Title

Constructing protein polyhedra via orthogonal chemical interactions

Journal

Nature, 578(7793)

ISSN

0028-0836

Authors

Golub, Eyal
Subramanian, Rohit H
Esselborn, Julian
et al.

Publication Date

2020-02-06

DOI

10.1038/s41586-019-1928-2

Peer reviewed

Constructing protein polyhedra via orthogonal chemical interactions

https://doi.org/10.1038/s41586-019-1928-2

Received: 9 May 2019

Accepted: 19 November 2019

Published online: 22 January 2020

Eyal Golub[1], Rohit H. Subramanian[1], Julian Esselborn[1], Robert G. Alberstein[1], Jake B. Bailey[1], Jerika A. Chiong[1], Xiaodong Yan[2], Timothy Booth[2], Timothy S. Baker[2] & F. Akif Tezcan[1,3]*

Many proteins exist naturally as symmetrical homooligomers or homopolymers[1]. The emergent structural and functional properties of such protein assemblies have inspired extensive efforts in biomolecular design[2–5]. As synthesized by ribosomes, proteins are inherently asymmetric. Thus, they must acquire multiple surface patches that selectively associate to generate the different symmetry elements needed to form higher-order architectures[1,6]—a daunting task for protein design. Here we address this problem using an inorganic chemical approach, whereby multiple modes of protein–protein interactions and symmetry are simultaneously achieved by selective, 'one-pot' coordination of soft and hard metal ions. We show that a monomeric protein (protomer) appropriately modified with biologically inspired hydroxamate groups and zinc-binding motifs assembles through concurrent Fe^{3+} and Zn^{2+} coordination into discrete dodecameric and hexameric cages. Our cages closely resemble natural polyhedral protein architectures[7,8] and are, to our knowledge, unique among designed systems[9–13] in that they possess tightly packed shells devoid of large apertures. At the same time, they can assemble and disassemble in response to diverse stimuli, owing to their heterobimetallic construction on minimal interprotein-bonding footprints. With stoichiometries ranging from [2 Fe:9 Zn:6 protomers] to [8 Fe:21 Zn:12 protomers], these protein cages represent some of the compositionally most complex protein assemblies—or inorganic coordination complexes—obtained by design.

Cage-like architectures have featured prominently in supramolecular design, owing to their aesthetically appealing structures and isolated interiors, which enable them to encapsulate molecular cargo and to perform selective chemical transformations[14–18]. Inspired by naturally occurring polyhedral assemblies, protein engineers have combined principles of symmetry with the proper design and arrangement of non-covalent interfaces to build diverse supramolecular architectures[9–13]. However, some of the key structural features of natural protein cages have been difficult to emulate (Fig. 1a). First, each cage is invariably composed of asymmetric protomers, which possess multiple self-associative patches to simultaneously satisfy the symmetry requirements necessary to build polyhedral assemblies (that is, concurrent generation of at least C_2 and C_3 symmetries, in addition to C_4 or C_5 symmetries for octahedra or icosahedra)[1,6]. Second, these self-associative patches collectively occupy a large fraction of the surface area on each protomer, enabling the formation of tightly packed shells with small apertures to enable the influx and efflux of select species[8]. Third, although the inter-protomer interfaces in natural protein cages are extensive, to ensure stable and selective association, they are also often conformationally flexible and chemically tunable, allowing the cages to undergo cooperative motions or disassembly in response to external cues[7,19].

Given the difficulty of designing multiple, selectively associative surfaces on a protomer, construction of artificial cages has relied exclusively on using natively oligomeric proteins or designed peptides with $C_{n\geq 2}$ symmetries as building blocks and the design of a single type of binary protein–protein interaction (PPI) through computation[10], genetic fusion[9,11], disulfide bond formation[12] or metal coordination[11,13]. Although these strategies can yield polyhedral symmetries, the resulting architectures are highly porous, do not display externally controllable assembly or disassembly (with two exceptions)[11,13] and cannot be easily modified to adopt alternative structures (that is, they are not modular or flexible). Inspired by previous work on bimetallic supramolecular coordination cages[20,21], we investigated whether these design problems could be addressed using an inorganic chemical approach, wherein a protomer is equipped with chemically orthogonal coordination motifs to self-assemble into polyhedral architectures.

Design of bimetallic protein cages

Previously, we have taken advantage of the simultaneous strength, lability and directionality of metal coordination bonds (particularly those formed by late first-row, low-valent transition metal ions) to effect the self-assembly of discrete protein complexes[5] and extended one-, two- and three-dimensional arrays[22,23]. Typically, selective nucleation sites for metal-mediated PPIs are formed by pairs of metal-binding amino acids (mainly His, Asp and Glu residues) (Fig. 1b) or non-native bidentate

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. [2]Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. [3]Materials Science and Engineering, University of California, San Diego, La Jolla, CA, USA. *e-mail: tezcan@ucsd.edu

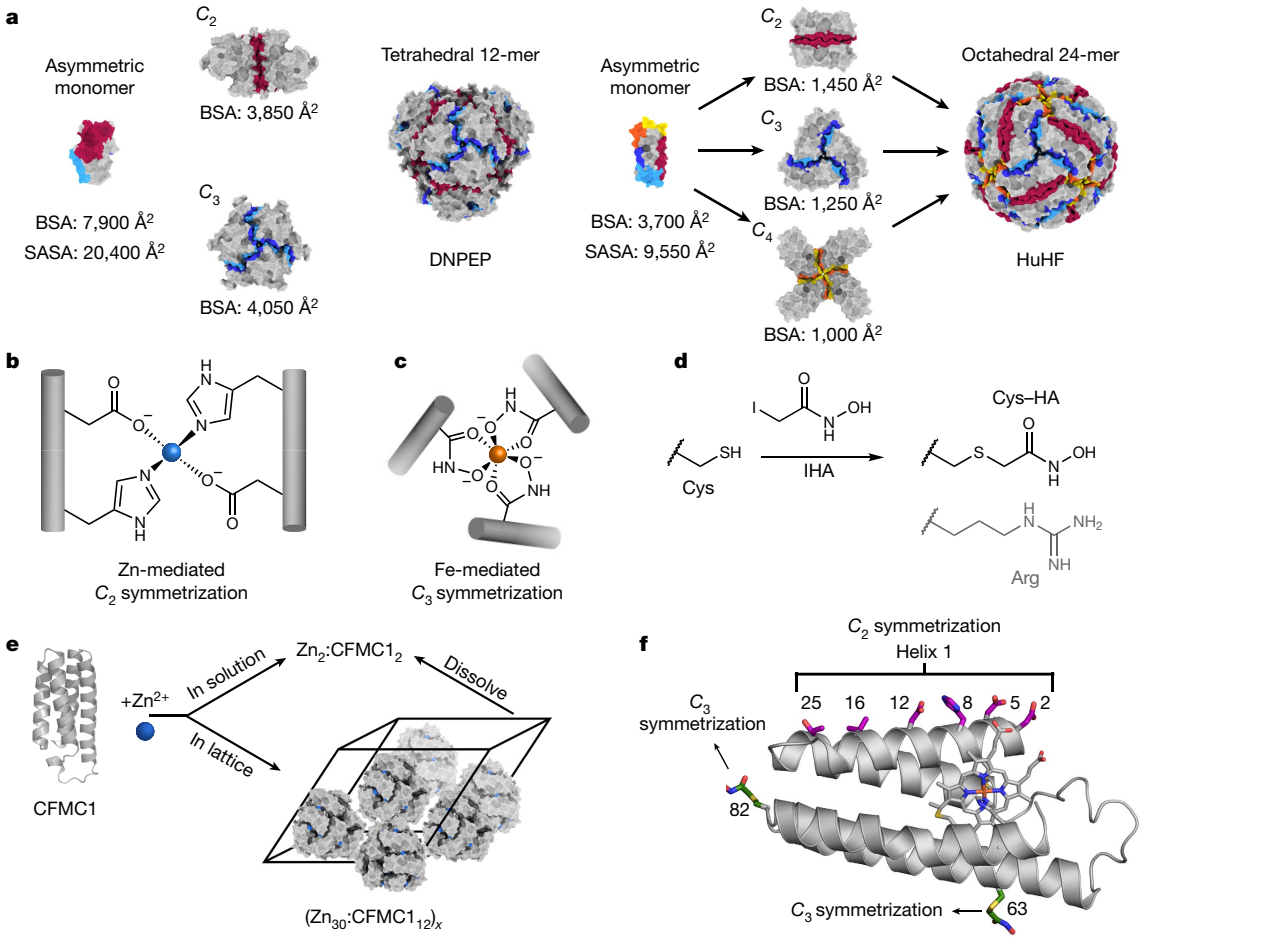


Fig. 1 | Design of protein cages. a, Representative examples of natural protein cages (DNPEP, aspartyl aminopeptidase; HuHF, human heavy-chain ferritin) and their assembly from asymmetric protomers. Per-protomer solvent-accessible surface areas (SASA) and buried surface areas (BSA) are indicated. Associative surfaces on the protomers are coloured red for homologous interactions and orange/yellow or blue/cyan for heterologous interactions. **b**–**f**, Design of artificial protein cages by metal coordination. **b**, C_2-symmetric protein dimerization induced by tetrahedral Zn^{2+} coordination of native amino acid side chains. **c**, C_3-symmetric protein trimerization induced by octahedral Fe^{3+}-tris-hydroxamate coordination. **d**, Scheme showing modification of native Cys side chains with IHA to yield Cys–HA, which is isosteric with arginine (light grey). **e**, Zn-mediated solution dimerization and crystallization of CFMC1. **f**, Structural overview of the cytochrome cb_{562} scaffold. Salient structural elements are shown as sticks.

functionalities (for example, 2,2′-bipyridine, 1,10-phenanthroline and 8-hydroxyquinoline)[5,24]. However, all of these natural or synthetic coordination motifs can be considered as soft (or intermediate-soft) according to the Hard-Soft Acid-Base (HSAB) classification[25] and have considerable overlap in terms of their coordination preferences for soft, low-valent transition metal ions. Owing to this lack of chemical discrimination, it has not been possible to design a heterometallic protein complex for which the self-assembly is selectively guided by multiple metal ions that mediate different PPIs.

To achieve this goal, we turned to a bidentate chelating motif, hydroxamate (HA, the conjugate base of hydroxamic acid), a common functional group found in bacterial siderophores to enable exceptionally stable coordination of Fe^{3+} ions[26,27]. HA groups preferentially form octahedral Fe^{3+} complexes with an inherent C_3 symmetry that we sought to impose on protein oligomerization (Fig. 1c). Notably, the formation constants of Fe^{3+}:$(HA)_3$ complexes (>10^{28} M^{-3}) are vastly higher than those of other metal–HA complexes, such that they can be considered as orthogonal to the aforementioned soft metal–ligand combinations[26,27]. For protein derivatization, we synthesized a small reagent, iodo-hydroxamic acid (IHA), which selectively reacts with Cys residues (Fig. 1d, Extended Data Fig. 1). The resulting Cys–HA side chain is isosteric with that of arginine and devoid of bulky aromatic moieties, furnishing a pseudo-natural amino acid functionality with the ability to chelate hard metal ions and induce C_3 symmetric oligomerization on a single-residue footprint.

