2155 ± 3	2186.5 ± 4
	IntL	979 ± 6.4	608 ± 7.1	240.8 ± 1.3	211 ± 0.8	658 ± 0.5
<Number of Interface atoms>	IntLX	340 ± 3.9	380 ± 5.7	766.7 ± 5.8	803.3 ± 3.4	2546 ± 1.9
	IntN	310 ± 5.2	120 ± 8	212 ± 4.6	214 ± 1.3	1328 ± 1.5
	IntL	239 ± 1.9	161 ± 0.9	66 ± 1	66 ± 0	60 ± 0
<Number of Interface residues>	IntLX	89 ± 1.2	93.3 ± 1.9	200.7 ± 1.6	197.7 ± 0.9	210.3 ± 0.5
	IntN	95 ± 2.1	39 ± 2.2	59.9 ± 0.9	59.8 ± 0.4	62.5 ± 0.5
	IntL	0.49	0.49	0.46	0.7	0.66
<f non-polar>	IntLX	0.57	0.67	0.43	0.53	0.61
	IntN	0.46	0.41	0.33	0.53	0.42
	IntL	24.6 ± 0.2	24.6 ± 0.4	22 ± 0.1	19.7 ± 0.1	21.2 ± 0.1
<Local Atomic Density > Ld	IntLX	23.8 ± 0.1	26.9 ± 0.1	31 ± 0.3	32 ± 0.2	31.8 ± 0.1
	IntN	17.1 ± 0.4	12.3 ± 0.7	22.7 ± 0.3	22.6 ± 0.2	22.6 ± 0.2
	IntL	61 ± 19.9	18 ± 0	5.5 ± 1.5	4 ± 0	6.0 ± 0

	Interface type	Mesophilic MnCatalase		Thermophilic MnCatalase (2)	Extremophile MnCatalase	
		Lactobacillus MnCatalase	Anabaena MnCatalase		Geobacillus thermopakistanensis (AF)	Archaea (AF)
<Number of H- bonds>	IntLX	17 ± 0.7	18.3 ± 1.7	54.5 ± 0.5	33.7 ± 0.5	41.7 ± 0.5
	IntN	15 ± 5	7.3 ± 0.5	10.3 ± 0.4	18 ± 0	15.2 ± 0.9
	IntL	-348.4 ± 7.6	-158.6 ± 4.2	-72.95 ± 0.65	-68.5 ± 0.3	-90.4 ± 0.6
<ΔG> (Rosetta energy unit)	IntLX	-86.3 ± 0.8	-108.4 ± 2.2	-281.55 ± 3.15	-189.5 ± 2.4	-311.7 ± 0.3
	IntN	-49.2 ± 25.2	-21.6 ± 2.9	-65.85 ± 1.05	-59.8 ± 1.5	-55.6 ± 0.4
	IntL	0.78	0.7	0.74	0.6	0.7
<Sc>	IntLX	0.64	0.7	0.72	0.7	0.7
	IntN	0.68	0.6	0.75	0.8	0.8
	IntL	-951.9 ± 1.6	-664.9 ± 3.1	-950.4 ± 15	-1029.6 ± 0	-970.6 ± 1.2
<Elec> (kcal/mol)	IntLX	-837.1 ± 15.5	-638.8 ± 13.6	-1106.5 ± 17	-1156.1 ± 1.1	-1132.1 ± 2.9
	IntN	-802.6 ± 27.1	-588.2 ± 12.2	-970.3 ± 17.2	-1053.1 ± 1.3	-984.5 ± 0.6
	IntL	-3218.7 ± 8.5, 515.6 ± 9.4	-2718.8 ± 16.0, 531.1 ± 6.0	-3582.8 ± 9, 986.9 ± 73	-3543.8 ± 0.9, 934 ± 6.2	-3511.2 ± 2.4, 834.2 ± 4.8
<van der Waals energy> attr, rep (kcal/mol)	IntLX	-2939.2 ± 10.5, 740.9 ± 80.2	-2632.3 ± 15.6, 554.7 ± 2.4	-3828.5 ± 7.3, 851.6 ± 14	-3804.8 ± 1.2, 887.9 ± 6.7	-3796.9 ± 3.2, 734.8 ± 2.6
	IntN	-2900.5 ± 38.6, 1180 ± 137.6	-2528.1 ± 4.7, 612.7 ± 5.5	-3594.3 ± 20.5, 1015.1 ± 86.4	-3574.2 ± 1.3, 954 ± 5.7	-3530.1 ± 1.4, 835.3 ± 3.1

<ΔG>: Estimated energy of a particular conformation of a biomolecule calculated using the computational model Rosetta Energy Function (REF) 15¹⁸ in Rosetta energy units, which includes a weighted average of energy terms associated with different molecular conformations. The model uses experimental data of van der Waals interactions, electrostatics, hydrogen bonding, and solvation effects to obtain the energy of a biomolecule at different conformations.

<Elec>: Electrostatic energy (kcal/mol) calculated using REF 15¹⁸

<van der Waals energy > rep represents repulsive energy (kcal/mol); attr represents attractive energy (kcal/mol) calculated using REF 15¹⁸

<Sc>: Shape complementarity of the interface calculated using REF 15¹⁸, Interface data for C-ferritins and Dps proteins are taken from the author's previous paper¹³

< $f_{\text{non-polar}}$ >: fraction non-polarity represents the fraction of non-polar atoms present at the particular interface, averaged over all interfaces of the same type

2.4 Dynamic analysis of the structures of MnCatalase and ferritins using MD simulation

The MD simulation analysis through RMSD and RMSF graphs depicted structural stability of the MnCatalase proteins along with C-ferritin throughout the simulation time and throughout the amino acid range (Fig. 3A and 3B). It validates the ΔG_{ass} numbers obtained from static analysis with MnCatalases ranging from – 100 to -200 kcal/mol and Dps and C-ferritin within the range – 150 to -300 kcal/mol (Fig. 3D). The Dps structure showed initial instability, possibly owing to the absence of Fe ions in the force field (Supplementary Fig. 5). We were able to parameterise Mn ions in MnCatalases because those ions did not form coordination bonds with amino acids. The RMSF plot showed some fluctuation between residues 75 and 105. This fluctuating region between amino acids 75 and 105 depicts the flexible BC loop as seen in the protein visualization tool. BC loop is considered among the most mobile regions of ferritin and is involved in receptor recognition and receptor interaction^{19,20}. ConSurf tool²¹ confirmed the variability of the BC loop by giving a low conservation score to that region for all the MnCatalases (Supplementary Fig. 6) in the protein conservation scale. The radius of the gyration figure showed consistency throughout the time range (Fig. 3C). It shows retention of the protein's compactness even in a dynamic environment, confirming structural stability. Table 3 lists H-bonds present in > 95% of the simulation frames, contributing to stability even under dynamic conditions. These (stable) H-bonds are higher in number in LactoB, Anabaena, thermophilic and archaea MnCatalases and are comparatively less in GeoB MnCatalase (Table 3). This could contribute to the low ΔG_{ass} values for GeoB MnCatalase compared with other 6-mer counterparts. Although the number of stable H-bond interactions is less in C-ferritin as compared to other MnCatalases, it is compensated by the large number of chains that stabilise the overall 3D structure of the protein (Fig. 3D). Thus, although C-ferritin interfaces contain fewer hydrogen bonds than those of manganese catalases, the larger number of interacting subunits in its 24-mer assembly provides greater overall stabilisation of the quaternary architecture than the 6-mer assembly of manganese catalases.

