Manuscript number: NP-P200203A

Corresponding author and email address: F. Akif Tezcan, tezcan@ucsd.edu

Number of:

Figs: 8

Tables: 1

Boxes: 0

List of SI:

Supplementary Figure 1 | 1 H NMR spectrum of O-tritylhydroxylamine. (400 MHz, DMSO-d6)
Supplementary Figure 2 | 1 H NMR spectrum of 2-chloro-N-hydroxyacetamide. (300 MHz, DMSO-d6)
Supplementary Figure 3 | 13C NMR spectrum of 2-chloro-N-hydroxyacetamide. (500 MHz, DMSO-d6)
Supplementary Figure 4 | 1 H NMR spectrum of 2-iodo-N-hydroxyacetamide. (300 MHz, DMSO-d6)
Supplementary Figure 5 | 13C NMR spectrum of 2-iodo-N-hydroxyacetamide. (500 MHz, DMSO-d6)
Supplementary Figure 6 | 1 H NMR spectrum of N1 ,N4 -dihydroxyterephthalamide. (500 MHz, DMSO-d6)

Supplementary Figure 7 | 13C NMR spectrum of N1 ,N4 -dihydroxyterephthalamide. (500 MHz, DMSO-d6)

Supplementary Figure 8 | 1 H NMR spectrum of N2 ,N3 -dihydroxyterephthalamide. (500 MHz, DMSO-d6)

Supplementary Figure 9 | 13C NMR spectrum of N2 ,N3 -dihydroxyterephthalamide. (500 MHz, DMSO-d6)

CFIs: N/A

EDITORIAL SUMMARY

To design new multiprotein systems, Tezcan and coworkers describe how to combine natural metal-coordinating motifs and hydroxamic acid groups to direct metal-mediated assembly of polyhedral protein architectures and 3D crystalline protein frameworks.

TWEET

@TezcanLab

DATA AVAILABILITY

The authors declare that all the data supporting the findings of this study are available within the article. All the data analysis was performed using published tools and packages and has been provided with the paper.

RELATED LINKS

Key reference(s) using this protocol

[These are primary research papers where the protocol has been used. Limit of 5. The reference is an example of the format used for this citation.]

Sontz, P.A., Bailey, J.B. et. al. J. Am. Chem. Soc. 137 (36), 11598-11601 (2015). https://pubs.acs.org/doi/10.1021/jacs.5b07463

Bailey, J.B. et. al. J. Am. Chem. Soc. 139 (24), 8160-8166 (2017). https://pubs.acs.org/doi/10.1021/jacs.7b01202

Golub, E. et. al. Nature. 578, 172-176 (2020). https://www.nature.com/articles/s41586-019-1928-2

Bailey, J.B. et. al. J. Am. Chem. Soc. 142 (41), 17265-17270 (2020). https://pubs.acs.org/doi/10.1021/jacs.0c07835

Design of Metal-Mediated Protein Assemblies via Hydroxamic Acid Functionalities

Rohit H. Subramanian[1], Jie Zhu[1], Jake B. Bailey[1], Jerika A. Chiong[1], Yiying Li[1], Eyal Golub[1], F. Akif Tezcan[1, 2*]

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA

[2]Materials Science and Engineering, University of California, San Diego, La Jolla, CA 92093, USA

*e-mail: tezcan@ucsd.edu

Abstract

The self-assembly of proteins into sophisticated multicomponent assemblies is a hallmark of all living systems and has spawned extensive efforts in the construction of novel synthetic protein architectures with emergent functional properties. Protein assemblies in nature are formed via selective association of multiple protein surfaces through intricate noncovalent protein-protein interactions, a challenging task to accurately replicate in the *de novo* design of multiprotein systems. In this protocol, we describe the application of metal-coordinating hydroxamate (HA) motifs to direct the metal-mediated assembly of polyhedral protein architectures and 3D crystalline protein frameworks (protein-MOFs). This strategy has been implemented using an asymmetric cytochrome cb_{562} monomer through selective, concurrent association of Fe^{3+} and Zn^{2+} ions to form polyhedral cages. Furthermore, the use of ditopic HA linkers as bridging ligands with metal-binding protein nodes has allowed the construction of crystalline 3D protein-MOF lattices. The protocol is divided into two major sections: (1) the development of a Cys-reactive HA molecule

for protein derivatization and self-assembly of protein-HA conjugates into polyhedral cages and (2) the synthesis of ditopic HA bridging ligands for the construction of ferritin-based protein-MOFs using symmetric metal-binding protein nodes. Protein cages can be analyzed using analytical ultracentrifugation (AUC), transmission electron microscopy (TEM) and single-crystal X-ray diffraction (sc-XRD) techniques. HA-mediated protein-MOFs are formed in sitting-drop vapor diffusion crystallization trays and are probed via sc-XRD and multi-crystal small-angle X-ray scattering (SAXS) measurements. Ligand synthesis, construction of HA-mediated assemblies, and post-assembly analysis as described in this protocol can be performed by a graduate-level researcher within six weeks.

Introduction

The self-assembly of proteins into higher-order structures is a cornerstone of all cellular functions.[1] Biological processes as diverse as the conversion of light into chemical energy in photosynthesis[2] or the packaging of DNA into nucleosomes[3] require large, multicomponent protein architectures and extended arrays. Given the sophistication of such natural protein assemblies and their central roles in biology, a fundamental goal in biomolecular engineering has been the development of new design tools and strategies for the construction of artificial protein assemblies, which possess structural and functional properties that match or even surpass those produced by natural evolution[4-7].