As a model system, we used cytochrome cb_{562}, a monomeric four-helix-bundle protein that has proved to be a versatile building block for metal-directed protein self-assembly[5]. A variant of cyt cb_{562} (CFMC1), which was designed and observed to form Zn-mediated dimers in solution, crystallizes into rhombohedral lattices in which the protomers arrange into dodecameric, cage-like units via Zn-mediated crystal packing interactions[28] (Fig. 1e). Whereas Zn-mediated interactions were not sufficiently strong to maintain the tetrahedral dodecamers upon crystal dissolution, we envisioned that these lattice units could serve as a structural model to engineer the protomers such that they would form self-standing cages. Looking first to stabilize the C_2 symmetric interfaces, we incorporated a bidentate His8–Asp12 motif to mediate the antiparallel association of two protomers along their helices 1 via tetrahedral Zn^{2+} coordination (Fig. 1f). Given that C_3 symmetric interfaces are small and heterologous (that is, they involve two different patches on each protomer; Extended Data Fig. 2a), they were unsuitable for stabilization by noncovalent interactions. Therefore, we focused on the central pores in each C_3 symmetric substructure and identified positions 63 and 82 as suitable locations for installing Cys–HA functionalities, which would stabilize trimeric substructures by forming Fe^{3+}:$(HA)_3$ centres (Fig. 1f). Thus, we prepared two CFMC1 variants

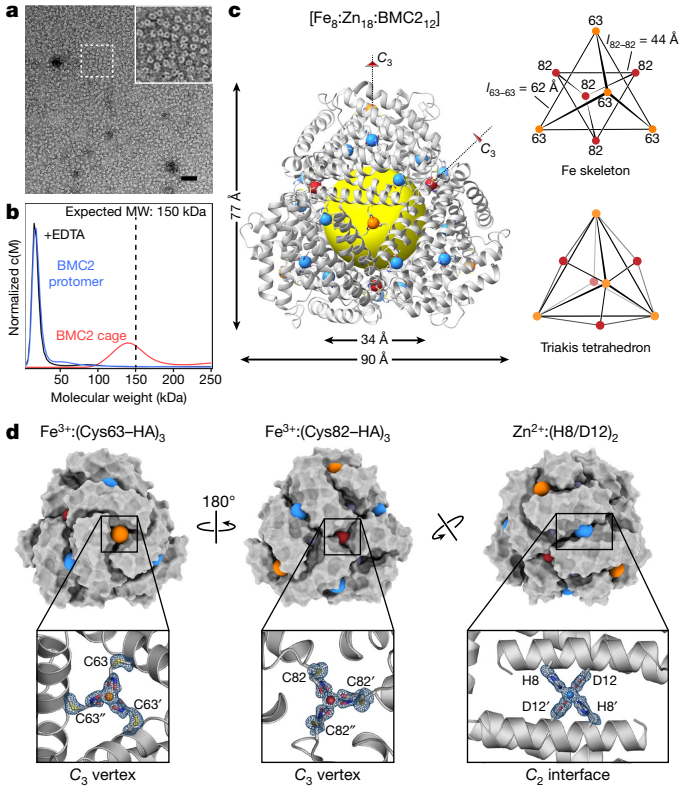


Fig. 2 | Characterization of BMC2 cages. a, ns-TEM of BMC2 cages obtained by the dissolution of three-dimensional crystals. Inset, close-up of the boxed region. Scale bar, 50 nm. **b**, AUC characterization of BMC2 protomers and BMC2 cages after crystal dissolution and after subsequent treatment with EDTA. **c**, Crystal structure of the BMC2 cage. Fe and Zn ions are represented as orange/red and blue spheres, respectively. The central cavity is highlighted by a yellow sphere. Two types of C_3 vertices formed by Fe:(Cys63–HA)$_3$ and Fe:(Cys82–HA)$_3$ coordination motifs form two superimposed tetrahedra to generate a triakis tetrahedron. **d**, Surface representations of the BMC2 cage, with metal ions shown as coloured spheres. Insets show atomic details of each metal coordination site, with the $mF_o - DF_c$ electron density omit map (blue mesh) contoured at 3σ.

designated bimetallic cage 1 and bimetallic cage 2 (BMC1 and BMC2; Extended Data Fig. 2b). Both BMC1 and BMC2 bear the His8–Asp12 motif on helix 1 and Cys63–HA along with the native peripheral Zn coordination sites (Ala1$_{N\text{-term}}$, Asp39 and His77) of the parent CFMC1 structure. BMC2 additionally contains Cys82–HA (Extended Data Figs. 1, 2b).

Crystals of BMC1 and BMC2 were obtained in the presence of near equimolar $ZnCl_2$ and $FeSO_4$. These crystals were isomorphic (*R*32 space group; $a = b = 126 \pm 1$ Å, $c = 167 \pm 1$ Å) with those of CFMC1[28], indicating that they possessed the same underlying lattice structure composed of dodecameric units (Extended Data Table 1, Extended Data Fig. 2). Crystals were dissolved in a solution lacking the precipitating agent (PEG-400) and then analysed by negative-stain transmission electron microscopy (ns-TEM; Fig. 2a, Extended Data Fig. 3). The images revealed uniform particles with a diameter of 8.4 ± 0.8 nm in the case of BMC2 but not BMC1, implying that two HA coordination motifs are necessary for cage stability. Analysis of the same BMC2 solution by analytical ultracentrifugation (AUC) indicated a predominant species with a molecular weight (MW_{obs}) of about 140 kDa (Fig. 2b), approximating the calculated value (MW_{calc}) of 150 kDa for a dodecamer. BMC2 particles dissociated upon treatment with ethylenediamine tetraacetic acid (EDTA), confirming their metal-dependent self-assembly (Fig. 2b, Extended Data Fig. 3).

We determined the crystal structure of the BMC2 cage at 1.4 Å resolution (Extended Data Table 1), revealing a compact structure with the shape of a truncated tetrahedron, outer dimensions of 80 × 90 Å and a cavity volume of 32,700 Å^3 (Fig. 2c, Extended Data Fig. 4). Like natural protein cages, the shell is tightly packed and the largest opening measures less than 4 Å across. Nearly 30% of the surface area of each protomer (1,700 Å^2 out of 6,500 Å^2) is buried in interfaces despite a design footprint of only four amino acids (His8, Asp12, Cys63 and Cys82).

The full complement of metal ions, comprising eight Fe ions (four each in the C_3 symmetric pores) and eighteen Zn ions (six in C_2 interfaces and twelve in peripheral sites) are clearly resolved (Fig. 2c). Anomalous X-ray diffraction data collected at and below Fe and Zn K-edges indicate that the designed Fe- and Zn-coordination sites exclusively bind to their cognate ions with no evidence of crosstalk (Extended Data Fig. 5, Supplementary Tables 3–6), which establishes that the metal-dependent self-assembly of BMC2 cages occurs with absolute chemical selectivity. The Fe centres form the eight C_3 vertices of a triakis tetrahedron, a Catalan solid with twelve equivalent faces (Fig. 2c). It can be viewed as the superposition of two tetrahedra: four Fe centres that are coordinated by Cys63–HA motifs generate the larger of these two tetrahedra (with an edge length (l_{edge}) of 62 Å), and four Fe centres coordinated by Cys82–HA motifs produce the smaller one (l_{edge} = 44 Å) (Fig. 2c). The BMC1 structure, in comparison, has a regular tetrahedral arrangement of four Fe centres as it lacks the Cys82–HA group (Extended Data Table 1, Extended Data Fig. 2).

As designed, the edges of the BMC2 tetrahedra are formed by six Zn ions located centrally in C_2 interfaces (Fig. 2d). The Fe^{3+}:(Cys63–HA)$_3$ and Fe^{3+}:(Cys82–HA)$_3$ motifs display near-ideal octahedral geometries (Fig. 2d), with the former in Λ (left handed) and the latter in Δ (right handed) configuration (Extended Data Fig. 5). Notably, the Fe^{3+}:(Cys82–HA)$_3$ centre also adopts an alternative conformation (20% abundance) owing to the flexibility of the Cys–HA side chain (Extended Data Fig. 5k). All Fe–O bond distances are in the range of 1.95–2.1 Å, which are typical of Fe^{3+}:(HA)$_3$ complexes[29]. Given that a Fe^{2+} precursor was used to initiate self-assembly, and as Fe:HA$_3$ centres have low reduction potentials ($E_{red} < -400$ mV)[26], this observation suggests that the protein self-assembly involves the initial formation of Fe^{2+}:HA centres, followed by the thermodynamically favoured oxidation of these species into Fe^{3+} either by the Fe^{3+}–haem centres embedded in each protomer (see Methods) or directly by ambient O_2.

Reversible assembly of dodecameric cages

Despite the stability of isolated BMC2 cages, their formation required an initial crystallization step. We reasoned that slow crystal nucleation or growth kinetics and the high attendant protein and metal concentrations probably increased the fidelity and yield of the complex self-assembly process to produce a discrete supermolecule consisting of 12 protomers and 26 metal ions of two different kinds. Reasoning that strengthened Zn-mediated interactions across the C_2 interface could increase the efficiency of cage self-assembly in solution, we generated two second-generation variants based on BMC2: BMC3 and BMC4 (Extended Data Figs. 1, 2b). In BMC3, the helix 1 surface was engineered to form two Zn-coordination sites (composed of His5, His8, Asp12 and His16) across the C_2 interface, whereas in BMC4 three potential Zn-coordination sites were engineered (one central site composed of two His8–Asp12 pairs as in BMC3 and two peripheral sites composed of Glu2, Glu5, His16 and Glu25). In BMC4, we also removed the Cys63–HA group with the purpose of eliminating any potential undesired assembly products that involve heteromeric Fe^{3+} coordination by Cys63–HA and Cys82–HA.

BMC3 indeed formed dodecameric cages in solution with high yields (>80%) as determined by ns-TEM and AUC measurements (Fig. 3a, Extended Data Fig. 3). The 1.85 Å-resolution crystal structure confirmed the eight Fe centres in the vertices and the two Zn coordination sites in each C_2 interface (twelve in total) as well as nine of the twelve possible

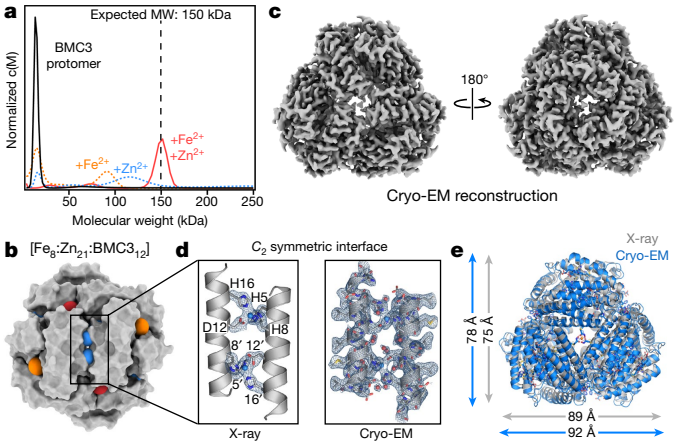


Fig. 3 | Characterization of BMC3 cages. a, AUC characterization of BMC3 self-assembly. **b**, Surface representation of the BMC3 cage (as derived from the crystal structure), oriented to show the incorporation of two Zn ions at the C_2 symmetric interface. **c**, The 2.6 Å-resolution cryo-EM electron density map for the BMC3 cage. **d**, Atomic details of both Zn-binding sites of the two-fold interface overlaid with the electron density $mF_o - DF_c$ omit map from the crystal structure (left) and as observed for the cryo-EM structure (right). Additional side chains and waters are shown for the cryo-EM structure to emphasize the structural robustness of the interface. **e**, Overlay of the BMC3 X-ray and cryo-EM structures to highlight the isotropic expansion of the cage in the absence of crystallographic packing interactions.

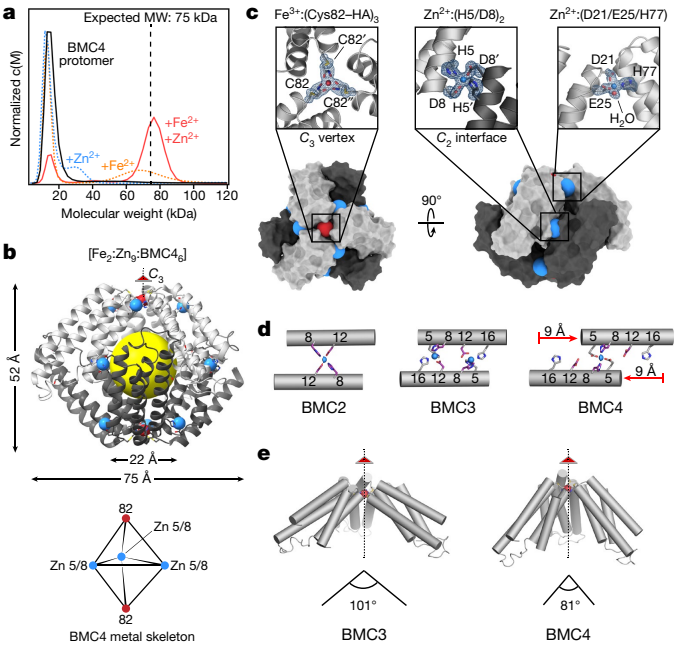


Fig. 4 | Characterization of BMC4 cages. a, AUC characterization of BMC4 self-assembly. **b**, Crystal structure of the BMC4 cage. Fe and Zn ions are represented as red and blue spheres, respectively. The central cavity is highlighted by a yellow sphere. The structural skeleton formed by Fe and Zn ions is shown below the structure. **c**, Surface representations of the BMC4 cage, with metal ions shown as coloured spheres. Atomic details of each metal coordination site are shown in insets, with the $mF_o - DF_c$ electron density omit maps (blue mesh) contoured at 3σ. **d**, Comparison of the C_2 symmetric protein interfaces in different BMC constructs. Residues 8 and 12, which are common to all constructs, are coloured purple. The slippage of the two-fold helix interface to accommodate the hexameric architecture of BMC4 is indicated with red arrows. **e**, Comparison of the apical angle formed by the Fe:(Cys82–HA)$_3$-mediated vertices in BMC3 and BMC4 cages.

peripheral Zn sites, which complete a [8 Fe:21 Zn:12 protomers] architecture (Fig. 3b). Notably, the self-assembly of BMC3 cages in solution was dependent on the presence of both Fe and Zn ions. The absence of either metal ion or the addition of various other first-row transition metal ions instead of Fe led to smaller oligomeric forms of BMC3 or non-specific assemblies (Fig. 3a, Extended Data Fig. 6a). BMC3 cages also formed with a Fe^{3+} precursor, Fe(acetylacetonate)$_3$ (Extended Data Fig. 6b). Consistent with self-assembly under thermodynamic control, the formation of dodecameric cages was independent of the order of addition of Fe or Zn ions. We determined the single-particle cryo-electron microscopy (cryo-EM) structure of isolated BMC3 cages at a resolution of 2.6 Å (Fig. 3c, Extended Data Table 2). A major portion of the assembly could be resolved at 2.0 Å or less (Extended Data Fig. 7). At this resolution, nearly all side chains, Zn coordination sites and some ordered water molecules are clearly distinguished (Fig. 3d). Consistent with the crystallographically observed flexibility of Fe^{3+}:(Cys–HA)$_3$ coordination sites, electron densities in the C_3 vertices are diffuse and some side chains that display high temperature factors in the crystal structure are found in alternative conformations in the cryo-EM structure (Extended Data Fig. 7). These observations confirm that the solution architecture of the BMC3 cage closely reflects the solid-state structure. Probably owing to lattice packing, the latter is isotropically compressed by around 2–3 Å compared to the former (Fig. 3e), which can be accommodated by slight changes in interfacial metal coordination.