Line graph of **A.** Backbone Root Mean Square Deviation (RMSD) in nm for 200ns, **B.** Root Mean Square Fluctuation (RMSF) for the residues and **C.** Radius of Gyration (nm) for MnCatalases and C-ferritin assembly. LactoB graph is denoted in pink, Thermophilic MnCatalase in grey, Anabaena in yellow, *Geobacillus thermopakistaniensis* in Mustard, Archaea in teal and C-ferritin in dark blue. (D) Barplot depicting ΔG_{ass} obtained using static calculation, the colour code as pink for *Lactobacillus* MnCatalase, lime green for Anabaena, grey for thermophilic MnCatalase, mustard yellow for *Geobacillus* MnCatalase and Blue for Archaea MnCatalase.

Table 3

List of stable H-bonds (between A chain and neighbouring chains) present in at least 95% of the MD simulation frames

Mesophilic (PDB:1jku)	Thermophilic (PDB:2v8t)	Extremophile		C-ferritin (PDB:1r03)
		Archaea (AF)	GeoB(AF)	
A0005-THR B0001-MET	A0141-SER F0164- ARG	A0277-GLU C0161-ARG	A0131-GLY B0129-TRP	A0085-ILE W0085- ILE
A0128-ALA B0126- TRP	A0280-GLU D0164- ARG	A0005-ILE B0001- MET	A0129-TRP B0131-GLY	A0083-GLN W0032-TYR
A0134-SER F0157- ARG	A0164-ARG F0141- SER	A0130-TRP B0132-GLY	A0054-LYS B0269- ASP	A0085-ILE W0085- ILE
A0177-ARG F0215-VAL	A0275-GLY B0051- ARG	A0001-MET B0005-ILE	A0276-GLU E0160-ARG	A0032-TYR W0083-GLN
A0215-VAL F0177- ARG	A0005-ILE B0001- MET	A0003-LEU B0003-LEU	A0006-ILE B0002- VAL	A0067-GLU W0039-TYR
A0126-TRP B0128- ALA	A0004-ARG B0047- SER	A0003-LEU B0003-LEU	A0002-VAL B0006-ILE	A0060-ARG W0067-GLU
A0003-LYS B0003- LYS	A0051-ARG B0275- GLY	A0111-TYR C0124-ASP	A0052-ARG B0116-ALA	A0174-THR L0168- TYR
A0005-THR B0001-MET	A0003-LEU B0003- LEU	A0138-SER F0161-ARG	A0109-THR E0124-SER	A0039-TYR W0074-ASN
A0007-LYS B0001- MET	A0236-PHE B0004- ARG	A0135-ILE F0033- ARG	A0160-ARG C0276-GLU	A0168-TYR N0174- THR
A0004-HIS C0219- PHE	A0003-LEU B0003- LEU	A0132-GLY B0130-TRP	A0160-ARG C0115-THR	A0084-ASP W0087-LYS
A0218-ASN F0061-ASP	A0001-MET B0005- ILE	A0120-ALA B0049-ASN	A0001-MET B0076-GLU	A0039-TYR W0067-GLU
A0004-HIS C0220- SER	A0164-ARG E0280- GLU	A0033-ARG F0135-ILE	A0116-ALA B0052-ARG	A0074-ASN W0039-TYR
A0057-ASP B0006- ARG	A0051-ARG B0277- ASP	A0019-ASP E0105-LYS	A0005-ARG B0068-THR	A0042-SER N0146- LYS
A0212-PHE F0177- ARG	A0053-LYS B0272- PHE	A0005-ILE B0001- MET		A0079-ARG W0044-ASP
A0201-PRO F0152- ARG	A0051-ARG B0120- GLY	A0007-ARG B0001-MET		A0146-LYS L0042- SER
A0137-LEU C0264- HIS	A0272-PHE B0053- LYS	A0049-ASN B0120-ALA		A0044-ASP W0079-ARG
A0006-ARG C0221-ASP	A0284-GLU D0212- LYS	A0124-ASP E0111-TYR		
A0250-GLY F0008- LEU	A0004-ARG B0236- PHE	A0108-ARG E0110-THR		

Mesophilic (PDB:1jku)	Thermophilic (PDB:2v8t)	Extremophile		C-ferritin (PDB:1r03)
		Archaea (AF)	GeoB(AF)	
A0061-ASP F0218-ASN	A0283-PRO B0144-LEU	A0164-GLU E0101-LEU		
A0220-SER F0061-ASP	A0005-ILE B0001-MET	A0004-ARG B0234-SER		
F0224-GLY A0057-ASP	A0212-LYS E0284-GLU	A0004-ARG B0233-PHE		
	A0120-GLY B0051-ARG	A0161-ARG E0115-ASN		
	A0049-ASN B0082-ASN	A0049-ASN B0084-ASN		
	A0004-ARG B0237-SER	A0234-SER B0004-ARG		
	A0047-SER B0004-ARG			
	A0082-ASN B0049-ASN			
	A0237-SER B0004-ARG			

2.5 Cavity analysis of MnCatalase cage with respect to other cage proteins of ferritin superfamily like Dps and C-ferritin

2.5.1 Cavity geometry

Cavity study of MnCatalase protein shows that the cavity volume sum for LactoB Mn Catalase ($16644 \text{ \AA}^3$), Anabaena MnCatalase ($16087.5 \text{ \AA}^3$), thermophilic MnCatalase ($17242 \text{ \AA}^3$) and extremophile MnCatalase (Archaea = $23331 \text{ \AA}^3$ and GeoB = $30631.9 \text{ \AA}^3$), as much smaller than that of Dps proteins ($45346.5 \text{ \AA}^3$) and C-ferritins ($61217.5 \text{ \AA}^3$). Extremophile MnCatalases have larger cavity volume sums than both mesophilic and thermophilic MnCatalases but less than that of Dps and C-ferritin. A similar trend is observed with respect to the sum of cavity surface area with total surface area in LactoB MnCatalase ($14930 \text{ \AA}^2$), Anabaena MnCatalase ($14184.8 \text{ \AA}^2$), thermophilic MnCatalase ($15322.1 \text{ \AA}^2$) and extremophile MnCatalase (Archaea = $19468.5 \text{ \AA}^2$ and GeoB = $24799.5 \text{ \AA}^2$) being lower than Dps proteins ($32885.3 \text{ \AA}^2$) and C-ferritin ($50446.1 \text{ \AA}^2$). However, when we calculated the surface area/volume ratio, we found the fraction to be higher in MnCatalase (LactoB = 0.9, Anabaena = 0.88, Thermo = 0.89) than in Dps proteins (0.73) and C-ferritin (0.86) (Table 4, Fig. 4D), explaining a higher surface area exposed to catalysis in MnCatalase than in C-ferritin and Dps protein. We observed surface area/volume ratio in extremophile Mncatalases (Archaea = 0.83 and GeoB = 0.81) to be much lesser than mesophilic and thermophilic MnCatalases highlighting more compactness and less exposure to the external environment. These results give us an insight into the cavity architecture disparity between MnCatalase, Dps, and C-ferritin. MnCatalase lacks a central cavity and has multiple small cavities, possibly increasing its cavity surface area/ cavity volume ratio, which would contribute to its major function of catalysis. On the other hand, Dps has to

accommodate a central cavity like C-ferritin despite only 12 chains in place of 24 chains; hence, the possibility of its cavity surface area/ cavity volume ratio (0.73) to be distinctly lower than MnCatalases and C-ferritins.