The simple composition of polynucleotides from four building blocks coupled with the high specificity of Watson-Crick base pairing has enabled the programmable assembly of DNA or RNA into virtually any nanoscale architecture[8]. By contrast, the complex chemical composition and 3D structures of proteins pose an enormous challenge in terms of predictably constructing

desired multiprotein arrays and architectures. Natural protein assemblies are built through the selective association of protein monomers (protomers). Predominantly, the contact points are multiple, extensive patches of each protein surface (surface patches) held together by heterogeneous, noncovalent protein-protein interactions [1]. In light of the difficulty of designing (or evolving) such associative patches on protein surfaces from scratch, a powerful strategy exploited both by nature and protein designers has been to create new structures by the symmetric arrangement of protein components (symmetrization)[9]. Applying symmetry principles enables the engineering of fewer associative surface patches to generate sophisticated multimeric assemblies; these principles are used to develop geometric design rules to generate discrete protein oligomers or periodic/crystalline protein arrays with predictable structures.

The first step, however, still involves designing stable and extensive protein-protein interactions. There are many approaches to address this challenge; our group has previously shown that the strength, reversibility, and directionality of metal-coordination interactions could be used to bypass the necessity of designing large, noncovalent protein interfaces while also imposing symmetry[7]. These advantages in turn have enabled the construction of many protein assemblies with unique structural, functional and dynamic properties[10-14].

In this Protocol, we describe the development and applications of the versatile hydroxamic acid (HA) functionality, which is a bidentate chelate that is capable of binding many metal ions with high affinity and is exploited in bacterial siderophores for selective Fe^{3+} capture[18,19] (**Fig. 1a**). In particular, we focus on two classes of HA-based reagents and synthetic linkers, which have enabled the construction of 1) cage-like, polyhedral protein assemblies with unique structural and stimuli-responsive properties[10] (**Fig. 1b**), and 2) a series of chemically-designed, crystalline 3D protein networks (Protein-Metal-Organic Frameworks or protein-MOFs) with tunable symmetries

and unit cell metrics[11,20,21] (**Fig. 1c**). Both types of protein assemblies are distinguished from other artificial protein architectures and arrays by their ease of design, modularity, reversible formation and dynamic features.

Development of the Protocol

Given the challenges of *de novo* protein design, many construction strategies have relied on linking natively oligomeric proteins via binary protein-protein interactions to form multidimensional assemblies[5]. One approach has been to create genetic fusions of natively oligomeric proteins to position proteins into higher-order structures and promote *in vitro* and *in vivo* assembly without further manipulation (**Fig. 2a**). Early reports from Yeates and coworkers implemented the fusion of two symmetric components to generate polyhedral protein cages and 1D filaments[22]. Natively dimeric and trimeric proteins were covalently tethered using alpha-helical linkers at different orientations to afford 0D cage-like assemblies and bundles of 1D protein filaments, serving to validate genetic fusion of symmetric components as a viable protein design strategy. Further efforts, using both peptide and protein components as structural nodes, have produced tetrahedral, octahedral, and icosahedral protein cages[23-26] as well as 2D crystalline arrays[27]. These studies couple symmetric elements, through a rigid or flexible linker, to afford modular control of protein assemblies. In the meantime, significant advances in computational design have enabled the creation of tight, associative interfaces (consisting of electrostatic and/or hydrophobic interactions similar to those present in biological assemblies) between symmetric proteins to form megadalton-scale protein cages[28,29] and extended 2D assemblies[30] (**Fig. 2b**). As an alternative approach, the introduction of directional bonding interactions (*e.g.*, disulfide bond

formation[31,32] between Cys residues and metal coordination[12,33,34]) between pairs of symmetric modules has been used to generate robust 0-, 1-, 2-, and 3D protein assemblies (**Fig. 2c, d**). From this it is clear that there are several methods to generate protein oligomers by installing a C_2 symmetric or binary protein-protein interaction. However, achieving symmetric association via higher order symmetries (*e.g.*, introducing C_3 symmetric nodes) has been relatively unexplored.

Higher-order symmetry achieved using metal coordination

One approach that could facilitate the introduction of multiple symmetric elements with specificity is metal coordination. Metal ions perform vital functions in biological systems[35] (*e.g.*, as catalytic centers, cofactors, and structural anchors for protein folds) and indeed, metal binding provides many desirable properties for protein design, including strong and directional bonding, chemical tunability and reversibility (*e.g.*, by pH, metal chelators, and redox potential).

With these advantages in mind, our group developed a strategy termed Metal-Directed Protein Self-Assembly (MDPSA), whereby metal-coordination motifs are incorporated onto protein surfaces to promote oligomerization upon the addition of late-first-row transition metal ions (*e.g.*, Ni^{2+}, Cu^{2+}, and Zn^{2+}). Such metal-mediated assemblies have largely relied on the surface installment of natural metal-coordinating amino acid functionalities like histidine (His), cysteine (Cys), glutamic acid (Glu) and aspartic acid (Asp). This method was first implemented on a monomeric four-helix bundle protein, cytochrome cb_{562}, by installing a pair of bis-His "clamps" at *i* and *i+4* positions on an α-helix to promote oligomerization upon the introduction of Zn^{2+} ions[38]. Further experiments conducted with other first-row transition metals revealed that the coordination preference of the metal ion could directly influence the oligomerization state and symmetry of the protein scaffold (*e.g.*, square planar Cu^{2+} binding yielded C_2 symmetric protein

dimers and octahedral Ni^{2+} binding produced C_3 symmetric trimers)[39]. Optimization of this strategy, through the introduction of associative interfaces via computational design[36] and additional chemical bonding via disulfide formation[40], enabled the creation of *in vivo* assembling oligomers[41], infinite 1D helical nanotubes and 2D crystalline arrays[12,42], hydrolytic enzymes through the introduction of distinct structural and catalytic Zn^{2+} sites[13,43], and allosteric assemblies via strained intermolecular disulfide bonding coupled to Zn^{2+} binding[14,44]. These results demonstrated that a diverse set of protein oligomers can be obtained from a single, monomeric protein building block through the judicious incorporation of metal coordinating residues.