Next, we examined the assembly and disassembly behaviour of BMC3 cages in response to different stimuli. BMC3 cages readily disassemble upon treatment with EDTA (Extended Data Fig. 6c). They were stable at 50 °C but dissociated upon incubation at 70 °C (Extended Data Fig. 6d). A key feature of siderophores is that their cellular release of Fe is promoted by the destabilization and labilization of their Fe^{3+}:(HA)$_3$ centres through reduction to the Fe^{2+} form in the cytosol[27]. Along these lines, the treatment of BMC3 cages with a strong reductant (dithionite; $E_{red} < -500$ mV)[30] led to their disappearance and the emergence of monomeric species (Extended Data Fig. 6e). By contrast, a weaker reductant (ascorbate; $E_{red} > -100$ mV at pH 7)[31] with a reduction potential higher than that of Fe^{3+}:(HA)$_3$ had considerably less effect (Extended Data Fig. 6e), suggesting that the disassembly of BMC3 cages occurs through the reduction of the Fe centres. These observations establish BMC3 cages as a distinctive system among natural and artificial protein architectures the assembly and disassembly of which can be controlled through multiple stimuli: chemical, thermal or redox. BMC3 cages can passively encapsulate small fluorogenic molecules in either their lumen or inter-protomer interfaces, retain them for several days and release them upon treatment with EDTA (Extended Data Fig. 8).

Formation of a hexameric cage

Unexpectedly, AUC measurements indicated that the other second-generation variant, BMC4, self-assembled as a hexamer upon Fe and Zn coordination, with yields exceeding 70% (Fig. 4a). The 1.50 Å resolution crystal structure of the BMC4 assembly revealed a D_3 symmetric, cage-like architecture with a composition of [2 Fe:9 Zn:6 protomers], outer dimensions of 75 Å × 50 Å and a cavity volume of more than 7,800 Å^3 (Extended Data Fig. 4). The overall shape is a trigonal bipyramid (Fig. 4b), which is the smallest polyhedral architecture with a sizeable interior cavity that can be constructed from an asymmetric building block. The apical vertex of each pyramidal half is formed by a Fe^{3+}:(Cys82–HA)$_3$ motif shared by three protomers (Fig. 4b). These C_3 symmetric vertices are further reinforced by Zn^{2+} ions that link pairs of protomers through Asp21, Glu25 and His77 coordination (Fig. 4c). The pyramids are joined by three equatorial, C_2 symmetric vertices mediated by Zn centres coordinated to Glu5 and His8 (Fig. 4b, c). A comparison to the BMC2 and BMC3 cages indicates that this unexpected coordination motif requires an approximately 9 Å slip of each protomer along the C_2 symmetric interfaces (Fig. 4d). The shift markedly reduces

the C_2 symmetric contact area between protomers, effectively transforming the edges in the tetrahedral BMC2 and BMC3 cages to vertices in the trigonal bipyramidal BMC4 cages. BMC4 cages exhibited similar thermal stability to BMC3 cages, with both species disassembling at below 70 °C. The thermal robustness of the BMC3 and BMC4 cages appear to be limited, at least in part, by the relative instability of the individual protomers (Extended Data Fig. 6d).

The large structural transformation is accompanied by a reduction in the apical angle formed at the Fe^{3+}:$(Cys82–HA)_3$-mediated vertices from 101° in the BMC2 and BMC3 cages to 81° in the BMC4 cage (Fig. 4e). This observation highlights the conformational adaptability of the Fe^{3+}:$(Cys82–HA)_3$ coordination motif, enabling it to accommodate different polyhedral geometries. Such behaviour is reminiscent of the interfacial flexibility in some icosahedral virus capsids in which the same protomer can form hexamers on capsid faces and pentamers on capsid vertices[7]. It is worth noting that BMC4 contains all of the Zn-coordinating residues on helix 1 to form the C_2 symmetric interfaces observed in the dodecameric BMC3 cage, indicating that the self-assembly process selects an alternative interfacial arrangement of lower free energy, enabled by the reversibility of metal coordination interactions. In terms of protein design, a caveat of interfacial flexibility is that it may lead to nonspecific or unintended self-assembly products, although it can also allow error correction during self-assembly and increase tolerance to design imperfections.

Conclusions

The self-assembly and function of biomolecular systems are predicated upon their specificity, stability and adaptiveness, which, in turn, are enabled by extensive networks of non-covalent interactions. Here, we have shown that fundamental concepts in inorganic coordination chemistry can be applied to achieve all of these attributes in protein self-assembly and, specifically, to construct complex polyhedral protein architectures from a simple, asymmetric building block. Despite their minimal design footprints, these cage-like architectures are distinguished by their structural compactness and responsiveness—hallmarks of evolved systems such as viral capsids. Key to our construction strategy was the reimagination of a biological coordination motif, hydroxamic acid, within a new structural context: as a new amino acid side chain with the ability to chelate hard metal ions. This example expands the growing lexicon of post-translational modifications that broaden the chemical scope of proteins.

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/s41586-019-1928-2.

1. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem*. **84**, 551–575 (2015).
2. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98**, 2217–2221 (2001).
3. Bai, Y., Luo, Q. & Liu, J. Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev*. **45**, 2756–2767 (2016).
4. Hamley, I. W. Protein assemblies: nature-inspired and designed nanostructures. *Biomacromolecules* **20**, 1829–1848 (2019).
5. Churchfield, L. A. & Tezcan, F. A. Design and construction of functional supramolecular metalloprotein assemblies. *Acc. Chem. Res*. **52**, 345–355 (2019).
6. Yeates, T. O. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys*. **46**, 23–42 (2017).
7. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol*. **269**, 665–675 (1997).
8. Lawson, D. M. et al. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–544 (1991).
9. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
10. Bale, J. B. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
11. Cristie-David, A. S. & Marsh, E. N. G. Metal-dependent assembly of a protein nano-cage. *Protein Sci*. **28**, 1620–1629 (2019).
12. Fletcher, J. M. et al. Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
13. Malay, A. D. et al. An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438–442 (2019).
14. Pluth, M. D., Bergman, R. G. & Raymond, K. N. Acid catalysis in basic solution: a supramolecular host promotes orthoformate hydrolysis. *Science* **316**, 85–88 (2007).
15. Chakrabarty, R., Mukherjee, P. S. & Stang, P. J. Supramolecular coordination: self-assembly of finite two- and three-dimensional ensembles. *Chem. Rev*. **111**, 6810–6918 (2011).
16. Yoshizawa, M., Klosterman, J. K. & Fujita, M. Functional molecular flasks: new properties and reactions within discrete, self-assembled hosts. *Angew. Chem. Int. Ed*. **48**, 3418–3438 (2009).
17. Mal, P., Breiner, B., Rissanen, K. & Nitschke, J. R. White phosphorus is air-stable within a self-assembled tetrahedral capsule. *Science* **324**, 1697–1699 (2009).
18. Liu, Y., Hu, C., Comotti, A. & Ward, M. D. Supramolecular Archimedean cages assembled with 72 hydrogen bonds. *Science* **333**, 436–440 (2011).
19. Mateu, M. G. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys*. **531**, 65–79 (2013).
20. Sun, X., Johnson, D. W., Caulder, D. L., Raymond, K. N. & Wong, E. H. Rational design and assembly of $M_2M'_3L_6$ supramolecular clusters with C_{3h} symmetry by exploiting incommensurate symmetry numbers. *J. Am. Chem. Soc*. **123**, 2752–2763 (2001).
21. Smulders, M. M. J., Jiménez, A. & Nitschke, J. R. Integrative self-sorting synthesis of a $Fe_8Pt_6L_{24}$ cubic cage. *Angew. Chem. Int. Ed*. **51**, 6681–6685 (2012).
22. Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem*. **4**, 375–382 (2012).
23. Suzuki, Y. et al. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373 (2016).
24. Radford, R. J., Nguyen, P. C. & Tezcan, F. A. Modular and versatile hybrid coordination motifs on α-helical protein surfaces. *Inorg. Chem*. **49**, 7106–7115 (2010).
25. Pearson, R. G. Hard and soft acids and bases. *J. Am. Chem. Soc*. **85**, 3533–3539 (1963).
26. Wong, G. B., Kappel, M. J., Raymond, K. N., Matzanke, B. & Winkelmann, G. Coordination chemistry of microbial iron transport compounds. 24. Characterization of coprogen and ferricrocin, two ferric hydroxamate siderophores. *J. Am. Chem. Soc*. **105**, 810–815 (1983).
27. Crumbliss, A. L. Iron bioavailability and the coordination chemistry of hydroxamic acids. *Coord. Chem. Rev*. **105**, 155–179 (1990).
28. Ni, T. W. & Tezcan, F. A. Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew. Chem. Int. Ed*. **49**, 7014–7018 (2010).
29. Failes, T. W. & Hambley, T. W. Crystal structures of tris(hydroxamato) complexes of iron(III). *Aust. J. Chem*. **53**, 879–881 (2000).
30. Mayhew, S. G. The redox potential of dithionite and $SO^{\cdot-}_2$ from equilibrium reactions with flavodoxins, methyl viologen and hydrogen plus hydrogenase. *Eur. J. Biochem*. **85**, 535–547 (1978).
31. Borsook, H. & Keighley, G. Oxidation-reduction potential of ascorbic acid (vitamin C). *Proc. Natl Acad. Sci. USA* **19**, 875–878 (1933).

Methods

Synthesis of the IHA ligand

O-tritylhydroxylamine was synthesized as previously described[32]. Chloroacetyl chloride (0.58 ml, 7.3 mmol) was dissolved in 2 ml CH_2Cl_2 and added dropwise to a suspension of *O*-tritylhydroxylamine (2.0 g, 7.3 mmol) and *N*,*N*-diisopropylethylamine (2.5 ml, 14.5 mmol) in 15 ml CH_2Cl_2 at 0 °C. The reaction mixture was gradually warmed to room temperature and stirred at room temperature for an hour. An additional 15 ml CH_2Cl_2 was added and the reaction was extracted with H_2O (3×30 ml). The CH_2Cl_2 solution was collected and evaporated to dryness. A solution containing 15 ml of CH_2Cl_2 with 10% (v/v) trifluoroacetic acid was added and the solution was stirred for 30 min. The crude product was purified by silica gel chromatography using a gradient of 0–100% ethyl acetate in hexanes as the eluent. The product was visualized using a $FeCl_3$ stain. Yield, 55%. Measured molecular weight (*m*/*z*): 108.37 [M − H^+]; calculated: 107.99 [M − H^+]. ^{1}H NMR: (300 MHz, DMSO-d_6) δ 10.88 (s, 1H), δ 9.15 (s, 1H), δ 3.93 (s, 2H). ^{13}C NMR: (500 MHz, DMSO-d_6) δ 162.88, δ 40.45. 2-chloro-*N*-hydroxyacetamide (400 mg, 3.7 mmol) and NaI (2.7 g, 18.3 mmol) were refluxed in 30 ml acetone for 1 h. The reaction mixture was purified by silica gel chromatography with 100% ethyl acetate as the eluent and dried in vacuo. Yield, >90%. Measured molecular weight (*m*/*z*): 223.85 [M + Na^+]; calculated: 223.95 [M + Na^+]. ^{1}H NMR: (300 MHz, DMSO-d_6) δ 10.81 (s, 1H), δ 9.09 (s, 1H), δ 3.51 (s, 2H). ^{13}C NMR: (500 MHz, DMSO-d_6) δ 164.83, δ −2.01.