Table 4
Cavity geometry of MnCatalase, Dps proteins and C-ferritin cages

	Mesophilic MnCatalase		Thermo MnCatalase (PDB:2v8t, 2cwl)	Extremophile MnCatalase		Dps	C- ferritin
	LactoB (PDB:1jku)	Anabaena (PDB:6j42)		Archaea (AF)	GeoB (AF)		
Number of chains	6	6	6	6	6	12	24
MW (kDa)	180	160	180	200	200	225	475
Protein diameter (Å)	80	90	80	80	80	90	120
Cavity volume (CV)(Å³)	16644	16087.5	17242.3 ± 3984.7	23331	30631	45346± 14689	61217± 44410
Cavity surface area (CSA)(Å²)	14930	14184.8	15322.1 ± 2448.5	19468	24799	32885± 10388	50446± 30073
Cavity surface area / Cavity vol ratio (CSA/CV) (Å⁻¹)	0.9	0.88	0.9	0.83	0.81	0.73	0.86
Cavity Lining Polarity (%NP)	47.9%	48.6%	53.1%	52.4%	48.4%	48.5%	45.6%

2.5.2 Cavity polarity

Polarity of the cavity surface lining was analyzed for MnCatalases (LactoB PDB:1jku, Anabaena PDB:6j42, Thermo PDB:2cwl, Archaea AlphaFold structure, GeoB AlphaFold structure, Dps PDB:1ji5 and C-ferritin PDB:1h96. It was observed that Mesophilic MnCatalases (LactoB and Anabaena) had a higher proportion of polar residues in the cavity lining along with Dps and C-ferritin. On the other hand, thermophilic MnCatalase and Archaea Mn Catalase have higher proportion of non polar residues in the cavity lining as compared to mesophilic mn catalases, Dps and ferritin. Surprisingly, GeoB MnCatalase has a higher proportion of polar residues despite being an extremophile (Table 4). A higher proportion of smaller cavities and presence of non-polar residues in the cavity lining could confer an evolutionary advantage to thermophilic MnCatalase and archaea MnCatalase by providing it with a tighter and more compact core.

2.5.3 Catalytic site placement

Placement of catalytic sites was analyzed with respect to the substrate entry point and interfaces. Dps and C-ferritin proteins having larger cages denoted greater distance between cavity entry point and catalytic site than the MnCatalases, explaining one reason for faster catalysis in MnCatalases ^{22]} than in Dps and ferritin. The catalytic site was closer to the lateral interfaces than angular interfaces across all MnCatalases (Fig. 4E).

A. Cylindrical cartoon figure showing C-ferritin's (PDB:1r03) major central cavity in purple, Dps's (PDB:1ji5) major central cavity in green and one of the minor cavities of Lactobacillus MnCatalase (PDB:1jku) in pink, also showing the absence of a central cavity. Barplot depicting **B.** Cavity counts with respect to volume bins. **C.** Percent polar and non-polar residues in the cavity lining using the colour code as pink for Lactobacillus MnCatalase, lime green for Anabaena, grey for thermophilic MnCatalase, mustard yellow for Geobacillus MnCatalase, and blue for Archaea MnCatalase. Green for Dps and purple for C-ferritin. **D.** Barplot depicting Cavity Surface Area/ Volume ratio for MnCatalases with colour code as in B and Dps in green and C-ferritin in purple. **E.** Barplot depicting the distance between COM of catalytic site and COM of interfaces (in Å) with colour code as in B. **F.** Distance between metal entry point and catalytic site (in Å) for proteins color coded as in C.

3. Discussion

The study highlights a clear structural distinction in the secondary structure architecture within MnCatalases as well as between MnCatalases and other ferritin subfamilies. Within MnCatalases, the presence of a higher content of structured components, such as α -helices and β -sheets, in extremophiles could contribute to their enhanced structural rigidity. Also, presence of large content of coils and turns in 6-mer MnCatalases as compared to 12-mer Dps and 24-mer C-ferritin could help MnCatalases in forming a compact cage with fewer building blocks²³. These amino acid segments are known to bring distant parts of the protein closer together, helping the protein fold into a compact three-dimensional shape²⁴. These loops and turns could also confer flexibility to the chains for chain-chain interactions as well as for catalytic operations by allowing for conformational adjustments²⁵. The amino acid composition study of the MnCatalase monomers highlight the presence of higher content of non polar residues in thermophilic MnCatalases and extremophiles compared to its mesophilic counterpart. It conforms with the existing knowledge that higher content of hydrophobic residues is associated with increased protein thermostability²⁶. The hydrophobic core is known to provide a rigid structural framework and is the primary driving force for protein folding and maintenance of folded state²⁷. Moreover, a cooperative effect between hydrophobic interactions and interactions like salt bridges, van der Waals interactions and hydrogen bonding together often determine the protein's resistance to thermal unfolding²⁸. Apart from the variability in polarity, another divergence has been observed in MnCatalases contributed by the variable terminal extensions that confer distinction and functional uniqueness to the proteins while the 4-helical catalytic core stays conserved within MnCatalases as well as across ferritin subfamilies.

The interface analysis emphasizes the role of long C-terminal extension of MnCatalases in bringing the chains together with one of the lateral interfaces showcasing higher BSA, H-bond interactions and ΔG_{int} as compared to Dps and C-ferritins. We believe that, as MnCatalase has only six subunits in comparison to ferritin and Dps, which has 24 subunits, the contribution of each of the MnCatalase interfaces in stabilizing the structure becomes very important. Hence, we observe that one of the lateral interfaces in MnCatalase has exceptionally high BSA, Local atomic density, H-bond number and ΔG_{int} , possibly playing a role in stabilizing the interface and generating cage architecture in MnCatalases. All manganese catalases, irrespective of the organism in which they occur, exhibit these properties (Table 2, Supplementary Table 4). It supports the idea that the existence of these long lateral interfaces caused by long C-terminal extensions in manganese catalases reflects not only structural adaptation, functional selection but also evolutionary drift from other ferritin subfamilies. A previous study on the ribonucleotide reductase (RNR) subfamily within the ferritin superfamily also highlights the long C-terminal extension as an evolutionary innovation that serves as a structural or functional module that diversifies protein

properties without altering the ancestral catalytic mechanism²⁹. Also, a clear distinction was observed in the interface electrostatics between thermophilic and extremophile MnCatalase, on the one hand, and mesophilic MnCatalase, on the other. Thermophilic and extremophile MnCatalase showcases higher van der Waals attraction, Electrostatic energy, and ΔG_{int} (Rosetta Energy Unit), underscoring the structural features that support cage architecture in extreme environments. This data underscores the importance of non-covalent interactions in strengthening subunit interfaces across all MnCatalases, particularly in thermophilic and extremophile MnCatalases, thereby ensuring the stability of cages with just 6 subunits at high temperatures.

The simulation analysis emphasized the overall stability of the MnCatalase proteins in dynamic conditions. The RMSF fluctuation in the 75–105 region signifies the mobile BC loop region which is flexible and dynamic and plays a role in receptor identification and interaction. Also, a list of very stable H-bonds (present in > 95% of the simulation frames) could give an insight into the regions that stabilize the MnCatalase cage even in a dynamic environment.