More complex architectures require additional metal-binding sites

In studies germane to this Protocol, we set out to construct cage-like, polyhedral protein assemblies and crystalline, 3D protein arrays using MDPSA. Cage-like architectures have been particularly attractive targets for protein design due to their highly symmetric structures as well as their potential uses in encapsulation, delivery and biocatalysis[45,46]. Similarly, the ability to rationally design 3D protein crystals would not only expand their ever-growing applications as porous materials for catalysis[47] and encapsulation[48], but it also constitutes an important goal in terms of X-ray protein crystallography, where obtaining protein crystals is generally a rate-limiting step[49]. However, the construction of both cage-like protein assemblies and 3D crystals is a considerably more complex task for MDPSA (compared to simple metal-mediated oligomers) due to the necessity to impose multiple symmetry elements simultaneously. Indeed, an examination of naturally occurring protein cages (*e.g.*, virus capsids, ferritin) reveals that they are invariably composed of asymmetric protomers that present multiple associative interfaces to satisfy the symmetry requirements necessary to build polyhedral structures. For example, a tetrahedral complex must at least possess C_2 and C_3 symmetric interfaces, whereas octahedral or icosahedral

architectures additionally feature C_4 or C_5 symmetries[1,4]. Furthermore, natural protein cages often display dynamic behavior or reversible assembly/disassembly as necessitated by their biological functions, meaning that their protein-protein interfaces must also be responsive to external stimuli[50,51].

To further broaden the structural and functional scope of such metal-directed protein assemblies, we and others have endeavoured to employ non-natural, metal-chelating functionalities to mediate protein-protein interactions[15-17]. In order to satisfy the stringent design criteria for cage-like protein assemblies and 3D crystals (*i.e.*, simultaneous generation of multiple, reversible protein-protein interfaces that impose different symmetries for self-assembly), we developed an alternative MDPSA strategy, which takes advantage of a fundamental concept in inorganic chemistry, namely the Hard-Soft Acid-Base (HSAB) theory[52]. Natural metal coordinating amino acids, such as His, Asp, Glu, or Cys residues, can be considered as soft or intermediate-soft bases according to the HSAB classification and have considerable overlap in terms of their coordination preferences for soft, low-valent transition metal ions such as Ni^{2+}, Cu^{2+}, and Zn^{2+}. Due to this overlap, it is essentially impossible to design a protein building block for MDPSA such that it can selectively coordinate two different soft metal ions on its surface based solely on natural amino acids.

Hydroxamic acid enables selective metal coordination

Therefore, we surmised that if a hard, metal chelating motif could be introduced onto the protein surface, it could work in concert with a soft metal-binding motif composed of natural amino acids to assemble into a complex architecture through the coordination of two different metal ions. Hydroxamic acid (HA), a bidentate chelating motif capable of binding many metal ions with high

affinity, is present naturally in bacterial siderophores where it is exploited for selective Fe^{3+} capture[18,19] (**Fig. 1a**). HA is a hard ligand that forms highly stable octahedral Fe^{3+}:$(HA)_3$ complexes with high specificity and affords C_3 symmetry. To implement our strategy, we synthesized a Cys-reactive HA reagent (2-iodo-*N*-hydroxyacetamide or IHA) and incorporated it onto the monomeric cytochrome cb_{562} scaffold, which was also tailored with native metal binding residues to enable C_2-symmetric metal coordination. The resulting cytochrome cb_{562} variants were observed to self-assemble into tetrahedral (dodecameric) or trigonal bipyramidal (hexameric) protein cages through concurrent Fe^{3+} and Zn^{2+} coordination[10]. Importantly, these tightly packed cages were capable of reversible assembly/disassembly due to their metal-dependent construction. For protein derivatization, we chose iodo-functionalization in lieu of commonly used maleimide or thiopyridine functional groups. This allowed us to minimize the number of bonds between the Cys reactive group and the HA motif, generating a pseudo-amino acid with a side chain isosteric with that of arginine. Additionally, maleimides have been shown to undergo undesired hydrolysis[53] and thiopyridine modification of Cys is a reversible, redox-sensitive process (which may potentially interfere with reversible redox-mediated assembly and disassembly of protein cages).

Ditopic HA linkers form bridges between proteins

In parallel, we used the HA motif to develop synthetic, ditopic linkers, which served as C_2-symmetric bridges to promote the formation of 3D protein lattices. In that case, rather than constructing a cage-like protein assembly, we took advantage of an already-existing 24meric protein cage (human heavy-chain ferritin or HuHF)[54] as a symmetric building block. HuHF was first engineered on surface locations with tris-His metal coordinating groups to create octahedral, metal-coordinating nodes. Upon addition of HA-based linkers, the HuHF nodes self-assembled

into the desired body-centered protein lattices with synthetically programmable unit cell parameters[11,20,21]. Continued pursuit of fundamental studies to further understand the effects of protein node symmetry, linker symmetry and metal ion identity will enable the generation of designer 3D protein materials towards molecular capture and information storage applications.

Overview of the procedure

In this Protocol, we discuss the incorporation of the HA group as a tool to direct protein self-assembly in two different modalities: (1) through direct conjugation of IHA onto proteins to site-specifically direct metal coordination for the formation of protein polyhedra, and (2) through the use of ditopic, HA-based ligands that act as metal-chelating bridges between proteins to construct 3D protein lattices.

HA-mediated protein cages

An overview of the procedure and timeline from the synthesis of IHA to the formation of protein cages is shown in **Fig. 3**. The major steps involved are: (1) chemical synthesis of an IHA molecule for labeling surface exposed Cys residues, (2) IHA-protein conjugation and subsequent purification, (3) incubation with metal ions under anaerobic conditions to promote self-assembly, and (4) assessment of self-assembly products.

A crucial component for successful self-assembly lies in choosing the correct protein partner to the HA motif and careful consideration of a surface-exposed site for HA conjugation. For the generation of discrete protein polyhedra, different surface positions of the Cys residue can be tested to assess their effects on protein self-assembly upon conjugation to HA and incubation with metal ions. It is also important to consider a protein scaffold that can accommodate multiple

metal-binding sites, should they be required to form a solution-stable oligomer. For instance, the cytochrome cb_{562} proteins used in our initial study were capable of housing native metal binding residues (His, Asp, and Glu) as well as Cys-HA motifs. Concurrent binding at both metal coordinating sites was necessary to form solution-stable protein cages[10]. Furthermore, the presence of multiple HA motifs can affect the assembly outcomes; the formation of dodecameric (required two HA motifs) vs. hexameric (required one HA motif) cages was, in part, determined by the number of surface HA sites on our protein scaffolds.