Protein expression and purification

All constructs (Supplementary Table 1) were derived from the parent pET-20b(+) plasmid containing the *CFMC1* gene via site-directed mutagenesis as previously described[28,33,34]. The appropriate plasmids were transformed into BL21(DE3) *Escherichia coli* cells (New England Biolabs) housing a CCM (cytochrome C maturation) cassette containing a chloramphenicol-resistance marker and expressed as previously described[35] with minor adjustments. Multiple 2.8-l flasks containing 1.5 l of LB medium were shaken at 200 rpm for 12 h at 37 °C and then at 100 rpm for an additional period of around 7 h. Cells were collected by centrifugation (5,000 rpm for 10 min at 4 °C), resuspended in a buffered solution containing 5 mM sodium acetate (NaOAc) (pH 5.0) and 2 mM dithiothreitol (DTT) and lysed via sonication. The pH of the crude lysate was first raised to 10 using NaOH to precipitate cellular contaminants, then reduced to pH 4.5. After centrifugation (12,000 rpm for 20 min at 4 °C), the clarified supernatant was decanted and diluted 15-fold with additional buffer. This solution was applied to a CM sepharose gravity column (GE Healthcare) pre-equilibrated with the aforementioned buffer and subjected to multiple buffer washes before elution using a stepwise-gradient of NaCl (0–0.5 M). Peak elution fractions were combined and concentrated using a 400-ml Amicon Stirred Cell (Millipore) and buffer-exchanged by overnight dialysis against a buffered solution containing 10 mM phosphate (pH 8.0) at 4 °C. Next, the protein was purified via a DuoFlow workstation station fitted with a Macroprep High Q-cartridge column (BioRad) and eluted using a linear gradient over 0–0.5 M NaCl. Fractions that exhibited an RZ ratio (A_{421}/A_{280}) > 4.4 were pooled, treated with 2 mM EDTA for 1 h, concentrated, and buffer-exchanged into 20 mM tris(hydroxymethyl) aminomethane (Tris) (pH 7.5) pretreated with Chelex 100 resin (Bio-Rad), via desalting column (Econo-Pac 10DG pre-packed columns, BioRad). Demetallated and purified proteins were concentrated to around 2 mM and stored at 4 °C.

Protein labelling and post-labelling purification

Purified protein solutions were treated with a 100-fold excess of DTT and placed in an anaerobic Coy chamber for approximately 2 h for slow degassing to remove dissolved oxygen. The fully reduced protein solution was buffer-exchanged into 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5) via desalting column to remove DTT, and the concentration of the resulting protein solution was determined spectroscopically (Agilent 8452 spectrophotometer) using the $\varepsilon_{421\,(red)} = 162{,}000\ M^{-1}\,cm^{-1}$ (ref. [34]). Solid iodohydroxamic acid (IHA) was dissolved in 100 µl degassed DMF to generate solutions containing a 15-fold excess IHA per protein monomer, which were then added to protein aliquots and incubated overnight. The HA-functionalized variants were removed from the Coy chamber and separated from unreacted or partially reacted protein via FPLC using a Q-column equilibrated with 10 mM *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES) (pH 9.3) and 2 mM DTT and eluted using a linear gradient over 0–0.5 M NaCl. Protein functionalization was verified using electrospray ionization mass spectrometry (ESI-MS; Extended Data Fig. 1) and the resulting protein solutions were buffer-exchanged into demetallated 20 mM Tris (pH 7.5) via desalting column, concentrated to around 2 mM and stored at 4 °C for further use.

Redesign of CFMC1 interfaces

To render the CFMC1 protomer competent for the bimetallic design strategy, we first performed the following mutations to remove potential competitive interactions: C67E, H59S and H73N. A negative design strategy was then used to disrupt a noncovalent dimerization interface found in CFMC1, leading to the mutations A34Q and A38Q. We further identified the dearth of protein–protein interactions within the core and periphery of the three-fold axis engulfing the 82 position as a likely contributor to poor cage assembly and crystallization in general. Accordingly, as a means to facilitate cage formation, we adopted Rosetta-prescribed mutations at the following positions: A24T, Q25(T/E), N80K and E81Q.

Crystallography

Screening and crystallization of all BMC variants were conducted via sitting drop vapour diffusion. In brief, solutions containing 2.1–2.2 mM BMC protomer were mixed with mother liquor (1 µl + 1 µl) and equilibrated against 200-µl reservoir volumes. Supplementary Table 2 details the experimental conditions for crystal growth. Protein solutions of BMC1 and BMC4 were first incubated with $FeSO_4$ for 1 h before mixing with $ZnCl_2$. Solutions of BMC2 and BMC3 were mixed with $FeSO_4$ and $ZnCl_2$ stock solutions and were immediately combined with the mother liquor (to prevent rapid aggregation of the proteins functionalized with two HA units). Crystals for all mutants typically appeared within several hours and were collected within a week of maturation. Crystals were cryoprotected by submersion into perfluoropolyether cryo oil (Hampton Research) for a few seconds and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K at either the Advanced Light Source (ALS) beamline BL 8.3.1 (using 1.12 Å radiation for BMC3 and 1.33 Å radiation for BMC4) or at the Stanford Synchrotron Radiation Lightsource (SSRL) beamlines 9-2 (using 0.98 Å radiation for BMC2) and 12-2 (using 0.98 Å radiation for BMC1). Data integration was performed using the XDS Program Package, truncated at $CC_{1/2} > 0.5$ (ref. [36]). Datasets of the same structure recorded at different wavelengths were scaled to the highest resolution dataset with XSCALE[37,38]. Phaser-MR[39] was used to carry out molecular replacement with search models based on the CMFC1 monomer (PDB ID: 3M4B) containing the expected side chain mutations (generated in Pymol[40]) but lacking HA. Rigid-body and structure refinement was performed using multiple rounds of Phenix.refine[39], interspersed with manual model rebuilding and metal/ligand placement with Coot[41]. Restraint files for the Cys-hydroxamic acid conjugates were generated using phenix.eLBOW to maintain the distances Cys-SG–HA-C1 (1.816 Å ± 0.02 Å) and angles Cys-CB–Cys-SG–HA-C1 as well as Cys-SG–HA-C1–HA-C2 (both 109° ± 3°) during refinement. Where necessary, the metal binding geometry of the hydroxamic acids was restrained to the distances Fe–HA-O1 (1.98 Å ± 0.05 Å) and Fe–HA-O2 (2.057 Å ± 0.05 Å) as well as through a planarity constraint for the atoms Fe, HA-O1, HA-O2 and HA-C1 following data from a high-resolution structure of Fe(III)-tris-benzhydroxamate trihydrate[42]. Simulated

annealing omit maps (metal atoms and side chain ligands) were generated for each metal binding site and model accuracy was assessed critically against these omit maps. Electron density maps were generated using Phenix and all molecular graphics images were produced with either Pymol or the UCSF ChimeraX package from the Computer Graphics Laboratory, University of California, San Francisco[43].

Crystallographic metal content analysis

Metal ions, with their relatively high-energy inner electrons, can absorb and resonate with soft X-rays; this leads, among other effects, to differences in the intensity of otherwise centro-symmetric Bragg diffraction peaks used for X-ray crystallography. Density maps calculated from these differences are routinely used to locate and identify metal ions in protein crystals. The magnitude of this anomalous X-ray diffraction varies with the X-ray energy, with stark differences around the energies of the K- and L-shell electrons of the respective elements allowing one to discern between elements at a position in question, if diffraction datasets are measured at the appropriate wavelengths. For a visual analysis of the bound metals, the scaled datasets of different wavelengths were used separately as an input for a single phenix.refine run, each with the final model of the highest resolution dataset. Importantly, only the B-factor or occupancy were allowed to change during refinement, resulting in anomalous difference density maps for each wavelength. Using these maps, isomorphous difference maps from data at wavelengths above and below the respective element K-edges were generated (if applicable) with Phenix and were inspected manually (Extended Data Fig. 5). To gain a more quantitative understanding of the identity of the bound metals for each site, the anomalous difference signal of each dataset was used to generate CCP4 format maps with phenix.mtz2map. The generated maps were used subsequently as inputs to calculate the mean signal in a sphere of 1 Å radius centred on each metal atom with the program MAPMAN (Uppsala Software Factory). For each pair of datasets above and below a metal-absorption edge, the ratio of the anomalous signal above and below the edge for every metal atom was tabulated. The experimental ratio was compared to the theoretical ratio for both Fe and Zn (Extended Data Fig. 5) according to http://skuld.bmsc.washington.edu/scatter, as calculated using the Cromer and Liberman approximation. Theoretical ratios were also calculated for hypothetical mixed occupancy Fe/Zn metal sites and compared to experimentally observed values (Supplementary Tables 3–6).

Protein cage sample preparation

Self-assembled cages. All samples were prepared in a low-O_2 atmosphere (Coy glovebox) to minimize undesired oxidation of Fe^{2+} ions before self-assembly. Protein solutions containing 20 μM BMC3 or 100 μM BMC4 in 20 mM Tris (pH 8.5) were incubated with either 20 μM $FeSO_4$ and 60 μM $ZnCl_2$ for BMC3 or 50 μM $FeSO_4$ and 200 μM $ZnCl_2$ for BMC4 for 2–3 h to yield the metallated cages. We note that the addition of $FeSO_4$ was followed by a small but observable change in the colour of the solution from red to pink, attributed to a shift of the haem Soret band to longer wavelengths, which suggested reduction of the haem by the ferrous ions and generation of ferric ions in close proximity to HA group(s). The final BMC3 solutions were then concentrated sevenfold before overnight incubation to improve the total cage yield. After self-assembly, the resulting solutions were diluted back to their original concentrations with the self-assembly buffer before characterization.

Dissolved crystals. Fe:Zn:BMC1 and Fe:Zn:BMC2 crystals were dissolved using buffer containing 100 mM HEPES (pH 7.5), 200 mM $MgCl_2$ and 800 μM $ZnCl_2$. Mature crystals were removed from their pedestal droplet, briefly submerged in fresh buffer to remove uncrystallized protein and surface-bound precipitates, and transferred into a new sitting drop crystallization well containing 8 μl buffer solution. The crystals were physically crushed with a small metal scalpel and vigorously pipetted until a large portion of the crystals dissolved. Undissolved crystals were removed by centrifugation (10,000 rpm for 5 min at 25 °C), yielding a light-red supernatant and dark-red precipitate.

Negative-stain transmission electron microscopy

A 4-μl droplet of BMC cages (either self-assembled or from dissolved crystals) was deposited onto formvar/carbon-coated Cu grids (Ted Pella) (pretreated by negative-mode glow discharge up to 15 min beforehand) and allowed to bind for 5 min. The grids were then washed with 50 μl MilliQ water, blotted using Whatman filter paper and stained using 2% uranyl acetate solution in water and blotted again. Grids were imaged using a FEI Sphera transmission electron microscope operating at 200 keV, equipped with an LaB_6 filament and a Gatan 4K CCD camera. Micrographs were collected using objective-lens underfocus settings ranging from 250 nm to 2 μm and analysed using Fiji (http://fiji.sc/Fiji).

Oligomerization state determination using AUC

Sedimentation velocity measurements were performed at 41,000 rpm and 25 °C using an XL-1 analytical ultracentrifuge (Beckman Coulter) equipped with an AN-60 Ti rotor. Data processing was performed using Sedfit[44] with the following parameters as calculated using SEDNTERP: viscosity: 0.01000 poise, density: 0.9988 g/ml (self-assembled samples) or viscosity: 0.0113191 poise, density: 1.0196 g/ml (dissolved crystals), and a partial specific volume of 0.7313 ml/g for all samples. All reported results correspond to a confidence level of 0.95.

Preparation of samples involving crystal dissolution

Dissolved crystal samples (BMC1 and BMC2), prepared as described above at ambient conditions, were diluted to 350 μl with 10 mM HEPES (pH 7.5), 200 mM $MgCl_2$ and 800 μM $ZnCl_2$. The solution was clarified via brief centrifugation in order to remove crystal debris and the supernatant was placed inside the cells.

Calculation of BMC void volumes

Structures of complete cage assemblies for BMC2, BMC3 and BMC4 were generated via the application of crystallographic symmetry operations to the fully refined asymmetric unit of each construct. These coordinates were recentred at the origin and stripped of waters, hydrogens, alternative conformations and crystallization reagents (PEG-400). Volumetric maps and volumes for the internal cavity of each cage were calculated using VOIDOO[45], and are reported as the solvent-accessible volume for a 1.4 Å rolling probe on a 0.25 Å grid spacing for all constructs. The cavity volumes using these parameters were determined to be approximately 32,700 $Å^3$ (BMC2), 32,700 $Å^3$ (BMC3) and 7,900 $Å^3$ (BMC4).