The cavity analysis study of MnCatalase shows the total cavity volume of the cage to be smaller in MnCatalase and much larger in Dps and C-ferritin. This is expected as MnCatalase is composed of only 6 chains while Dps has 12 chains and C-ferritin has 24 chains enhancing the Dps and C-ferritin cavity to have a larger enclosed space and conferring them the ability to store Fe ions. Conversely, the cavity surface area by volume ratio (SA/V) is higher for mesophilic and thermophilic MnCatalases as compared to Dps and C-ferritins. The high surface area could expose active sites in MnCatalases, facilitating the catalytic reaction. Interestingly, extremophile MnCatalases (Archaea and GeoB) have a low SA/V ratio, possibly to enhance compactness and reduce exposure to the external environment. Polarity analysis of the cavities underscored a higher proportion of non-polar residues in the thermophilic and Archaea MnCatalase, highlighting the hydrophobic nature of the cavity surface in thermophilic proteins. The cavity polarity may also influence the diffusion of peroxide molecules through the MnCatalase cage. Polar or hydrophilic residues lining the channel can stabilise peroxide molecules through hydrogen bonding and electrostatic interactions, facilitating their movement toward the active site³⁰. On the other hand, highly hydrophobic cavities may reduce favorable interactions with peroxide molecules and alter diffusion kinetics. Thus, while LactoB and Anabaena MnCatalase have polar cavity linings with stronger substrate interactions via H-bonds, thermophilic and extremophile MnCatalases have less reactive cavity surfaces, focusing more on thermostability and structural integrity in extreme environments. These analyses give an insight into the trade-off between enzyme activity and structural rigidity to ensure optimal functional utility.

4. Methods

4.1 Dataset

The RCSB Protein Data Bank (PDB) was used to retrieve information on the deposited structures of MnCatalase. These structures were categorised as mesophilic MnCatalases using PDB:1jku from *Lactobacillus plantarum* and 6j42 from *Anabaena*, and as thermophilic MnCatalases using PDB:2v8t and 2cwl from different strains of *Thermus thermophilus*. BLAST was undertaken to obtain homologous sequences by using PDB: 1jku and 2v8t as reference sequences and applying the selection criteria of > 60% sequence identity and > 80% coverage. The top 30 hits in our dataset were selected, and organism-wise. Ferritin datasets (C-ferritin and Dps) were taken from our previous study¹³. To further expand our analysis, MnCatalases from extremophiles such as the archaea

Pyrobaculum calidifontis and *Geobacillus thermopakistanensis* were modelled using AlphaFold 3.0³¹ and included in our dataset.

4.2 Monomer sequence characteristics– GRAVY, Secondary structure fraction, Instability Index

The physicochemical properties of MnCatalase and ferritin monomers were studied using the ProtParam tool¹⁴. Amino acid composition, aliphatic index, instability index and grand average of hydropathicity (GRAVY) were generated to compare these properties at a monomer scale for MnCatalases, Dps proteins and Classical ferritins. Monomer secondary-structure composition was analysed from PDB structures using the DSSP tool¹⁷.

4.3 Subunit-subunit Interface characteristics– interface area, compactness, and geometry

The PDB structures of MnCatalase and ferritins were downloaded from RCSB PDB. Due to the absence of crystal structure information of Archaea (*Pyrobaculum calidifontis*) MnCatalase and *Geobacillus thermopakistanensis* MnCatalase, their modelled 3-D structure was utilized. The best-predicted three-dimensional structures of the catalase proteins were generated using AlphaFold 3.0³¹ and visualized in PyMOL. All the structures were validated using the SAVES 6.1 server³², utilizing the ERRAT tool which is widely considered essential for optimal validation and quality check of the modelled proteins. Interface predicted template modelling score (ipTM) and predicted template modelling scores (pTM) were used to validate the quality of the modelled structure and pursue it for further analysis. To study the protein interfaces, each protein complex was split into individual subunits/chains. The interface area or Buried Surface Area (BSA) was calculated by estimating the sum of solvent Accessible Surface Area (ASA) of the two individual subunits in unbound form, minus that of the pair together in the complex/ bound form¹³. Those residues or atoms that lose ASA after complexation are considered interface residues/atoms. The NACCESS program³³ that uses the Lee and Richards rolling ball algorithm³⁴ was used for ASA calculation. The compactness or the atomic packing of the interface was estimated by calculating the local atomic density L_D ³⁵, which is the number of interface atoms within a radius of 12Å, averaged out for each of the interfaces of the same type. The interfaces were grouped by manual inspection (by visualizing the structure using PYMOL) and categorized into three types based on their position and geometry: a) lateral interface IntL, b) a second lateral interface IntLX, and c) an angular interface IntN. The distinction between IntL and IntLX is made based on the existence of an antiparallel β sheet contributed by both chains forming the interface. While antiparallel β sheets contributed by each chain are absent in IntL, they are part of IntLX and hence distinguished in the structure of the catalases. For C-ferritin and Dps, the interfaces are categorized as a) anti-parallel interfaces (IntA), b) perpendicular interface with N-terminus extension (IntB), and c) perpendicular interface with E-helix and C-terminus extension (IntC), as reported previously¹³.

4.4 H-bond interactions, electrostatics, and stability analysis of the interfaces

The HBPLUS tool³⁶ was used to calculate the number of H-bonds between protein subunits. An electrostatics and stability study of subunit interactions was performed using Rosetta¹⁸, an all-atom macromolecule energy function. Rosetta uses a scoring function REF15 that includes weighted sums of energy terms, like electrostatics and van der Waals' interactions, and statistical terms like the probability of finding the torsion angles in

Ramachandran space. Thus, for each interface in MnCatalases and ferritin complexes, electrostatic energy, van der Waals' energy, and free energy of binding/ energy (ΔG_{bind}) were obtained. The energy unit using REF 15 is described in "Rosetta Energy Unit" and is used to understand the energetics of interface interactions. All the energy terms used are estimated in kcal/mol. PDBePISA tool³⁷ was used to calculate ΔG_{bind} for the entire protein complexes.

4.5 MD simulations

We performed all-atom molecular dynamics (MD) simulations of the apo proteins using the GROMACS 2024 package with the Optimized Potentials for Liquid Simulations CHARMM27 all-atom force field. It required the 6-mer MnCatalases, 12-mer Dps structures and 24-mer C-ferritin to be prepared before forcefield application. A production run of 100 ns each was carried out for the protein systems in explicit solvent. For each titratable residue, the default protonation state was used. The complex was placed at the center of a cubic box, which was then solvated with TIP3P water molecules. The system's net charge was neutralized by adding Na⁺ and Cl⁻ ions, and energy minimization was performed using the steepest descent algorithm to remove any unfavorable contacts. The system was gradually heated to 300 K over 100 ps under constant number of particles, volume, and temperature (NVT) conditions, followed by 100 ps equilibration under constant number of particles, pressure and temperature (NPT) ensemble. Subsequently, a 100-ns final MD production run was performed and the simulations were repeated twice for each protein, and the simulation trajectory with minimum fluctuation in RMSD throughout, was selected for further analysis. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method, and periodic boundary conditions were applied in all directions. Trajectory analyses were performed using GROMACS analysis tools and in-house Python scripts.

4.6 Cage characteristics– cavity volume, cavity surface area, geometry, surface polarity

CavityPlus tool³⁸ was used for the generation of information related to the presence of cavities in MnCatalase, Dps, and C-ferritin cages. These cavity files were used for obtaining cavity volume, cavity surface area, and cavity surface information using in-house Python codes. These were utilized to generate cavity surface area by volume ratio and cavity polarity.

4.7 Protein catalytic site and its placement with respect to cavity entry point and subunit-subunit interfaces

Information on catalytic sites was obtained from the literature. The center of mass of the active site and cavity entry point was obtained using in-house Python code. The distance between the coordinates was calculated to determine the distance from the cavity entry point to the catalytic site. Also, an in-house Python program was written to obtain the shortest distance between the interface coordinates and the center of mass of the catalytic site.

5. Conclusion

This study provides structural insights into the functional forms of MnCatalases across diverse organisms and situates these observations within the broader context of the ferritin superfamily by comparing them with Dps and C-ferritins. Comparative analyses at the monomeric, interfacial, and cavity levels enabled clear discrimination between thermophilic and extremophilic MnCatalases and their mesophilic counterparts.