HA-mediated protein-MOFs

The generation of protein-MOFs is accomplished by combining ditopic HA-bearing linkers with metal-binding protein nodes, as shown in the overview in **Fig. 4**. The procedure involves (1) the synthesis of ditopic HA bridging linkers and (2) preparation of a protein node to effectively coordinate transition metal ions, which can be combined to form μm-scale crystalline lattices, and (3) using X-ray diffraction and scattering techniques for the characterization of protein-MOFs.

To facilitate the formation of 3D networks, the proteins contain tripodal metal coordination motifs that can tightly bind transition metal ions in solution while simultaneously offering a surface exposed open coordination site for binding to HAs. The construction of ferritin-MOFs was enabled using a metal-coordinating HuHF variant, generating an octahedral metal-protein node, that coordinated with bridging HA linkers to form ligand-mediated crystalline 3D networks. Despite a marked (nearly 10-fold in the longest dimension) size difference between the organic HA linkers and the ferritin node, the protein-MOF lattices are robustly interconnected solely by metal-HA interactions.

Applications of the method

Applications of protein-HA conjugation to generate protein cages

The straightforward synthesis and simple protein labeling methods used to covalently conjugate HA onto a protein scaffold present a powerful strategy for generating a selective metal coordination motif to induce protein trimerization. This approach can be further extended using symmetric building blocks (natively occurring or a product of *de novo* design) to generate different types of polyhedra (*e.g.*, octahedral and icosahedral cages) or extended 1D and 2D structures depending on the oligomerization state of the building block and the positioning of the HA motif. A single Cys residue, placed appropriately on the protein surface, can complement computational design, secondary metal coordination motifs, genetic fusion, or any other design strategies to provide structural and functional diversity in the construction of sophisticated protein assemblies for potential uses in the selective capture and release of cargo for drug delivery or therapeutic needs and providing confined cavities for improved catalytic activity[55-58].

The HA motif can be replaced with other metal chelators to diversify the metal coordination motifs used to drive protein self-assembly. Non-native metal-binding motifs[15,16] (*e.g.*, bipyridine, terpyridine, 1,10-phenanthroline, and 8-hydroxyquinoline), in addition to minimally explored siderophore-inspired metal-coordinating functional groups[59] (catechols or phenolates), can be used for protein derivatization and cage formation in a similar manner to the HA motif described in this Protocol. One advantage of exploring different metal binding groups is the ability to probe the effect of bidentate vs. tridentate ligands (*e.g.*, bipyridine vs. terpyridine) on self-assembly products. Depending on the positioning of these ligands and the choice of protein scaffold, it may be possible to achieve different cage symmetries by altering the ligand coordination or varying the order of metal ion addition when using a bimetallic scaffold.

Additionally, such chelates often give rise to metal complexes with strong electronic absorption[60,61] or luminescence properties[62], meaning that the formation of protein assemblies can be readily monitored. Diversification of the metal coordination motifs can also be achieved through the incorporation of unnatural amino acids (UAA) to enable *in vivo* formation of metal-driven protein cages. Existing UAAs (*e.g.*, BpyAla[63] and HQ-Ala[64]) can be incorporated onto self-assembling cytochrome cb_{562} protein scaffolds while parallel studies to generate a HA-bearing UAA can be performed to readily assemble protein cages *in vivo*.

Applications of linker-mediated 3D protein-MOFs

The use of ditopic HA bridging linkers to bridge protein nodes into ordered lattices can be implemented as a strategy for the ligand-mediated crystallization of symmetric building blocks by leveraging the strong metal coordination interactions that drive protein-MOF assembly. The versatility of protein-MOF construction can be expanded by increasing the scope of the protein building blocks and ditopic bridging ligands. Systematic modulation of protein-MOF components has already been shown to alter crystal behavior (*e.g.*, improved thermal stability in Ni^{2+}- vs. Zn^{2+}-ferritin-MOFs)[20] and further exploration into new proteins and ligands may yield unique bulk materials properties resulting from the underlying molecular arrangements. Investigating other proteins bearing alternative symmetries (*e.g.*, *T* or *I* symmetry) would alter the lattice patterning of the resultant protein-MOFs. In addition, using HA bridging ligands bearing functional molecules can impact the dynamic behavior of the resultant crystalline scaffolds. Incorporating new moieties (*e.g.*, fluorescent dyes[65], light-responsive azobenzenes[66], and large coiled-coil peptide or DNA biomolecules) onto the HA ligand scaffold will enable the formation of dynamic

frameworks with chemical tunability and functional versatility and serve to advance the design and construction of a new class of crystalline 3D frameworks.

Comparison with other methods

Construction of protein cages

Genetic fusion of symmetric proteins, or peptides, has proven to be an effective strategy for generating uniform protein cages[24-26]. In this strategy, a pair of oligomeric proteins or peptides with appropriate symmetries and topologies are selected and their monomeric components are subsequently fused with peptide linkers to create chimeric building blocks that self-assemble into cage-like architectures. However, genetic tethering of two proteins necessitates C-terminus to N-terminus linkages and may even require protein restructuring using circular permutation to link the proteins at an orientation optimal for self-assembly, which will require judicious selection of both linker placement and the target protein(s). Computational techniques for interface redesign between symmetric building blocks have enabled the generation of a diverse array of protein cages[28,29,67,68]. Protein design affords the creation of thousands of candidates towards a particular assembly motif which, in conjunction with high throughput screening, permits experimental validation of hundreds of potential targets. Thus far, a focus of interface design has required that interprotein interfaces often consist of extensive hydrophobic patches and electrostatic interactions that effectively "glue" the proteins together to create exceptionally stable complexes at the expense of modularity and flexibility. Some recent studies have incorporated responsive elements as part of a designed protein system[69] and continued improvements to computational design methods will perhaps enable the formation of more sophisticated stimuli-responsive assemblies similar to those present in nature. In addition to computational and genetic strategies, two recent reports describe

359 the use of reversible metal coordination motifs to generate protein cages, either by introducing Au-
360 thiol interactions between 11meric proteins[33] or fusing metal-binding coiled-coil peptides onto a
361 trimeric scaffold[34].