Solution self-assembly, disassembly and thermal stability of BMC3 and BMC4

Assembled samples were prepared as described above and placed inside the AUC measurement cells anaerobically (20 μM BMC3 and 100 μM BMC4). Disassembly of the cages via metal-ion removal was performed by treating the protein cages with 2 mM EDTA for 1 h. Redox-controlled disassembly of the protein commenced by the addition of either 5 mM sodium dithionite or 5 mM sodium ascorbate to the cage solution anaerobically and subsequent incubation of the samples at around 22 °C for 16 h. Samples were then loaded into the AUC measurement cell.

For thermal stability measurements, samples were placed in a thermoregulated chamber preequilibrated at the appropriate temperature for 2 h, and subsequently removed from the chamber and equilibrated at room temperature for 30 min before AUC analysis. Circular dichroism (CD) spectra were measured using an Aviv 215 spectrometer. CD measurements were performed using 10 μM protein in a buffered solution containing 20 mM Tris (pH 8.5). Thermal melts were measured at 222 nm at a 1 nm slit width, scanning at 1-nm intervals with a 1-s integration time. Measurements were taken from 25 °C to 85 °C at 2-degree intervals

with a 2 min equilibration at each temperature. Unfolding data were fit to a two-state model with van't Hoff's enthalpy using the CalFitter web server[46].

Cryo-EM sample preparation

Self-assembled BMC3 cages were removed from the anaerobic Coy chamber immediately before grid preparation. A 3.5-µl aliquot of self-assembled BMC3 cages was dropped onto holey carbon grids (Electron Microscopy Sciences, Quantifoil R1.2/1.3 holey carbon on 300 mesh copper) that had been freshly glow-discharged for 30 s. The initial application of the sample was side blotted manually with Whatman No. 1 filter paper immediately followed by a secondary application of a 3.5-µl aliquot, blotted for 3.5 s and plunge-frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI).

Cryo-EM data acquisition and image processing

Samples were imaged on a Titan Krios G3 transmission electron microscope (FEI) operating at 300 kV equipped with a K2 Summit direct electron detector (Gatan) and a GIF Quantum energy filter. The slit-width of the energy filter was set to 10 eV. Movies were collected at a magnification of 165,000× in EFTEM mode giving a physical pixel size of 0.84 Å/pixel. In total, 4,672 movie stacks (50 frames/movie) were collected using a 10 s exposure at a dose rate of 1.2 $e^-/Å^2$ per frame for a total electron dose of 60 $e^-/Å^2$ per movie. Objective-lens underfocus settings varied between 0.6 µm and 1.6 µm. Data collection was performed using software EPU (FEI). All image processing was performed in the Relion-3.0 pipeline[47]. Motion correction and dose weighting were performed using MotionCor2[48], and defocus values were estimated with Gctf[49] using a pixel size of 0.8 Å/pixel. A total of 3,513 movie stacks were selected following motion correction and CTF estimation, and 805,156 particles were auto-picked using RELION-3.0. Particle images were extracted and binned by 2 (1.6 Å/pixel, 100 pixel box size) and subjected to two-dimensional (2D) classification. A total of 444,247 particles were selected corresponding to good 2D class averages and subjected to three-dimensional (3D) classification imposing *T* symmetry and using an initial model generated from a subset of the particles. A total of 129,653 particles were chosen from a 3D class showing strong secondary-structural elements and subjected to 3D auto-refinement with *T* symmetry. The particles were re-centred and re-extracted to their original pixel size of 0.8 Å/pixel. These particles were subjected to 3D auto-refinement with *T* symmetry and the yield map was then postprocessed towards 2.6 Å resolution based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion. The pixel size of the map was manually adjusted using Relion image handler to match the physical pixel size of the images. Local resolution was calculated in Relion 3.0 using ResMap[50].

Model building and refinement

The BMC3 crystal structure (PDB ID: 6OT7) stripped of hydrogens and waters was used as an initial model and manually docked into the cryo-EM density using UCSF Chimera[51]. The structural model was subject to real space refinement in Phenix against the cryo-EM map with geometry restraints for the Fe-binding sites and molecular coordinates for the Cys–HA ligand. The atomic model was manually improved using Coot. Tightly bound waters were identified based on clear density in the EM density map. Whereas the structural flexibility of the hydroxamate sites manifested in poor electron density, the twofold interface was much more rigid and unambiguous density was observed for Zn-binding. A tryptophan at the 66 position, which had shown high-temperature factors in the BMC3 crystal structure, was identified in multiple conformations in the EM density map. The final model was subjected to real space refinement using Phenix[39] and evaluated using MolProbity[52]. All molecular graphics images were rendered in PyMoL or UCSF ChimeraX.

Encapsulation of rhodamine in BMC3 cages

BMC3 cages were self-assembled in a low-O_2 atmosphere in the presence of rhodamine for the passive encapsulation of the dye. Solutions containing 20 µM BMC3 were incubated with 20 µM $FeSO_4$, 60 µM $ZnCl_2$ and 2 mM rhodamine. A control sample was prepared in the absence of added metal ions (20 µM BMC3 incubated with 2 mM rhodamine). Samples were incubated for 2–3 h and concentrated sevenfold before overnight incubation. Protein solutions were buffer exchanged on a PD-10 desalting column using a buffer containing 20 mM Tris (pH 8.5) (with 5 µM $FeSO_4$ and 10 µM $ZnCl_2$ supplemented for solutions already containing metal ions) to separate unassociated dye from protein. Cage solutions were split in two: one half was treated with 1 mM EDTA and incubated for 2 h before washing. All protein solutions were additionally washed three times using a centrifugal filter to completely remove any remaining free rhodamine.

Fluorescence measurements were performed using 6 µM protein solutions after the previously mentioned wash steps. For each sample, an excitation wavelength of 555 nm with a 2 nm slit width was used and emission was measured between 560 and 650 nm with a 2 nm slit width and 0.2 s integration time. For the time-course experiments, cages encapsulating rhodamine were washed three times after 4 days and after 7 days and diluted to 6 µM before fluorescence measurements. AUC measurements were performed at the λ_{max} of the cytochrome (415 nm) and at the λ_{max} of rhodamine (555 nm) to assess whether there was a sufficiently large rhodamine signal associated with BMC3 cages. Ultraviolet–visible light (UV-vis) absorbance measurements were performed on each solution to measure the protein and rhodamine concentrations. Difference spectra were taken between each rhodamine-incubated sample and BMC3 protomer to eliminate any background signal.

Statistics and reproducibility

All reported samples represent technical replicates. The ns-TEM micrograph of BMC2 cages after 3D crystal dissolution (Fig. 2a) is representative of experiments repeated independently four times. AUC experiments for BMC2 (Fig. 2b) were performed in duplicate. Self-assembly of BMC3 cages and subsequent AUC characterization (Fig. 3a) were performed the following number of times: BMC3 protomer ($n = 2$), $+Fe^{2+}$ ($n = 4$), $+Zn^{2+}$ ($n = 4$), $+Fe^{2}$, $+Zn^{2+}$ ($n = 6$). Self-assembly of BMC4 cages and subsequent AUC characterization (Fig. 4a) was performed the following number of times: BMC4 protomer ($n = 2$), $+Fe^{2+}$ ($n = 1$), $+Zn^{2+}$ ($n = 1$), $+Fe^{2}$, $+Zn^{2+}$ ($n = 5$). Mass spectra (Extended Data Fig. 1c–f) were collected in duplicate for native and HA-labelled proteins; AUC experiments were performed in duplicate. TEM characterization of BMC constructs (Extended Data Fig. 3) were performed the following number of times: dissolved BMC1 crystals ($n = 1$), dissolved BMC2 crystals ($n = 4$), BMC2 +EDTA ($n = 2$), self-assembled BMC3 cages ($n = 5$), BMC3 +EDTA ($n = 4$). AUC experiments following the incubation of BMC3 with first-row transition metals (Extended Data Fig. 6a) were performed in duplicate. Self-assembly of BMC3 in the presence of Fe(acetylacetonate)$_3$ (Extended Data Fig. 6b) was performed in duplicate. BMC3 cage disassembly in the presence of EDTA (Extended Data Fig. 6c) was performed in triplicate. AUC characterization of BMC variants after equilibration at different temperatures (Extended Data Fig. 6d) was performed the following number of times: BMC3 at 50 °C ($n = 2$), BMC3 at 70 °C ($n = 2$), BMC4 at 50 °C ($n = 3$), BMC4 at 70 °C ($n = 3$), BMC4 at 90 °C ($n = 3$). Thermal unfolding of BMC variants as measured by CD spectroscopy (Extended Data Fig. 6d) was performed in duplicate. Treatment of BMC3 cages with chemical reductants (Extended Data Fig. 6e) was performed in duplicate. Cryo-EM characterization of BMC3 cages was performed after collecting 4,672 movie stacks. Extended Data Figure 7a shows a representative micrograph and three representative 2D class averages. Fluorescence characterization of BMC3 samples incubated with rhodamine were performed (Extended Data Fig. 8a) in triplicate. AUC characterization

of BMC3 cages encapsulating rhodamine (Extended Data Fig. 8b) was performed in duplicate. UV-vis characterization of BMC3 samples incubated with rhodamine (Extended Data Fig. 8c, d) was performed in triplicate. Repeated fluorescence characterization of a solution containing BMC3 cages encapsulating rhodamine (Extended Data Fig. 8e) was performed in duplicate.

Reporting summary

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

Data availability

The principal data supporting the findings of this work are available within the figures and the Supplementary Information. Additional data that support the findings of this study are available from the corresponding author on request. Structural data obtained by X-ray crystallography and cryo-EM have been deposited into the RCSB PDB and EMDB data banks with the following accession codes: 6OT4 (BMC2), 6OT7 (BMC3), 6OT8 (BMC4), 6OT9 (BMC1) and 6OVH (BMC3 cryo-EM) in the PDB or EMD-20212 at The Electron Microscopy Data Bank.

32. Michalak, K., Wicha, J. & Wójcik, J. Studies towards dynamic kinetic resolution of 4-hydroxy-2-methylcyclopent-2-en-1-one and its *E*-*O*-trityloxime. *Tetrahedron* **72**, 4813–4820 (2016).
33. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. **8**, 91 (2008).
34. Faraone-Mennella, J., Tezcan, F. A., Gray, H. B. & Winkler, J. R. Stability and folding kinetics of structurally characterized cytochrome *c*-b_{562}. *Biochemistry* **45**, 10504–10511 (2006).
35. Bailey, J. B., Subramanian, R. H., Churchfield, L. A. & Tezcan, F. A. in *Methods in Enzymology* Vol. 580 (ed. Pecoraro, V. L.) 223–250 (Academic, 2016).
36. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
37. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
38. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr*. **66**, 125–132 (2010).
39. Terwilliger, T. C. et al. Phenix—a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
40. Schrodinger, LLC. The PyMOL Molecular Graphics System. version 1.3 (2010).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
42. Lindner, H. J. & Gottlicher, S. Die Kristall- und Molekulstruktur des Eisen(III)-benzhydroxamat-Trihydrates. *Acta Crystallogr. B* **25**, 832–842 (1969).
43. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. **27**, 14–25 (2018).
44. Schuck, P. A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys. Chem*. **108**, 187–200 (2004).
45. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).
46. Mazurenko, S. et al. CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Res*. **46**, W344–W349 (2018).
47. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
48. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
49. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol*. **193**, 1–12 (2016).
50. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
51. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem*. **25**, 1605–1612 (2004).
52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

Acknowledgements This work was supported by the US Department of Energy (Division of Materials Sciences, Office of Basic Energy Sciences; DE-SC0003844; for the design strategy, EM imaging and analysis, and biochemical analysis) and by the National Science Foundation (Division of Materials Research; DMR-1602537; for crystallographic analysis). E.G. acknowledges support by an EMBO Long-Term Postdoctoral Fellowship (ALTF 1336-2015). J.E. acknowledges support by a DFG Research Fellowship (DFG 393131496). R.H.S. was supported by the National Institute of Health Chemical Biology Interfaces Training Grant UC San Diego (T32GM112584). We acknowledge the use of the UCSD Cryo-EM Facility, which is supported by NIH grants to T.S.B. and a gift from the Agouron Institute to UCSD. Crystallographic data were collected either at Stanford Synchrotron Radiation Lightsource (SSRL) or at the Lawrence Berkeley Natural Laboratory on behalf of the Department of Energy.

Author contributions E.G. conceived the project, and designed and performed most experiments. R.H.S. and R.G.A. performed and processed the ns-TEM data and performed structural modelling and analysis. R.H.S. conducted encapsulation experiments. J.E. performed crystallographic analysis. J.B.B. and J.A.C. synthesized the IHA ligand. X.Y., T.B. and T.S.B. performed the cryo-EM data collection and processing. F.A.T. conceived and directed the project and wrote the manuscript with contributions from all authors.