Thermophilic and extremophilic MnCatalases exhibited a higher proportion of non-polar residues together with an increased content of ordered secondary structural elements like α -helices and β -sheets, suggesting enhanced structural rigidity and stability under extreme environmental conditions. At the oligomeric interface, thermophilic and extremophilic MnCatalases displayed stronger inter-subunit interactions, reflected by increased van der Waals attractions, electrostatic interactions, and more favourable ΔG_{int} (Rosetta Energy Units), compared with their mesophilic counterparts. These structural characteristics provide a molecular basis for the enhanced stability and functional adaptation of thermophilic and extremophilic MnCatalases relative to the mesophilic ones and in future could pave the way to engineering synthetic thermophilic proteins. Cavity analyses further revealed distinct adaptations associated with extremophilic enzymes. In particular, extremophilic MnCatalases possessed cavities with lower surface area-to-volume ratios, while thermophilic and archaeal MnCatalases exhibited a greater proportion of non-polar cavity-lining residues than mesophilic and GeoB MnCatalases. Interestingly, GeoB MnCatalase not only occupied a distinct position within the ferritin superfamily phylogenetic tree but also displayed a few structural features that differed from archaeal MnCatalases. These observations suggest that GeoB MnCatalase may represent a unique evolutionary lineage with distinct structural adaptations. Future studies investigating additional physicochemical properties of GeoB MnCatalase will help elucidate the basis of its structural divergence from other thermophilic and archaeal proteins. Finally, extending our comparative analyses to Dps (12-mer) and C-ferritin (24-mer) provided valuable insights into the chain economy of the hexameric MnCatalase. Despite comprising only six subunits, MnCatalase forms a stable protein cage while positioning its active sites in close proximity to the cavity entry points, thereby facilitating rapid substrate access and efficient catalysis. These findings advance our understanding of the structural principles underlying the stability, functional specialization, and evolutionary diversification of MnCatalases within the ferritin superfamily.

Supplementary data to this article can be found online at XXX.

Study Limitations

The structural analyses are based on the limited non-redundant wild-type MnCatalase structures available in the PDB database; hence, we have modelled structures using AlphaFold in some cases and we acknowledge the limitation associated with a small dataset. We also acknowledge that MD simulations were done for 100ns time and can be extended further if needed.

Authorship contribution statement

Amrita: Writing – original draft, Formal analysis, Data curation, Methodology, Investigation. **Srijita Acharya:** Methodology, Investigation. **Soumyananda Chakraborti:** Writing – review & editing, Conceptualization. **Sucharita Dey:** Writing – original draft, Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization, Supervision.

Declarations

Competing Interests

S.D. and S.C. are members of the Editorial Board of Scientific Reports. S.D. and S.C. had no involvement in the editorial handling or peer review of this manuscript. The authors declare no other competing interests.

Funding

This work was supported by the research grant from the Department of Biotechnology, Government of India (RLS grant to SD: BT/ RLF/Re-entry/10/2020, sanction order serial number 145), and ARG-MATRICES from Anusandhan National Research Foundation (ANRF) (ANRF/ARGM/2025/000752/TS), which are gratefully acknowledged.

Author Contribution

Amrita wrote the original draft and performed all the analysis, data curation, methodology, and investigation. S.A. worked on methodology and investigation. S.C. did review & editing, and conceptualisation. S.D. conceptualised the work, designed the experiments, provided resources, was responsible for funding acquisition, supervision and did review & editing of the manuscript. All authors reviewed the manuscript.

Acknowledgement

A acknowledges the CSIR-NET India fellowship. SA acknowledges the MHRD fellowship. SD acknowledges DBT India for the RLS grant, Anusandhan National Research Foundation (ANRF) and IIT Jodhpur, India, for infrastructure support. SC acknowledges DBT India for the RLS grant.

Data Availability

This study did not generate new experimental structural data. All crystal structures analysed were obtained from the publicly available RCSB PDB repository [<https://www.rcsb.org/>], and the PDB codes are given in the main text and in Supplementary Table 1. The datasets generated during the current study, namely the computationally modelled structures of Mn catalases from *Pyrobaculum calidifontis* (Arch) and *Geobacillus thermopakistanensis* (GeoB), are available in the Mendeley Data repository, <https://doi.org/10.17632/8vck63dgfd.2>. Additionally, all the raw data files and analysis code are available at <https://github.com/amrita-srinivasa/MnCatalase-Python-codes-Figures-and-Files.git>.

References

1. Chakraborti, S. & Chakrabarti, P. Self-Assembly of Ferritin: Structure, Biological Function and Potential Applications in Nanotechnology. *Adv. Exp. Med. Biol.* **1174**, 313–329. https://doi.org/10.1007/978-981-13-9791-2_10 (2019).
2. Kaushal, J., Mehandia, S., Singh, G., Raina, A. & Arya, S. K. Catalase enzyme: Application in bioremediation and food industry. *Biocatal. Agric. Biotechnol.* **16**, 192–199. <https://doi.org/10.1016/j.bcab.2018.07.035> (2018).
3. Bihani, S. C., Chakravarty, D. & Ballal, A. KatB, a cyanobacterial Mn-catalase with unique active site configuration: Implications for enzyme function. *Free Radic Biol. Med.* **93**, 118–129.

- <https://doi.org/10.1016/j.freeradbiomed.2016.01.022> (2016).
4. Antonyuk, S. V. et al. Three-dimensional structure of the enzyme dimanganese catalase from *Thermus Thermophilus* at 1 Å resolution. *Crystallogr. Rep.* **45**, 105–116. <https://doi.org/10.1134/1.171145> (2000).
 5. Barynin, V. V. et al. Crystal structure of manganese catalase from *Lactobacillus plantarum*. *Structure* **9**, 725–738. [https://doi.org/10.1016/s0969-2126\(01\)00628-1](https://doi.org/10.1016/s0969-2126(01)00628-1) (2001).
 6. Geiger, O. Biogenesis of Fatty Acids, Lipids and Membranes.
 7. Pedone, E., Fiorentino, G., Bartolucci, S. & Limauro, D. Enzymatic Antioxidant Signatures in Hyperthermophilic Archaea. *Antioxid. (Basel)*. 9. <https://doi.org/10.3390/antiox9080703> (2020).
 8. Shaeer, A., Aslam, M. & Rashid, N. A highly stable manganese catalase from *Geobacillus thermopakistaniensis*: molecular cloning and characterization. *Extremophiles* **23**, 707–718. <https://doi.org/10.1007/s00792-019-01124-5> (2019).
 9. Kono, Y. & Fridovich, I. Functional significance of manganese catalase in *Lactobacillus plantarum*. *J. Bacteriol.* **155**, 742–746. <https://doi.org/10.1128/jb.155.2.742-746.1983> (1983).
 10. Andrews, S. C. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim. Biophys. Acta.* **1800**, 691–705. <https://doi.org/10.1016/j.bbagen.2010.05.010> (2010).
 11. Whittaker, J. W. Non-heme manganese catalase—the other catalase. *Arch. Biochem. Biophys.* **525**, 111–120. <https://doi.org/10.1016/j.abb.2011.12.008> (2012).
 12. Aumiller, W. M., Uchida, M. & Douglas, T. Protein cage assembly across multiple length scales. *Chem. Soc. Rev.* **47**, 3433–3469. <https://doi.org/10.1039/c7cs00818j> (2018).
 13. Physicochemical features of subunit interfaces and their role in self-assembly across the ferritin superfamily Structure. (2024). <https://doi.org/10.1016/j.str.2024.12.004>
 14. Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552. <https://doi.org/10.1385/1-59259-584-7:531> (1999).
 15. Williams, R. M. et al. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.* 89–100. https://doi.org/10.1142/9789814447362_0010 (2001).
 16. Klotz, M. G. & Loewen, P. C. The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. *Mol. Biol. Evol.* **20**, 1098–1112. <https://doi.org/10.1093/molbev/msg129> (2003).
 17. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637. <https://doi.org/10.1002/bip.360221211> (1983).
 18. Alford, R. F. et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048. <https://doi.org/10.1021/acs.jctc.7b00125> (2017).
 19. Montemiglio, L. C. et al. Cryo-EM structure of the human ferritin-transferrin receptor 1 complex. *Nat. Commun.* **10**, 1121. <https://doi.org/10.1038/s41467-019-09098-w> (2019).
 20. Cho, K. J. et al. The crystal structure of ferritin from *Helicobacter pylori* reveals unusual conformational changes for iron uptake. *J. Mol. Biol.* **390**, 83–98. <https://doi.org/10.1016/j.jmb.2009.04.078> (2009).
 21. Yariv, B. et al. Using evolutionary data to make sense of macromolecules with a face-lifted ConSurf. *Protein Sci.* **32**, e4582. <https://doi.org/10.1002/pro.4582> (2023).
 22. Smejkal, G. B. & Kakumanu, S. Enzymes and their turnover numbers. *Expert Rev. Proteom.* **16**, 543–544. <https://doi.org/10.1080/14789450.2019.1630275> (2019).