362 In contrast to the approaches described above, the metal coordination approach described
363 in this Protocol requires a much smaller design footprint to generate stimuli-responsive bimetallic
364 protein cages from asymmetric monomers. Our approach requires additional manipulation of a
365 protein after expression (bioconjugation to a HA ligand, additional purification, and incubation
366 with metal ions to enable self-assembly) and thus, HA-mediated assemblies cannot be generated
367 *in vivo*. Whereas HA-mediated cages cannot sustain the extreme temperatures and chemical
368 conditions in which computationally designed cages are stable[10,29], fewer protein-protein contacts
369 using reversible chemical bonding interactions enable structural flexibility and modularity.
370 Flexibility, in particular, is a necessary component of self-assembly processes to minimize kinetic
371 traps and permit structural rearrangements as well as exhibiting more biologically representative
372 characteristics (*e.g.*, O_2 binding cooperativity of hemoglobin[70]).

373

374 *Construction of 3D protein lattices*

375 Whereas there have been many reports on the construction of *de novo* designed 0-, 1-, and
376 2D protein assemblies, there has been minimal progress in the predictable construction of 3D
377 lattices. Traditionally, protein crystals are formed in supersaturating solutions by vapor diffusion,
378 promoted by the introduction of precipitating agents (*e.g.*, salts and short polymers). However, it
379 remains a challenge to determine the solubility and crystallizability of a given protein based on its
380 sequence and folds, requiring extensive screening and optimization using decades-old strategies
381 to obtain diffraction-quality 3D protein crystals[71]. One rational method to improve crystallization

has relied on a concept termed surface entropy reduction (SER) wherein flexible residues or loops on the protein surface are replaced with residues and motifs with lower conformational entropy[72]. Alternative approaches have involved the use of designed protein-protein interactions, such as the introduction of disulfide bonds onto monomeric proteins to improve crystallization via symmetrization[73] or installation of electrostatic patches onto symmetric proteins to promote the formation of binary protein lattices[74]. Given that no general strategy has been devised for the predictive crystallization of proteins, the aforementioned methods all provide different approaches that one can adopt towards making 3D protein lattices. However, these strategies require considerable manipulation of a protein building block to promote 3D lattice formation and moreover, they provide little control over the molecular arrangements of the proteins within the 3D crystal. One advantage provided by HA-mediated crystallization of protein-MOFs is that lattice arrangements can be systematically varied by altering metal ion identity or organic HA linkers to afford synthetic modularity, permitting a limited predictive control over crystal packing behavior.

Limitations

HA-mediated protein cages

The formation of HA-mediated protein cages requires site-specific modification of purified protein and additional purification steps prior to performing self-assembly experiments. These steps necessitate that the protein building block is amenable to multiple rounds of purification and buffer exchange (often via centrifugal filtration). The protein must be devoid of non-engineered Cys residues to eliminate unwanted HA reactivity. Should a protein with internal disulfides be the desired building block, exploring incorporation of the HA motif as a UAA is the best course forward. It is important to note that the use of covalently tethered metal binding motifs to control

protein self-assembly is a relatively underexplored concept and the continued advances in protein design over the past few years lend themselves to the marriage of multiple protein engineering strategies, including the ones presented in this Protocol, in the design of novel protein assemblies. Additionally, our HA-mediated protein cages also contain Zn-binding sites introduced for induction of C_2 symmetry, achieved by positioning metal coordinating residues at *i* and *i+4* positions along an α-helical structural motif on the protein. This may limit the choice of building block to proteins that contain α-helical folds to accommodate metal binding chelates as well as the ability for two proteins to associate at the metal binding interface without steric clashes. In principle, β-sheet containing proteins can also accommodate chelating sites if metal-binding residues are placed at *i* and *i+2* positions. Similarly, proteins with well-defined folds can accommodate metal binding by carefully positioning His, Asp or Glu residues such that the side chains are properly oriented to bind transition metal ions in the desired geometry to enable metal-mediated protein-protein association. Such studies would require design of protein scaffolds using computational tools (*e.g.*, PyRosetta[75]) or judicious manual modeling using protein visualization tools, and inevitably, some trial-and-error. These alternative structural solutions present additional options to discover new potential scaffolds for metal-mediated self-assembly.

HA-mediated protein-MOFs

The formation of protein-MOFs is inherently favored through the use of a robust building block with internal 3D symmetry (*e.g.*, tetrahedral, octahedral, and icosahedral), which imposes a somewhat strong restriction on the number of potential building blocks that can be used to create similar 3D protein crystals. These proteins must be stable at pH values near 9 to deprotonate the HA motif and increase metal binding affinity. However, our prior work also indicates that a