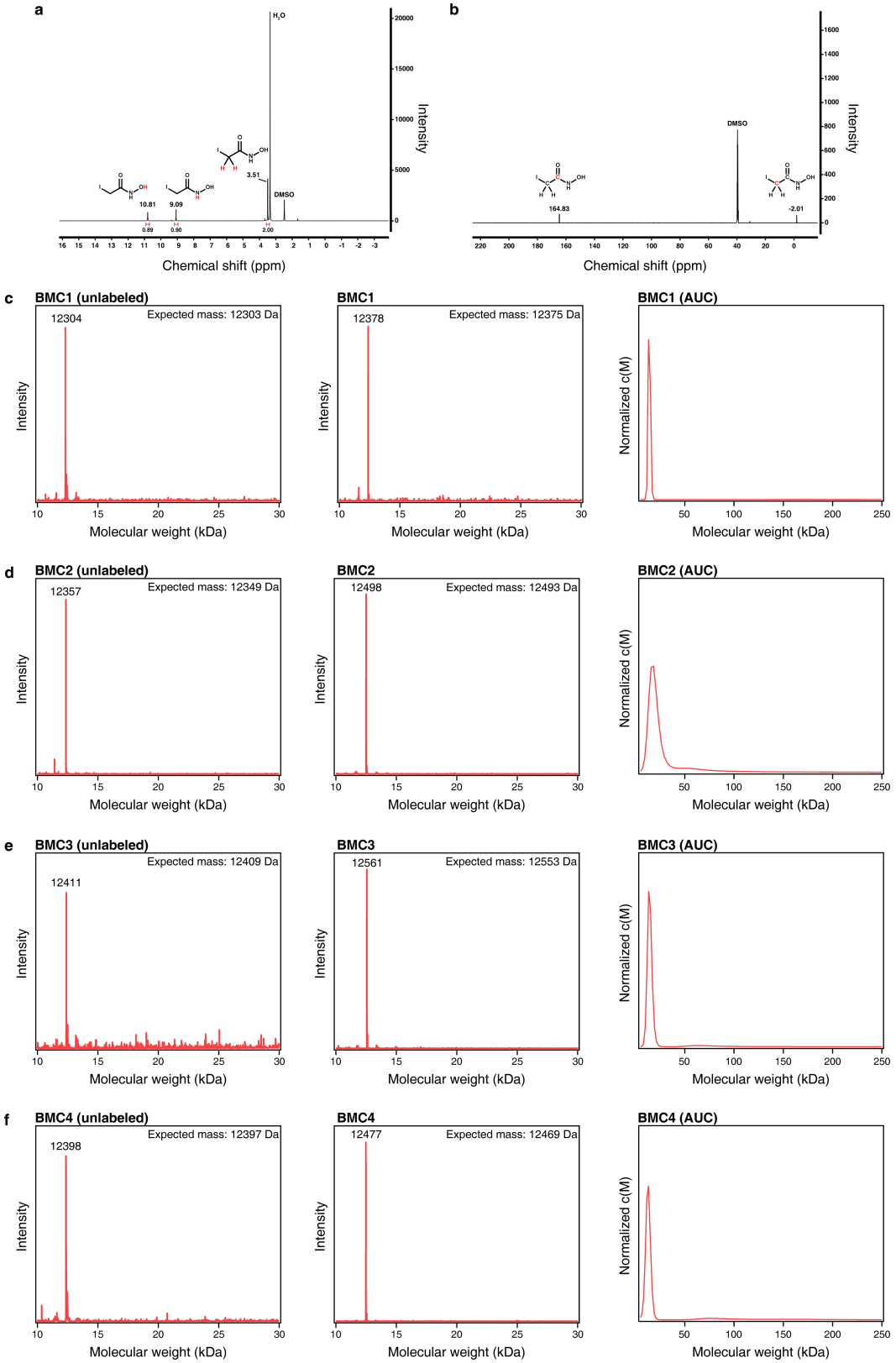
Competing interests The authors declare no competing interests.

Additional information

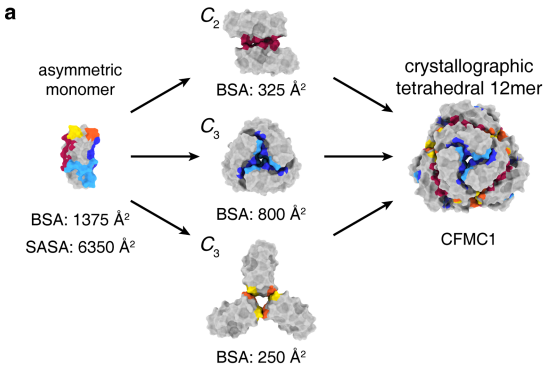
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1928-2.

Correspondence and requests for materials should be addressed to F.A.T.

Peer review information *Nature* thanks Jack Johnson, Todd Yeates and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

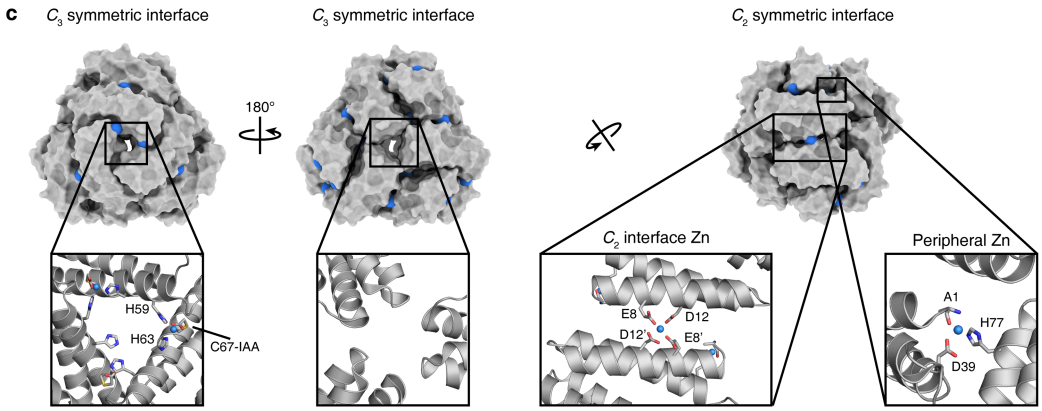


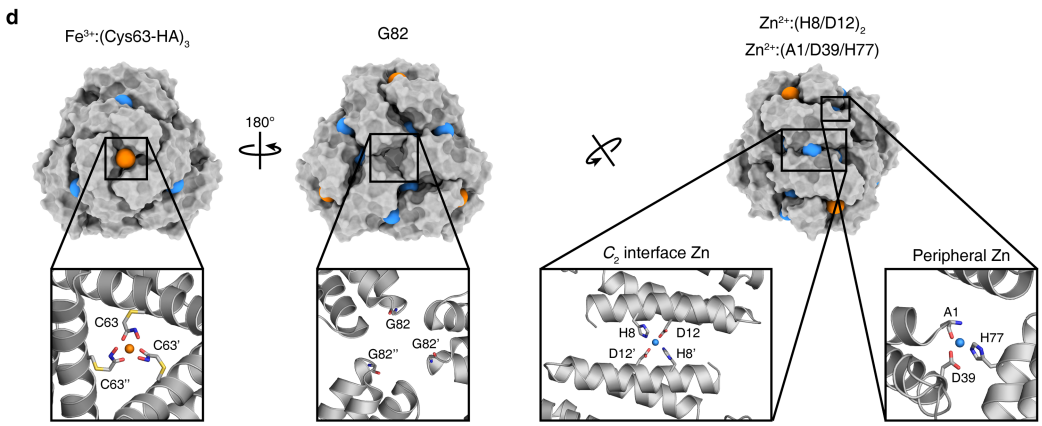
Extended Data Fig. 1 | Characterization of the IHA ligand and the BMC constructs. a, **b**, NMR spectra of *N*-hydroxy-2-iodoacetamide in DMSO-d_6 for ^{1}H (**a**) and ^{13}C (**b**). **c**–**f**, ESI-MS of as-isolated and HA-functionalized BMC constructs, and AUC profiles of HA-functionalized protomers for BMC1 (**c**), BMC2 (**d**), BMC3 (**e**) and BMC4 (**f**). The calculated masses for each unlabelled protein are determined by summing the mass of the polypeptide sequence and the *c*-type haem (618 Da) covalently linked to the cytochrome.

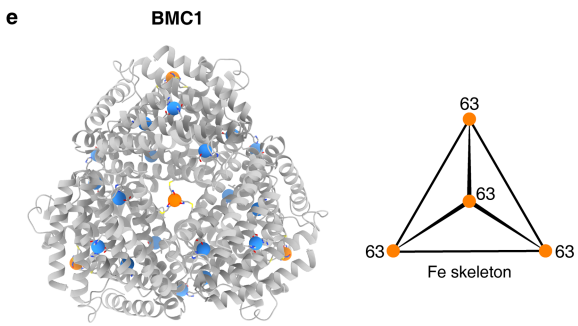


b

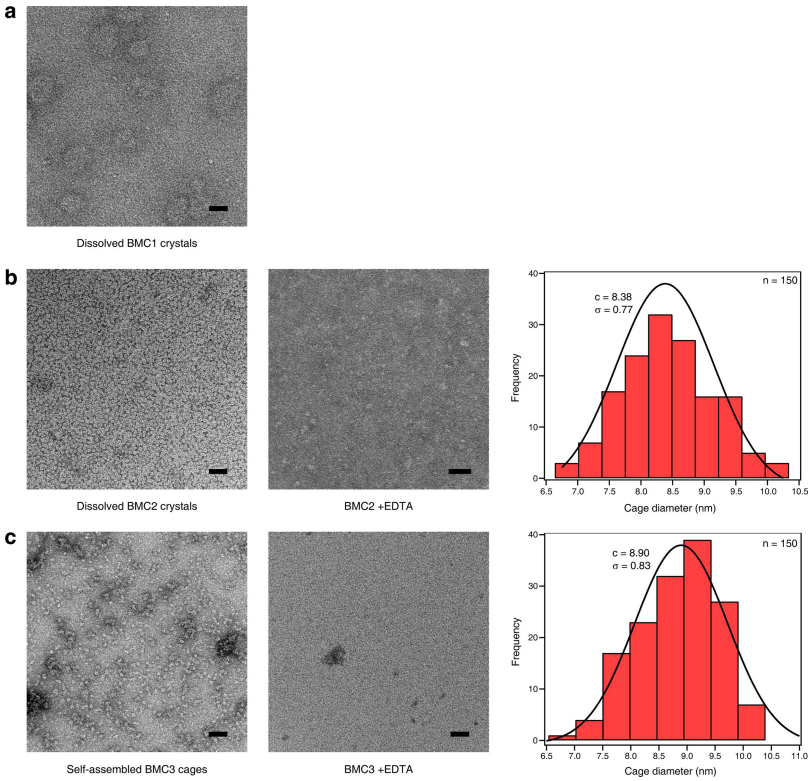
Cage variants	Zn-binding motifs	HA attachment sites
Generation 1		
BMC1	H8/D12	C63
BMC2	H8/D12	C63, C82
Generation 2		
BMC3	H5/H8/D12/H16	C63, C82
BMC4	H8/D12 E2/E5/H16/E25	C82



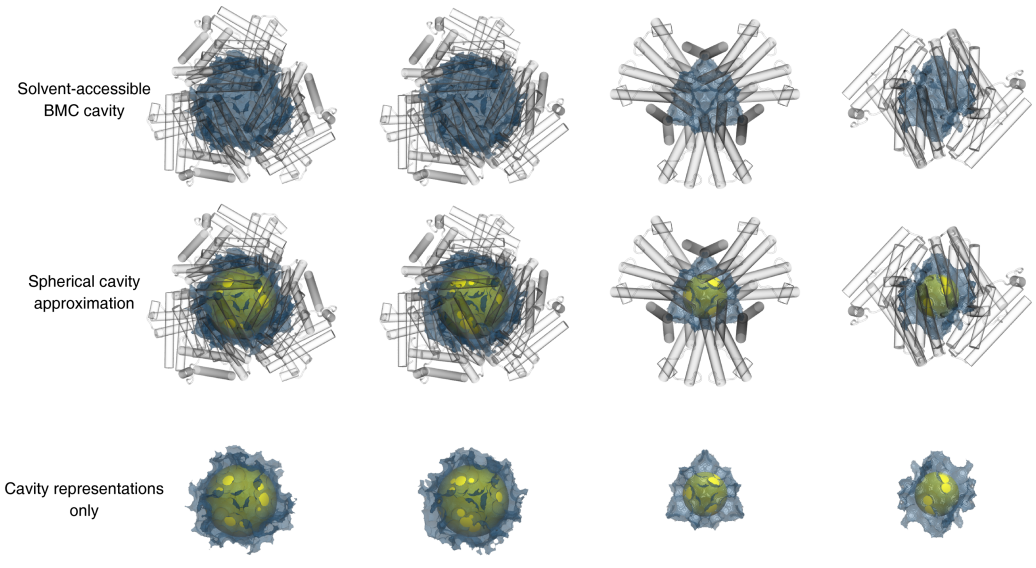




Extended Data Fig. 2 | Structural comparison of CFMC1 and BMC1 cages. **a**, The symmetric substructures of the CFMC1 dodecameric unit and its per-protomer SASA and BSA values. Associative surfaces on the protomers are coloured red for homologous interactions and red/orange or blue/cyan for heterologous interactions (right). **b**, Summary of engineered metal-coordination motifs for BMC constructs (see Supplementary Table 1 for all mutations). **c**, **d**, Comparison of C_2 and C_3 symmetric interfaces and corresponding metal binding sites for CFMC1 (**c**) and BMC1 (**d**). Full cages are shown as surfaces; insets show details of each interface. Fe and Zn ions are represented as orange and teal spheres, respectively. **e**, Cartoon representation of a full-size BMC1 cage with all metal ions shown as spheres. PDB ID: 3M4B (CFMC1), 6OT9 (BMC1).