23. Marcelino, A. M. C. & Gierasch, L. M. Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **89**, 380–391. <https://doi.org/10.1002/bip.20960> (2008).
24. Fasman, G. D. (ed) Prediction of protein structure and the principles of protein conformation (Springer) (2012). <https://doi.org/10.1007/978-1-4613-1571-1>
25. Fetrow, J. S. Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.* **9**, 708–717 (1995).
26. Panja, A. S., Bandopadhyay, B. & Maiti, S. Protein Thermostability Is Owing to Their Preferences to Non-Polar Smaller Volume Amino Acids, Variations in Residual Physico-Chemical Properties and More Salt-Bridges. *PLoS One.* **10**, e0131495. <https://doi.org/10.1371/journal.pone.0131495> (2015).
27. Baldwin, R. L. Energetics of protein folding. *J. Mol. Biol.* **371**, 283–301. <https://doi.org/10.1016/j.jmb.2007.05.078> (2007).
28. Folch, B., Rooman, M. & Dehouck, Y. Thermostability of salt bridges versus hydrophobic interactions in proteins probed by statistical potentials. *J. Chem. Inf. Model.* **48**, 119–127. <https://doi.org/10.1021/ci700237g> (2008).
29. Burnim, A. A., Xu, D., Spence, M. A., Jackson, C. J. & Ando, N. Analysis of insertions and extensions in the functional evolution of the ribonucleotide reductase family. *Protein Sci.* **31**, e4483. <https://doi.org/10.1002/pro.4483> (2022).
30. Chelikani, P., Carpena, X., Fita, I. & Loewen, P. C. An electrical potential in the access channel of catalases enhances catalysis. *J. Biol. Chem.* **278**, 31290–31296. <https://doi.org/10.1074/jbc.M304076200> (2003).
31. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2> (2021).
32. Waterhouse, A. M. et al. The structure assessment web server: for proteins, complexes and more. *Nucleic Acids Res.* **52**, W318–W323. <https://doi.org/10.1093/nar/gkae270> (2024).
33. Hubbard, S. J. *naccess* (Dept Biohem Mol Biol UCL, 1993).
34. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400. [https://doi.org/10.1016/0022-2836\(71\)90324-x](https://doi.org/10.1016/0022-2836(71)90324-x) (1971).
35. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.* **336**, 943–955. <https://doi.org/10.1016/j.jmb.2003.12.073> (2004).
36. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793. <https://doi.org/10.1006/jmbi.1994.1334> (1994).
37. Wlodawer, A., Dauter, Z. & Jaskolski, M. *Protein Crystallography: Methods and Protocols* (Humana), (2017).
38. Xu, Y. et al. CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.* **46**, W374–W379. <https://doi.org/10.1093/nar/gky380> (2018).