428 singular protein building block can be used to create diverse protein-MOF structures by
429 interchanging the identity of the HA ligand, resulting in unique emergent materials properties[11,20,21].
430 Therefore, the relatively small space of highly symmetric, thermostable and soluble proteins could
431 still yield an array of protein-MOF structures with different structural and functional attributes (see
432 **Applications of the Method** section).
433
434 **Experimental design**
435 ***Selection of the protein building block***
436 In this Protocol, we focus on the assembly of HA-mediated protein cages from cytochrome
437 cb_{562} and protein-MOFs from HuHF. When considering the application of our procedures to other
438 protein building blocks, there are several criteria to consider.
439 The protein must be soluble and stable in aqueous buffers, ideally over a broad range of
440 pH values (5.0 – 10.0), to accommodate purification, chemical reduction and bioconjugation,
441 multiple centrifugal filtration steps, and incubation with metal ions at ambient temperatures for
442 many days. Biochemical and biophysical characterization of the protein (*e.g.* size-exclusion
443 chromatography, gel electrophoresis, circular dichroism, and analytical ultracentrifugation) to
444 assess protein purity, chemical and thermal stability, and oligomeric state is useful in determining
445 whether a given building block is amenable to our protocols.
446 Proteins are overexpressed in bacterial *E. coli* cultures, lysed to release soluble proteins,
447 and purified using column chromatography techniques. One of the most common strategies for
448 rapid and facile protein purification involves the use of polyhistidine tags, which are strong metal
449 chelators themselves and must therefore be removed when developing metal-binding protein
450 constructs. This is normally achieved by appending a cleavage site (*e.g.* TEV- or thrombin-

selective cut sites) followed by incubation with the appropriate enzyme after initial purification steps. Cleavage should be followed by additional purification steps to ensure that no extraneous metal-binding residues remain on the protein that may lead to off-pathway oligomerization.

Structurally, the protein must also be tolerant to the installation of metal coordinating residues (Cys for HA labeling, His, Asp, or Glu residues for metal coordination) on its surface without decreasing its solubility or stability. More details for each type of HA-mediated assembly are described in the following sections, commenting on both the specifics for the protein building blocks we have explored and considerations for alternative building blocks.

HA-mediated protein cages

We chose the monomeric four-helix bundle protein, cytochrome cb_{562} for our initial studies based on the aforementioned criteria and familiarity with using this protein in our lab. Since the protein consists almost entirely of α-helices, the precise placement of metal binding residues can be achieved with high specificity without concern for flexible domains altering the position of metal coordination.

Generally speaking, α-helices are a convenient structural motif for the installation of any metal coordinating residues, which is especially important for the coordination of transition metal ions to predictably form C_2 symmetric interfaces. Proteins with α-helical structural motifs are ideal candidates, when considering a bimetallic scaffold which requires both C_2 symmetric Zn^{2+} binding in addition to C_3 symmetric HA-mediated Fe^{3+} coordination.

In our experience, it is best to place metal-binding residues at rigid, surface exposed sites on a protein.[7] We previously installed native metal coordinating residues along Helix 1 of cytochrome cb_{562} to generate bimetallic protein cage (BMC) variants that were able to selectively

474 coordinate Zn^{2+} ions using native His, Asp, and Glu residues and coordinate Fe^{3+} ions at surface
475 Cys residues modified with HA (**Fig. 5a**)[10]. In principle, one can circumvent a bimetallic approach
476 via the selective installation of one or two Cys residues onto a natively oligomeric protein to
477 promote HA-mediated oligomerization. Crucially, the protein must be devoid of any native Cys
478 residues that are not engineered for HA conjugation in order to avoid unwanted modification of
479 multiple residues.

480 In our studies, we observed the formation of both hexameric and dodecameric cages with
481 our BMC variants due to flexibility at the HA site and structural adaptability in secondary
482 coordination to Zn^{2+} ions. Such adaptability is potentially lost when considering a symmetric
483 scaffold. When the building block is arranged into a predetermined geometry (*i.e.* a symmetric
484 building block), the forced symmetry element mitigates the possibility of unforeseen protein
485 arrangements arising from flexible components and unexpected metal coordination. Fortuitously,
486 we observed the formation of two different types of cages by using monomeric protein scaffolds
487 bearing metal binding residues. This resulted in unexpected Zn-binding modes due to the
488 flexibility afforded at the HA site and accommodated by a monomeric protein that did not enforce
489 a particular symmetry on the assembly product. Further studies must be performed to more
490 carefully probe the factors that determine assembly geometry to predictively incorporate flexible
491 components that can alter assembly products in the future.

492 Some factors to consider for new protein scaffolds when searching for the ideal location to
493 place a Cys-HA motif include:

494 1. the surface accessibility of the amino acid and its nearby residues (which could be
495 assessed using SASA calculations)

2. proximity to bulky neighboring amino acids (*e.g.*, potential negative effects on metal coordination efficacy by placing a Cys residue next to a bulky Trp or charged Arg residue)
3. the geometric positioning of additional metal binding sites on the protein (either additional Cys residues for multiple HA binding sites or native metal coordinating residues) to favor the formation of multiple metal nodes in a cooperative fashion to facilitate self-assembly.

When determining the placement of HA motifs for our BMC designs (**Fig. 5a**), surface exposed sites were chosen based on a *in crystallo* cage-like assembly observed for a cytochrome protein.[76] Based on our successes, we surmised that the placement of the HA motif must complement the location of a secondary metal coordination motif; in our case, the HA motifs shown in **Fig. 5a** are located on α-helices at the opposite face of Helix 1, the Zn^{2+} coordination interface. Furthermore, building a structural model of the desired assembly can help elucidate appropriate locations along a 3-fold symmetry axis for the placement of an HA motif (see Extended Data Figure 2 in Golub *et. al.*)[10]. Finally, while some predictive power is available in designing such structures, it will likely be necessary to test the placement of Cys residues at different surface positions to find the optimal assembly construct.