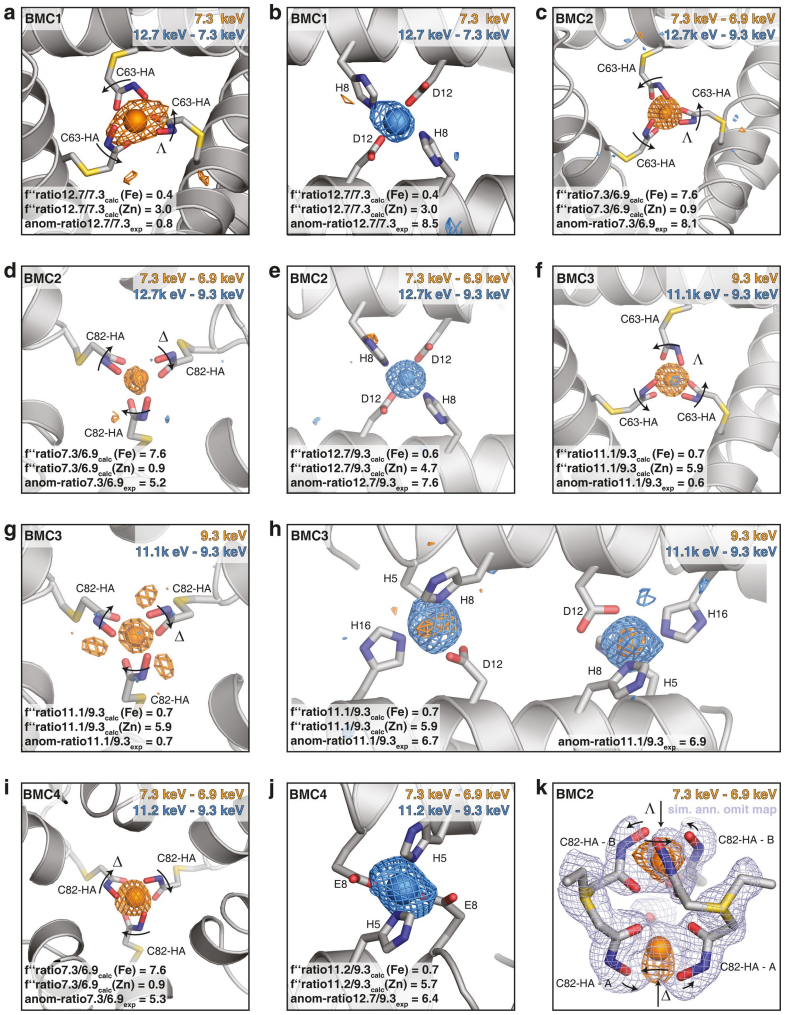


Extended Data Fig. 3 | ns-TEM characterization of BMC constructs. **a**, **b**, Dissolved Fe:Zn:BMC1 (**a**) and Fe:Zn:BMC2 crystals (**b**) in a buffer containing 100 mM HEPES (pH 7.5), 200 mM $MgCl_2$ and 800 μM $ZnCl_2$. **c**, Self-assembled Fe:Zn:BMC3 cages in a buffer containing 20 mM Tris (pH 8.5), 20 μM $FeSO_4$ and 60 μM $ZnCl_2$. Histograms in **b**, **c** reflect the size distributions of Fe:Zn:BMC2 and Fe:Zn:BMC3 cage diameters as measured from ns-TEM images. Gaussian fits to both distributions are drawn as solid lines along with their centres and standard deviations reported. Scale bars, 50 nm.

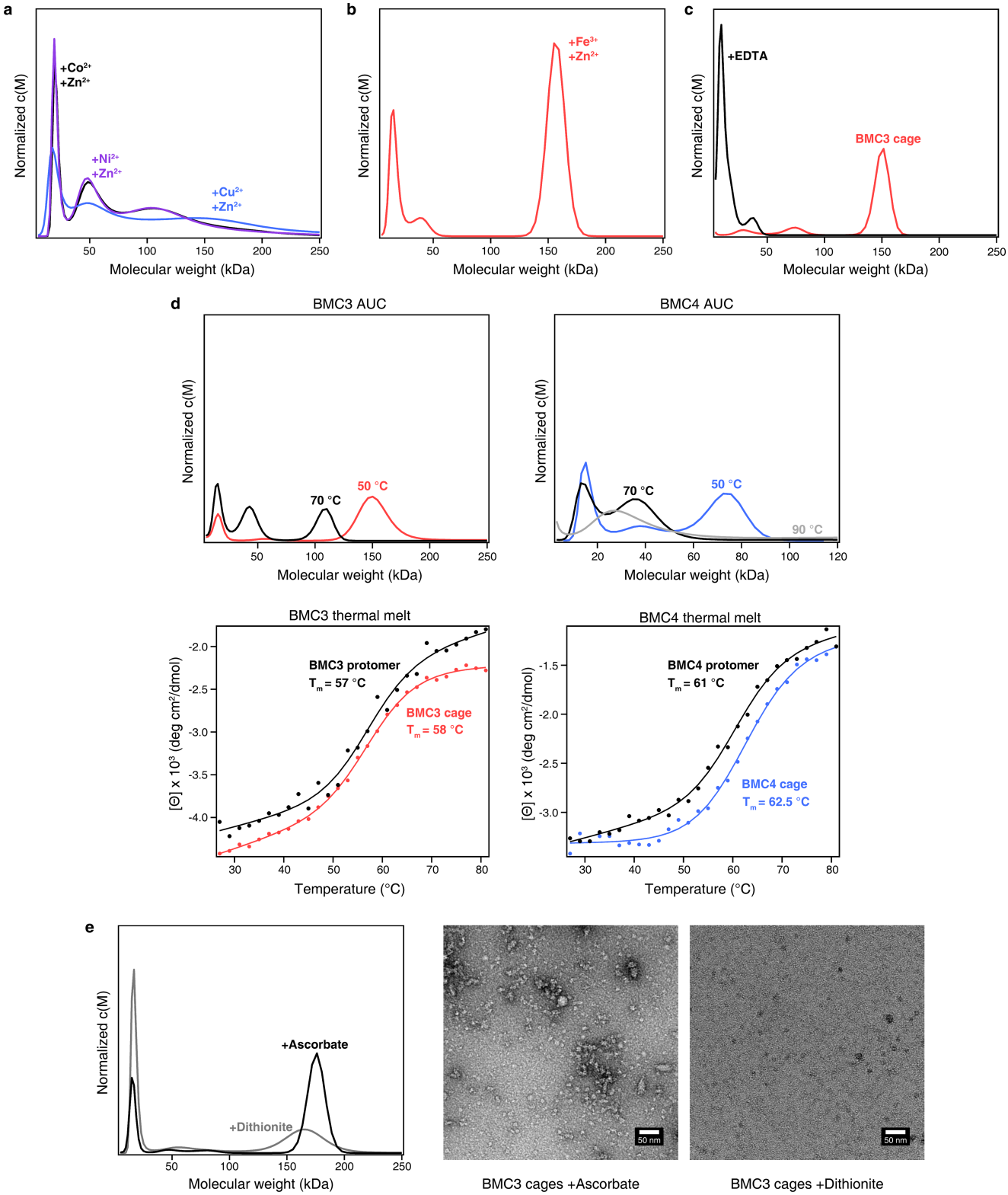


	BMC2	BMC3	BMC4 (top view)	BMC4 (side view)
Cavity volume (solvent-accessible)	32,700 Å^3	32,680 Å^3	7,880 Å^3	
Sphere volume (radius)	20,580 Å^3 (17 Å)	20,580 Å^3 (17 Å)	5,575 Å^3 (11 Å)	

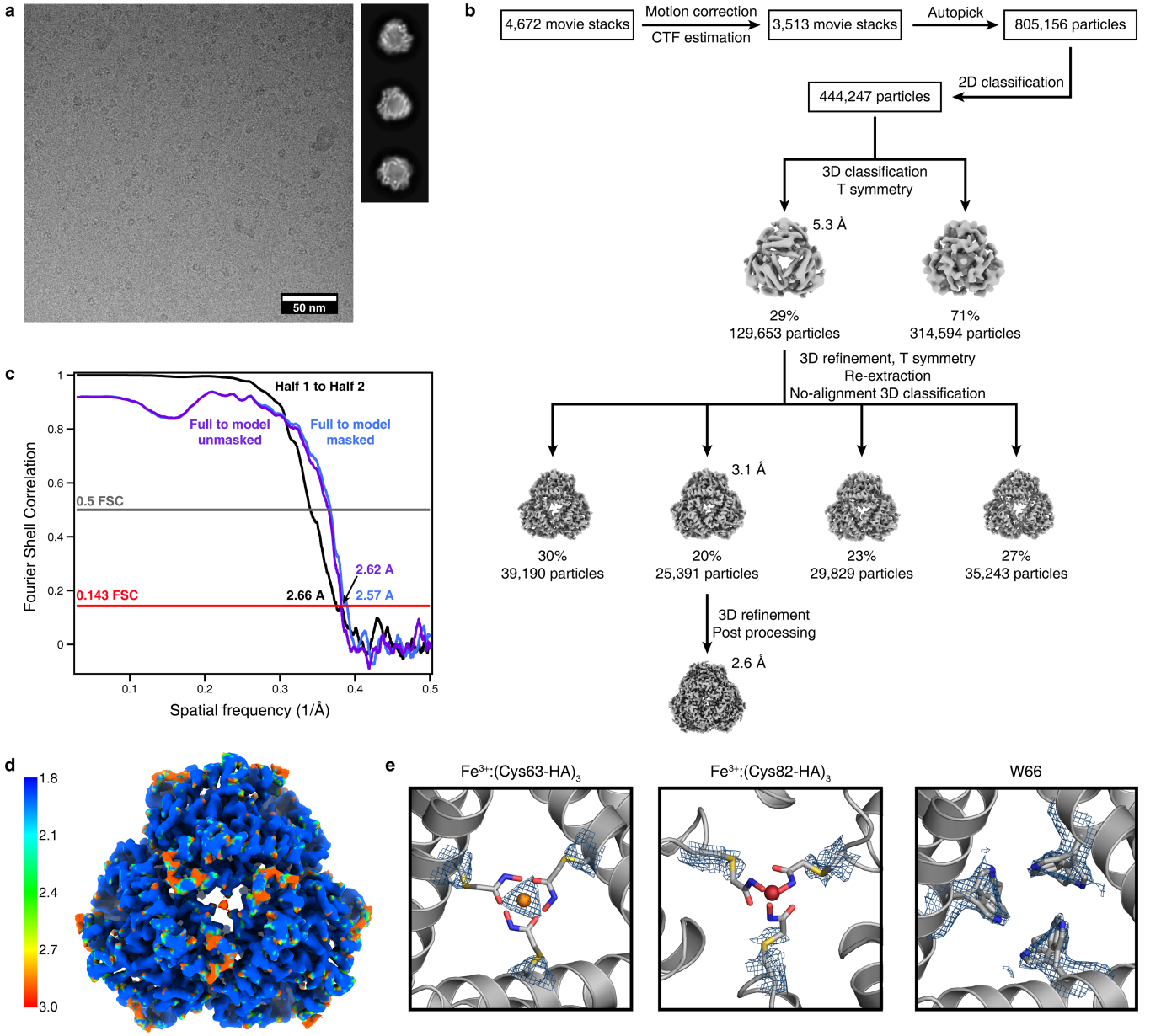
Extended Data Fig. 4 | Cavity volumes of BMC cages. Solvent-accessible cavity volumes within BMC cages as calculated by a 1.4 Å rolling probe are shown visually as blue meshes and reported numerically below. Spherical cavities, shown as yellow spheres in Figs. 2, 4, are reproduced for comparison to the calculated volumes. BMC proteins are represented as transparent cylinders.