Figures

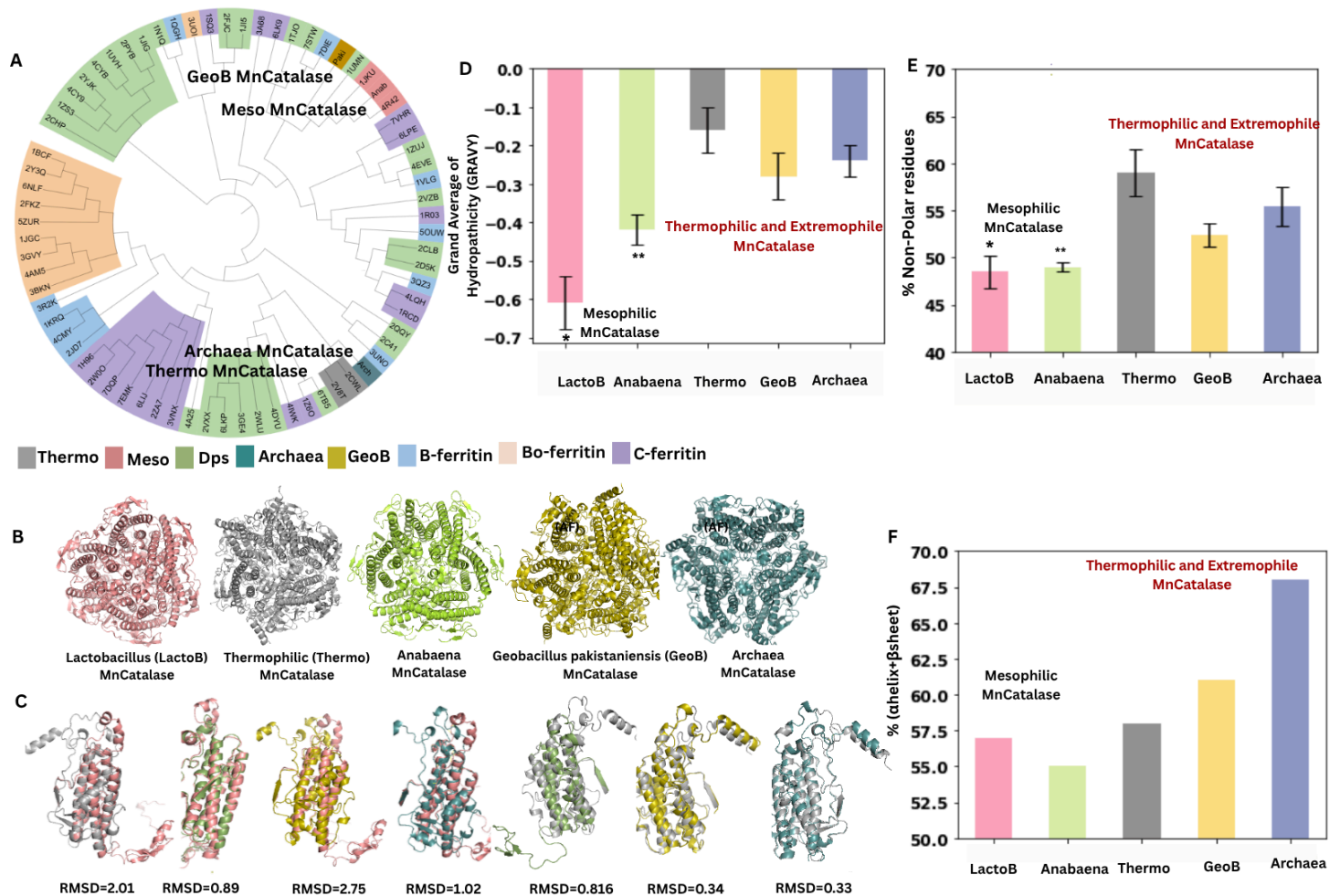


Figure 1

MnCatalase monomer and its structural attributes

A. Phylogenetic Tree of ferritin superfamily in circular layout that includes C-ferritin (purple), B-ferritin (blue), Bo-ferritin (orange), Dps proteins (green), mesophilic (LactoB and Anabaena) MnCatalase (pink), Archaea MnCatalase (navy blue), *Geobacillus thermopakistanensis* MnCatalase (mustard) and thermophilic MnCatalase (grey). **B** Cage structure representation of. Mesophilic MnCatalase (PDB:1jku) in pink, Thermophilic MnCatalase (PDB: 2v8t) in grey, Anabaena MnCatalase (PDB:6j42) in greenish yellow, *Geobacillus thermopakistanensis* MnCatalase in yellow mustard and Archaea MnCatalase in blue. **C.** Structural alignment of a single subunit between MnCatalases of *Lactobacillus* and *Thermus thermophilus*, between *Lactobacillus* and Anabaena, between *Lactobacillus* and *Geobacillus*, and between *Lactobacillus* and Archaea. Structural alignment between MnCatalases of *Thermus thermophilus* and Anabaena, between *Thermus thermophilus* and *Geobacillus*, and between *Thermus thermophilus* and Archaea, with their Root Mean Square Deviation. **D.** Barplot depicting Grand Average of Hydropathicity (GRAVY), **E.** % Non polar residues and **F.** % (α-helix and β-sheet) for MnCatalase proteins.

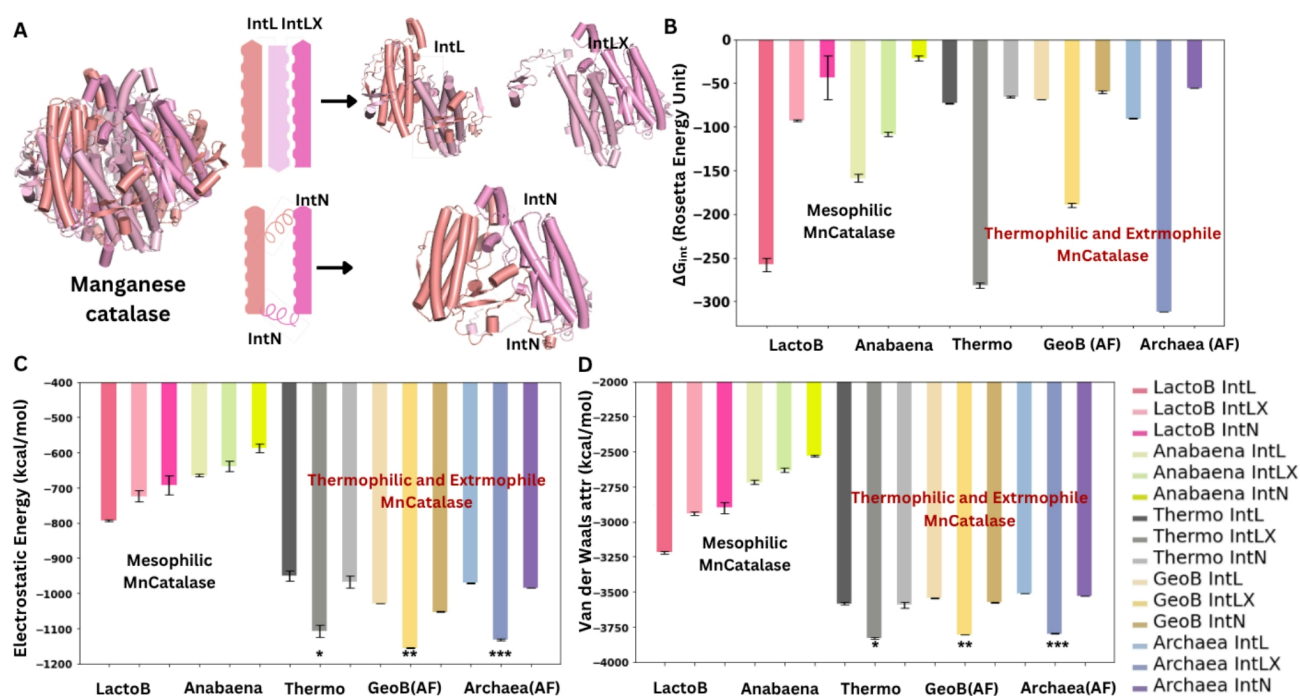


Figure 2

Lactobacillus (LactoB) MnCatalase interfaces and comparative study of its physicochemical attributes

A. Cylindrical cartoon representation of LactoB MnCatalase (PDB:1jku) showing three of the six chains comprising lateral interfaces IntL and IntLX and angular interface IntN. The distinction between IntL and IntLX is based on the presence of an antiparallel β -sheet contributed by each of the chains forming the interface (IntLX) and absence in IntL. Barplot distribution of **B**. Delta G of interface (ΔG_{int}) (Rosetta Energy Unit), **C**. Interface Electrostatic Energy in kcal/mol and **D**. Interface Van der Waals attraction in kcal/mol for the three interfaces (IntL, IntLX and IntN) of Mesophilic (Lactobacillus and Anabaena), Thermophilic (*Thermus thermophilus*), and extremophile (*Geobacillus thermopakistaniensis* and Archaea). In all the barplots, Lactobacillus MnCatalase interfaces data is shown in the shades of pink, Anabaena data is shown in the shades of greenish yellow, Thermophilic MnCatalase interfaces data is shown in the shades of grey, Geobacillus is shown in the shades of mustard yellow and Archaea MnCatalase interfaces data is shown in the shades of blue. AF denotes the AlphaFold structures of *Geobacillus thermopakistaniensis* (GeoB) and Archaea (*Pyrobaculum calidifontis*)

*, ** and *** symbols denote the significance level ($p < 0.05$) calculated using Welch's 2 sample t-test between Thermophilic and mesophilic MnCatalases, GeoB and mesophilic MnCatalases and Archaea and mesophilic MnCatalases.