HA-mediated protein-MOFs

We have relied on a symmetric building block to form 3D protein lattices bridged by dihydroxamate linkers. The 24meric, octahedral HuHF is engineered with a His residue at its C_3-symmetric pores to form a tripodal coordination motif for binding a transition metal ion (**Fig. 5d**)[11]. The tetrahedral metal coordination site affords stable binding of transition metal ions while

519 presenting a surface-exposed open site for HA binding. Furthermore, there are no discernable
520 protein-protein contacts near the site of HA binding, enabling free access for the HA ligand to
521 form bridging contacts. The addition of the HA bridging linkers connect ferritin molecules via the
522 C_3 sites to form μm-sized 3D ferritin-MOFs. An octahedral protein building block is not a
523 prerequisite to generate protein-MOF lattices, but the protein must be able to accommodate stable
524 metal coordination nodes at symmetrically positioned surface sites to connect with other proteins
525 and form 2- or 3D lattices.
526 The C_3 symmetric pore of HuHF was particularly useful in this context due to the facile installation
527 of a tripodal metal coordination motif through a single mutation (T122H). A three-coordinate
528 metal binding site is ideal in this instance due to tight coordination to a transition metal ion (a
529 feature most likely absent in monodentate or bidentate metal binding) while leaving an open
530 coordination site for HA binding (which is much more challenging to achieve in a four coordinate
531 site). When searching for alternative scaffolds, the presence of a C_3 symmetric axis would greatly
532 enhance the likelihood of identifying surface locations to easily generate a tripodal metal
533 coordination site and enable the self-assembly of protein-MOFs. Proteins containing internal 3D
534 symmetry (*e.g.*, octahedral or tetrahedral symmetry), like HuHF, are most likely to yield protein
535 MOFs in the current iteration. Such protein building blocks can be readily identified through a
536 survey of the Protein Data Bank (PDB). For instance, selecting for *T* symmetric proteins in the
537 PDB, one can search for proteins with inherent C_3 symmetries that could potentially be useful for
538 protein-MOF construction. One such protein we have discovered in our search is the *H. halophila*
539 dodicin (PDB ID: 2VXA), which could potentially accommodate a tripodal metal coordination site
540 via a Leu9His mutation.
541

542 ***Synthesis of HA ligands***

543 **HA-mediated protein cages**

544 The synthesis of IHA is performed in a straightforward procedure using commercially
545 available reagents (Fig. 6a).[10,77] Since IHA is both temperature and light sensitive, we recommend
546 performing a large-scale synthesis of the 2-chloro-N-hydroxamate and only converting a portion
547 of it to IHA as necessary. IHA can be stored protected from light at −20 °C. A similar synthetic
548 approach can be adopted for alternative chelating motifs, provided that there is an amino
549 nucleophile available to conjugate to chloroacetyl chloride. For some motifs such as 8-
550 hydroxyquinoline and 1,10-phenanthroline, there are published procedures for conversion into
551 Cys-reactive iodo ligands[16].

552

553 *HA-mediated protein-MOFs*

554 The synthetic scheme for *p*-H_2bdh and *m*-H_2bdh is shown in **Fig. 6b, c**[11,20]. The procedure
555 for both ligands is nearly identical, differing only in the use of dimethyl terephthalate for *p*-H_2bdh
556 and dimethyl isophthalate for *m*-H_2bdh. In both instances, hydroxylamine is first deprotonated by
557 the addition of NaOH followed by the addition of the appropriate "phthalate" molecule to yield
558 the desired product. We initially chose the phthalate-based linkers to maintain molecular rigidity
559 and only introduce flexibility at the site of HA. Different ditopic linkers, varying in the
560 hydroxamate-hydroxamate spacing and/or geometry (*e.g.*, a bent vs. collinear orientation), have
561 been used to demonstrate that the ferritin lattice arrangements can be dictated based on the
562 coordination preference of the transition metal ion and the bridging ligand. In addition to varying
563 HA geometry and altering linkers lengths of phthalate-based HA ligands, we have also explored
564 other aromatic linkers containing different heteroatoms (O, N, or S) that may affect the planarity

of the aromatic ring, as well as more flexible linkers containing PEG or carbon spacers in-between the HA motifs. The other bridging linkers can be synthesized as previously described[20,21].

Protein conjugation and purification

HA-mediated protein cages

Before conjugation, Cys-bearing proteins are reduced using a large (50-100-fold) excess of dithiothreitol (DTT) and transferred into an anaerobic chamber (Coy Laboratory Products). The protein solution is buffer exchanged into a freshly degassed reductant-free buffered solution using a 10 D/G column and incubated with a molar excess (15-fold) of IHA to generate the protein-HA conjugate. As an alternative to IHA conjugation in an anaerobic chamber, protein labeling can be performed in the presence of a low concentration of a reductant (*e.g.* 1 mM TCEP) in an O_2 atmosphere. We opted to perform the reaction anaerobically to minimize disulfide formation between protein Cys residues instead of the desired Cys - IHA conjugation.

ESI-MS analysis of the crude protein solution is recommended to ensure that the conjugation proceeded successfully prior to additional purification steps. An Ellman's assay can be performed to assess whether there are Cys residues that remain unmodified; in our experiments with cytochrome cb_{562}, we had to use ESI-MS instead of the Ellman's assay due to overlapping absorbance features[78] with the covalently tethered heme. While there is some batch-to-batch variability, we routinely observe very little unmodified protein after IHA labeling. If a large amount of unreacted protein persists, the conjugation procedure (reduction of Cys-bearing protein, buffer exchange into a reductant-free buffer, and addition of IHA) can be repeated 1-2 more times to improve conjugation yields.

Following conjugation, proteins are purified at pH ≥9.5 to deprotonate the amino group of HA and allow for separation from unconjugated protein using ion-exchange chromatography. Purification is performed under reducing conditions to eliminate a monomer:dimer equilibrium among unmodified proteins. As an additional precautionary measure, proteins are also treated with metal chelating agents (EDTA/DPA) prior to applying them onto an ion-exchange column to remove any metal-bound species from the solution.

Since cytochrome cb_{562} proteins are red, we apply a linear NaCl gradient until we see the protein start to move on the column and elute the protein by holding at that [NaCl]. This allows for better separation between functionalized and native proteins, especially when we use proteins bearing two Cys residues. After the protein band has traveled halfway down the column, the NaCl gradient is continued. When using an uncolored protein, a slowly ramping NaCl gradient is useful in separating unmodified proteins from the protein-HA conjugate. Following these procedures, we can successfully modify and purify single and double Cys-HA BMC variants (**Fig. 5b, c**). BMC3, which forms dodecameric cages, contains Cys63-HA and Cys82-HA; BMC4, which forms D_3 hexameric cages, contains Cys82-HA.