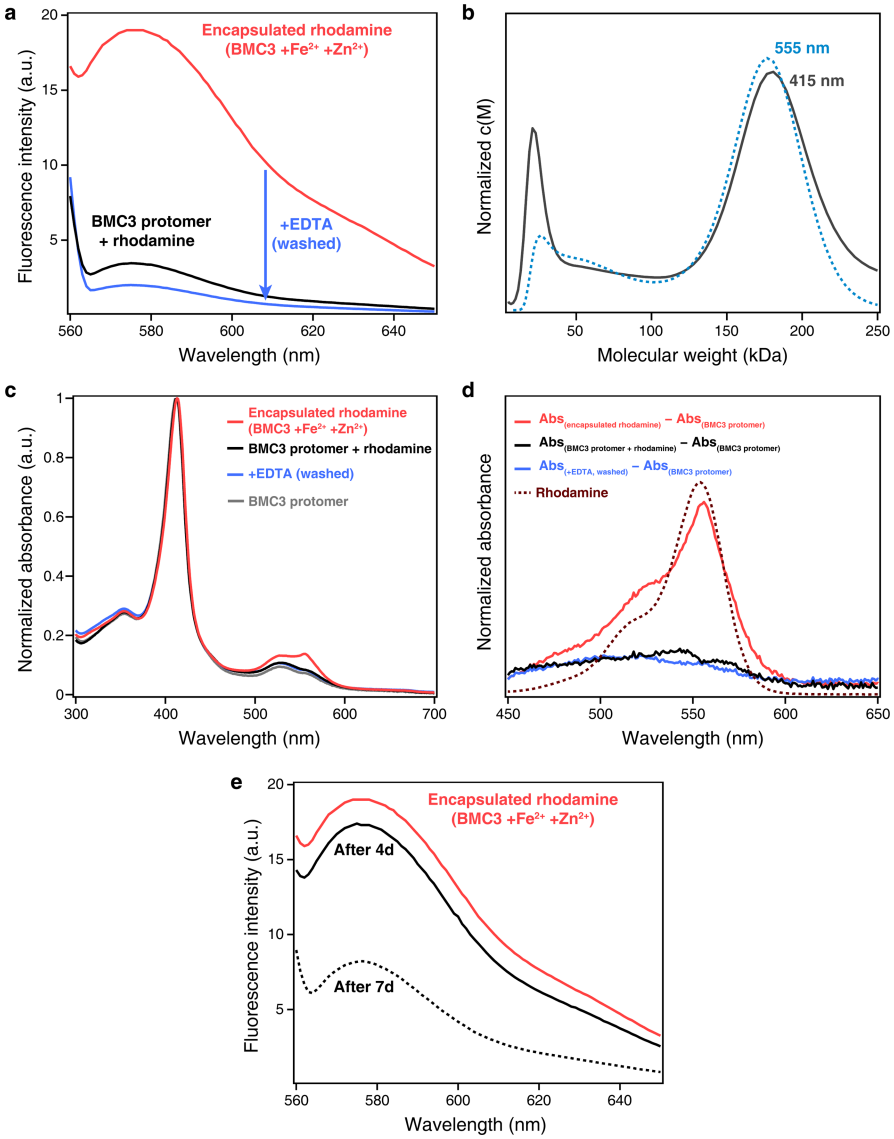
Extended Data Fig. 5 | Anomalous densities of engineered metal binding sites and conformational flexibility of Cys82–HA site. a–**j**, Cartoon and stick representations of the symmetric interfaces of BMC1 (**a**, **b**), BMC2 (**c**–**e**), BMC3 (**f**–**h**) and BMC4 (**i**, **j**) showing the engineered metal binding sites with the C63–HA ligands (**a**, **c**, **f**), C82–HA ligands (**d**, **g**, **i**) and Zn binding sites (**b**, **e**, **h**, **j**). The difference in the anomalous signal between pairs of datasets above and below the K-shell energy of Zn and Fe, respectively, are depicted as blue or orange meshes. A strong signal illustrates a strong change in anomalous signal across the respective edge, in turn suggesting the presence of the respective metal. The top right corner of each panel indicates the energies of the datasets used for the map of the respective colour. All anomalous difference maps are contoured at 3σ. As datasets around the Fe-edge were not available for BMC1 and BMC3 (necessitating calculations using anomalous difference density of singular datasets), the calculated f'' values for Zn at 7.3 and 9.3 keV are 0.82 and 0.52 (that is, non-zero) and thus some residual anomalous signal of the lower energy maps around the Zn atoms is expected to result even from strictly selective Zn loading. For a more quantitative analysis of the nature of the bound metal, ratios of the anomalous signal to the expected values (bottom left corner of each panel) were calculated as described in the Methods. **k**, Stick representation of the BMC2 Cys82–HA binding site in both alternative conformations with the anomalous difference density over the Fe-edge shown as orange mesh and a simulated annealing omit map (omitting all C82–HA atoms and Fe) of the normal electron density as light blue mesh contoured at 2σ. For all Cys–HA binding sites, arrows indicate the handedness of the binding site as Δ (right handed) or Λ (left handed). The reversion of handedness in **k** with the respective view angle is indicated by arrows. Colour code for atoms in all panels: Fe in orange, Zn in blue, S in yellow, O in red and N in dark blue.



Extended Data Fig. 6 | Solution characterization of self-assembled BMC3 and BMC4 cages. a–**c**, The oligomerization state of BMC3 cages as monitored by AUC measurements following incubation with various first-row transition metal ions (**a**), incubation with Zn^{2+} and Fe^{3+} ($Fe(acetylacetonate)_3$) (**b**) and disassembly via sequestration of metal ions by EDTA (**c**). **d**, AUC profiles of BMC variants after equilibration for 2 h at the indicated temperatures (top). Thermal unfolding of BMC variants as measured by circular dichroism spectroscopy at 222 nm (bottom). **e**, AUC profiles of BMC following treatment with chemical reductants of different reduction potentials (left). ns-TEM micrographs (middle and right) are shown for cage samples incubated with the corresponding chemical reductants.



Extended Data Fig. 7 | Cryo-EM analysis of BMC3 cages. a, Representative cryo-EM micrograph and 2D class averages. **b**, Flowchart detailing image processing from collected movie stacks to final map. Additional details can be found in the Methods. **c**, FSC curves calculated between the half-maps (black line), atomic model to the unmasked full map (purple line) and atomic model to the masked full map (blue line). Resolution values are indicated at the gold-standard FSC 0.143 criterion. **d**, Local resolution estimates of the final reconstruction calculated using ResMap. **e**, Electron density shown at BMC3 C_3 interfaces highlighting poorly resolved density (reflecting high flexibility) at hydroxamate sites and multiple conformations of W66.



Extended Data Fig. 8 | Encapsulation of rhodamine inside BMC3 cages. **a**, Fluorescence characterization of BMC3 samples incubated with rhodamine. Cages encapsulating rhodamine were treated with EDTA and washed before measuring fluorescence intensity. **b**, AUC profiles of cages encapsulating rhodamine monitored at the haem Soret absorption maximum (λ_{max} = 415 nm) and rhodamine absorption maximum (λ_{max} = 555 nm). **c**, UV-vis characterization of BMC3 samples incubated with rhodamine. **d**, Difference spectra of BMC3 samples and BMC3 protomer shown in **c**. Free rhodamine dissolved in buffer is shown as dark-red dashes. **e**, Repeated fluorescence characterization of a solution containing BMC3 cages encapsulating rhodamine over several days. The sample was washed three times before each fluorescence measurement.

Extended Data Table 1 | X-ray data collection, processing and refinement statistics

	BMC1	BMC2	BMC3	BMC4
Data collection				
Space group	R 3 2	R 3 2	R 3 2	P 6_3 2 2
Cell dimensions				
a, b, c (Å)	125.6, 125.6, 166.4	126.1, 126.1, 168.2	126.7, 126.7, 167.8	87, 87, 63.3
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	39.92 – 2.40 (2.46 – 2.40)	39.25 – 1.40 (1.44 – 1.40)	91.83 – 1.85 (1.90 – 1.85)	48.47 – 1.50 (1.54 – 1.50)
No. Reflections Observed	393414 (28956)	1863896 (88779)	885564 (66924)	734764 (18399)
No. Reflections Unique	38245 (2875)	195424 (14465)	85783 (6342)	37652 (1334)
R_{merge}	0.185 (2.986)	0.056 (1.830)	0.096 (2.377)	0.053 (1.936)
$I / \sigma I$	8.1 (0.96)	20.3 (0.84)	13.7 (1.04)	26.6 (1.23)
CC1/2	0.998 (0.501)	1.000 (0.410)	0.999 (0.468)	1.000 (0.480)
Completeness (%)	99.81 (100.00)	99.37 (98.9)	100.00 (100.00)	87.66 (42.2)
Redundancy	10.29 (10.01)	9.54 (6.138)	10.32 (10.54)	19.51 (13.94)
Wilson B $(Å)^2$	52	19	36	29
Refinement				
Resolution (Å)	36.37 – 2.40 (2.43 – 2.40)	31.54 - 1.40 (1.42 - 1.40)	91.83 - 1.85 (1.88 - 1.85)	48.47 - 1.50 (1.53 - 1.50)
No. reflections	38194 (1286)	195270 (8377)	85755 (3974)	37642 (810)
R_{work}/R_{free}	0.2174/0.2718	0.1659/0.1909	0.1826/0.2106	0.1900/0.2181
No. atoms				
Protein	3292	3502	3459	925
Ligand/ion	200	244	240	58
Water	19	786	342	124
B-factors $(Å)^2$				
Protein	70	24	41	42
Ligand/ion	68	22	43	35
Water	63	37	46	49
R.m.s. deviations				
Bond lengths (Å)	0.010	0.011	0.009	0.013
Bond angles (°)	1.29	1.33	1.15	1.33
Clashscore	7	5	7	10
Ramachandran favored (%)	100	100	100	97
Ramachandran allowed (%)	0	0	0	3
Ramachandran outliers (%)	0	0	0	0
Rotamer outliers (%)	1	2	2	1

Numbers in parentheses correspond to the highest-resolution shell.

Extended Data Table 2 | Cryo-EM data collection, processing, and refinement statistics

	BMC3 (EMDB ID: EMD-20212) (PDB ID: 6OVH)
Data collection and processing	
Magnification	165,000x
Voltage (kV)	300
Electron dose (e^-/Å^2)	60
Exposure rate (e^-/ Å^2/s)	6
Defocus range (μm)	0.84
Pixel size (Å)	0.3 – 2.7
Symmetry imposed	T
Total extracted particles (no.)	805,156
Final refined particles (no.)	25,391
Map resolution (Å)	2.57
FSC 0.143 (unmasked/masked)	2.62/2.57
Map resolution range (Å)	∞– 2.57
Applied B-factor (Å^2)	-79
Refinement	
Initial model used (PDB code)	6OT7
Model resolution (Å)	2.57
FSC 0.5 (unmasked/masked)	2.74/2.72
FSC 0.143 (unmasked/masked)	2.62/2.57
Model resolution range (Å)	∞– 2.57
Map sharpening *B* factor (Å^2)	-79
Model composition	
Non-hydrogen atoms	10895
Protein residues	1272
Ligand/ion	68
Water	171
B-factors (Å)^2	
Protein	51
Ligand/ion	59
Water	48
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.876
Validation	
MolProbity score	1.33
Clashscore	1.15
Rotamer outliers (%)	4.55
Ramachandran plot	
Favored (%)	99.04
Allowed (%)	0.96
Disallowed (%)	0.00

nature research

Corresponding author(s): F. Akif Tezcan

Last updated by author(s): Nov 8, 2019

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 Authors & Referees 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	X-ray crystallography: Custom software at SSRL and ALS Electron Microscopy: EPU (FEI)
Data analysis	X-ray crystallography: XDS ver. June 1, 2017 (BUILT=20170615) and XSCALE ver. Jun 17, 2015 (BUILT=20150617) (data integration/ scaling), Phaser-MR (molecular replacement), Phenix v.1.13-2998 (model building/refinement), Coot v.0.8.6.1 (visualization, real time refinement), Pymol version 1.3 (molecular graphics), ChimeraX v.0.9 (molecular visualization), VMD v.1.9.3 (molecular visualization), Voidoo v.3.3.4 (cavity measurements), Mapman v.7.8.5(manipulation and analysis of electron-density map), Fiji (image processing), MotionCor2 v.1.2.1 (motion correction), Gctf v.1.06 (defocus value estimation), Relion v.3.0 (particle picking, classification, refinement), ResMap v.1.1.4

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/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

Provide your data availability statement here.

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	No sample size calculation was performed. The various sample sizes (n) for the different experiments was selected as no significant difference was observed between technical replicates
Data exclusions	No data was excluded
Replication	All results from the experiments were successfully replicated.
Randomization	No randomization was performed as it is not relevant to the current study.
Blinding	No blinding was performed as it is not relevant to the current study.

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 and 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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

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

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.

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.*

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.