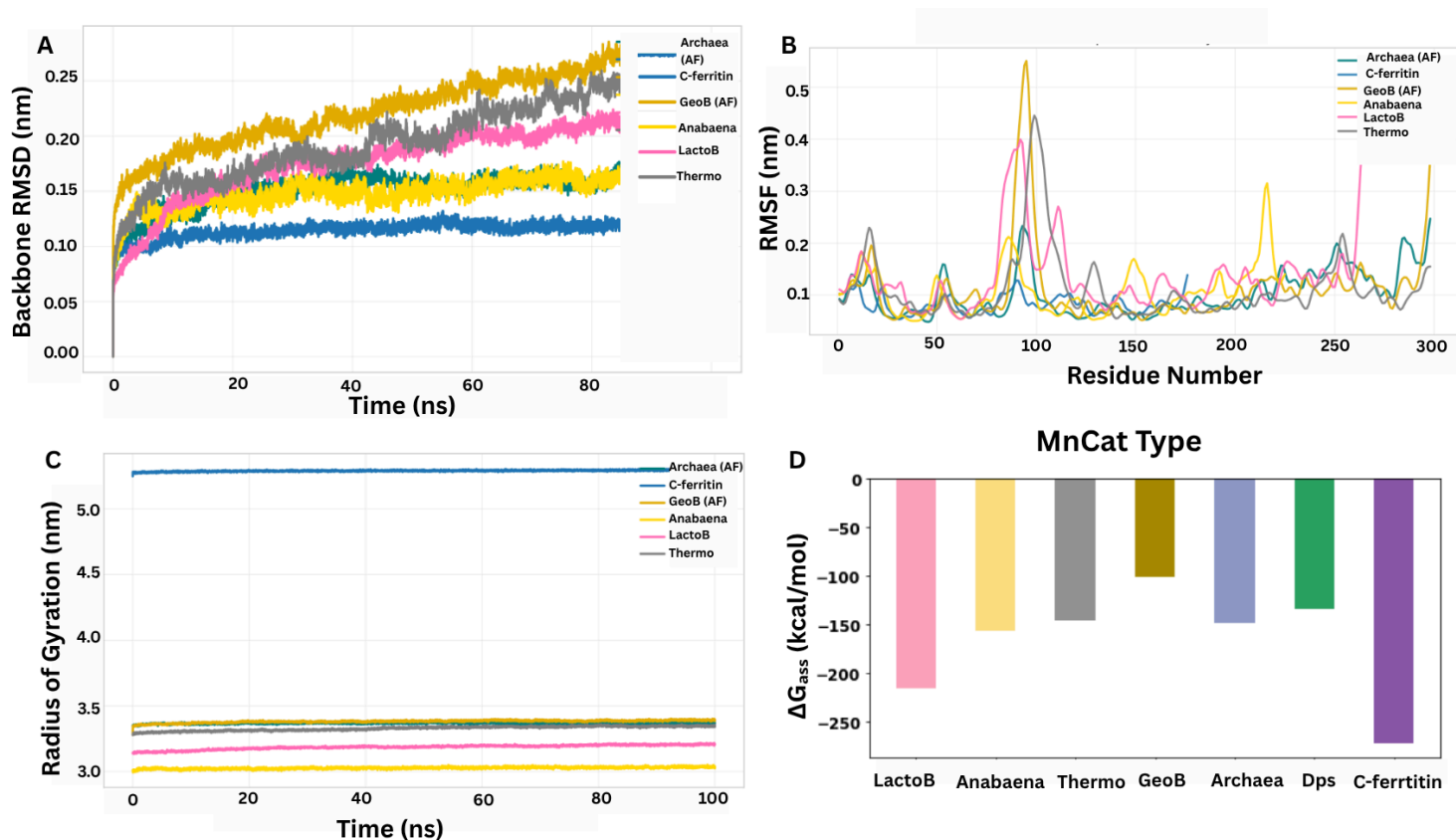


Figure 3

Dynamic analysis of MnCatalase structure and comparing it with static data

Line graph of **A**. Backbone Root Mean Square Deviation (RMSD) in nm for 200ns, **B**. Root Mean Square Fluctuation (RMSF) for the residues and **C**. Radius of Gyration (nm) for MnCatalases and C-ferritin assembly. LactoB graph is denoted in pink, Thermophilic MnCatalase in grey, Anabaena in yellow, *Geobacillus thermopakistanensis* in Mustard, Archaea in teal and C-ferritin in dark blue. (D) Barplot depicting ΔG_{ass} obtained using static calculation, the colour code as pink for *Lactobacillus* MnCatalase, lime green for Anabaena, grey for thermophilic MnCatalase, mustard yellow for *Geobacillus* MnCatalase and Blue for Archaea MnCatalase.

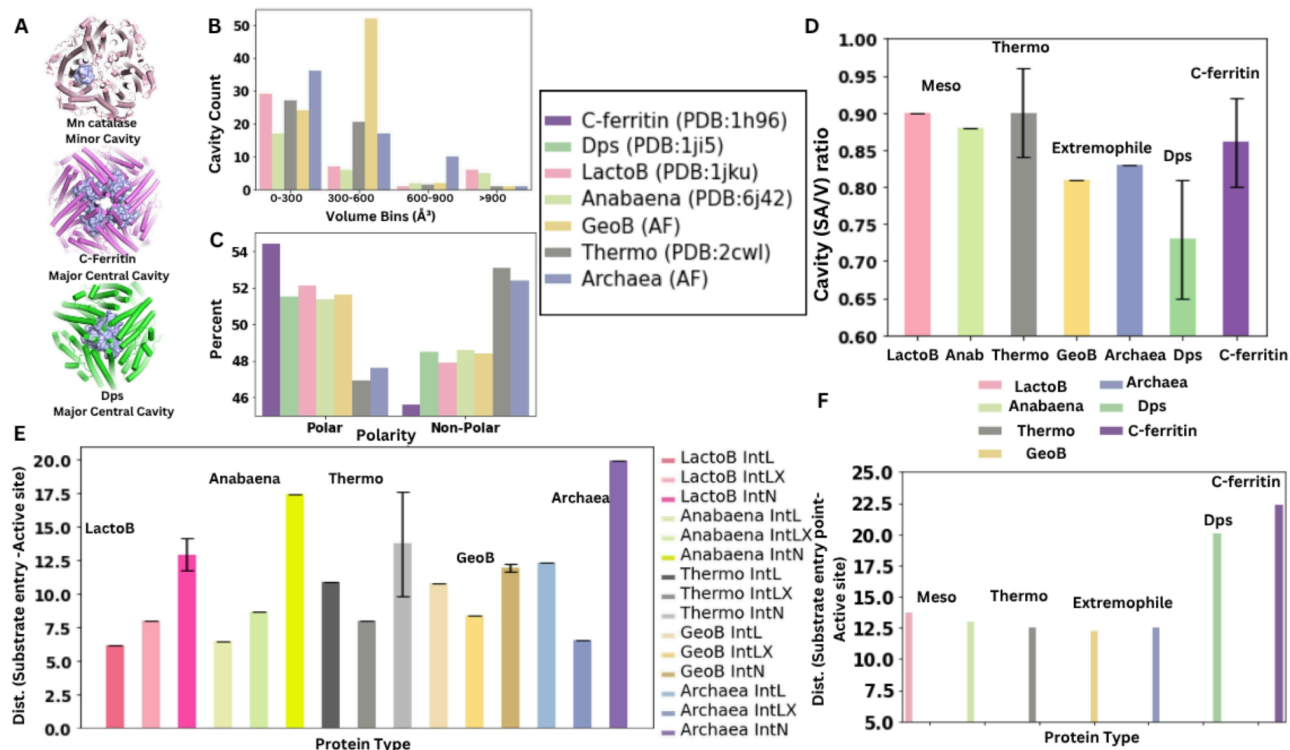


Figure 4

MnCatalase cavity with respect to Dps proteins and C-ferritin

A. Cylindrical cartoon figure showing C-ferritin's (PDB:1r03) major central cavity in purple, Dps's (PDB:1ji5) major central cavity in green and one of the minor cavities of Lactobacillus MnCatalase (PDB:1jku) in pink, also showing the absence of a central cavity. Barplot depicting **B**. Cavity counts with respect to volume bins. **C**. Percent polar and non-polar residues in the cavity lining using the colour code as pink for Lactobacillus MnCatalase, lime green for Anabaena, grey for thermophilic MnCatalase, mustard yellow for Geobacillus MnCatalase, and blue for Archaea MnCatalase. Green for Dps and purple for C-ferritin. **D**. Barplot depicting Cavity Surface Area/ Volume ratio for MnCatalases with colour code as in B and Dps in green and C-ferritin in purple. **E**. Barplot depicting the distance between COM of catalytic site and COM of interfaces (in Å) with colour code as in B. **F**. Distance between metal entry point and catalytic site (in Å) for proteins color coded as in C.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [02Supplementary.docx](#)