Metal-mediated protein oligomerization

HA-mediated protein cages Metal coordination of HA-bearing BMC proteins with Zn^{2+} and Fe^{3+} ions resulted in the self-assembly of discrete dodecameric and hexameric cages. The addition of iron salts must be performed anaerobically to minimize oxidation of iron species to form insoluble iron hydroxides. We have found that, with our cytochrome cb_{562} variants, protein cages will form even if Fe^{2+} ions are added. We attribute this behavior to Fe^{2+} oxidation to Fe^{3+} by the covalently

609 tethered heme of cytochrome *cb*$_{562}$, which can be observed in a shift in the Soret maximum (415
610 nm to 421 nm)[78].

611 Given that most proteins are not likely to oxidize Fe^{2+} species in solution, one must screen
612 multiple Fe^{3+} salts (*e.g.*, $FeCl_3$, $Fe(acac)_3$, or $Fe(NO_3)_3$) to determine which will produce the
613 highest yield of assembly products. For the formation of bimetallic cytochrome *cb*$_{562}$ cages
614 described in this protocol, Fe^{2+}/Fe^{3+} ions are first added to the protein solution followed
615 immediately by the addition of Zn^{2+} ions. We did not, however, observe any differences in cage
616 formation when the order of addition of Zn^{2+} and Fe^{2+}/Fe^{3+} ions was changed. To further improve
617 cage yields, the protein solution is concentrated 5-6-fold using an Amicon spin filter after an initial
618 3-4 h incubation of protein and metal. We found that we obtained better self-assembly yields when
619 we concentrated the protein after metal incubation than if we performed the reaction at a higher
620 starting protein concentration and omitted the spin filtering step.

621

622 **HA-mediated protein-MOFs**

623 Three components (protein, metal, and linker) must be combined to form protein-MOFs.
624 Incubation of ^{H122}HuHF with divalent transition metal ions generates protein nodes that can be
625 connected using ditopic HA linkers. We usually form protein-MOF crystals in sitting-drop
626 crystallization trays to more easily manipulate the crystals that form. The components are mixed
627 in the top well with a larger reservoir solution present underneath to promote vapor diffusion.
628 However, unlike traditional protein crystallization, the use of a crystallization tray is not necessary
629 and protein-MOFs can also be formed in solution in a glass or plastic vial. While the exact solution
630 conditions vary slightly, we provide a general set of conditions and recommendations for the
631 formation of high-quality ferritin-MOF crystals in this Protocol. Ferritin-MOFs formed in a range

of pH values (8.0 – 10.0). The bridging ligands used to form protein-MOFs suffer from low solubility in aqueous buffers, so lattice formation is generally performed in basic conditions to deprotonate the HA motif and promote metal coordination. We observe the formation of ferritin-MOFs using divalent transition metal ions (*e.g.*, Co^{2+}, Ni^{2+}, Zn^{2+}) wherein the coordination preferences of the metal ion could dictate the resulting 3D lattice symmetry. We recommend a broad screen of transition metal ions to probe the effect of coordination geometry, and possibly redox state of the metal ion, on the formation of protein-MOF lattices. Crystals generally appear in 12-24 h.

Characterization of self-assembly products

HA-mediated protein cages

For our experiments, we primarily used analytical ultracentrifugation (AUC) and transmission electron microscopy (TEM) techniques to observe the formation of protein cages. Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments allowed us to characterize protein oligomers in solution and determine the optimal conditions (*e.g.*, [protein]:[metal] ratio, pH, metal ion identity, protein concentration) necessary to form cages. Size-exclusion chromatography (SEC) might be useful as a complementary technique to AUC to reproducibly differentiate cages from protein monomers; based on preliminary experiments, our bimetallic protein cages were not stable in the column matrix so we did not pursue this further. AUC experiments are time-consuming (16-20 h per sample), so we also used negative-stain TEM experiments to search for ca. 10 nm protein cages. If HA-mediated protein oligomers form >5-10 nm assemblies, protein solutions can be rapidly screened by TEM to identify promising samples for further analysis using AUC or SEC. Cytochrome cb_{562} protein cages form μm-scale 3D crystals

655 in sitting-drop vapor diffusion crystallization trays, allowing us to probe the structure of solution-
656 formed protein cages at atomic resolution using single-crystal X-ray diffraction (sc-XRD)
657 techniques. If crystallization is not feasible, sufficiently large structures can be analyzed using
658 single-particle cryo-electron microscopy.

659 Ferric hydroxamate-bearing siderophores have absorption features at 425-435 nm, which
660 can be measured using a UV-vis spectrometer. It should, therefore, be possible to check for the
661 formation of Fe^{3+}:$(HA)_3$ complexes using circular dichroism techniques due to ligand chirality
662 around the metal center[79]. We did not, however, observe any strong features by UV-vis or CD
663 experiments with protein cages containing Fe^{3+}:$(HA)_3$ complexes, which we attributed to strong
664 interfering absorption of the cytochrome cb_{562} heme in the same spectral region. The appearance
665 of these features may be observable when using uncolored proteins and this would be a convenient
666 technique to observe the formation of Fe^{3+}:$(HA)_3$ complexes in solution.

667

668 **HA-mediated protein-MOFs**

669 After obtaining ferritin-MOF crystals, their molecular details can be probed with sc-XRD
670 experiments. Obtaining high-resolution crystal data can be challenging with ferritin-MOFs due to
671 sparse protein-protein interactions and flexibility at the linker-mediated contact regions. We
672 recommend screening several cryoprotectant solutions to identify optimal freezing conditions[80] or
673 collecting data at room temperature, as the ferritin-MOFs are sensitive to solution perturbations.
674 We have had success with perfluoropolyether, xylitol, and pentaerythritol propoxylate 5/4 PO/OH
675 (PEP) as cryoprotectants. In addition to sc-XRD experiments, protein-MOFs can be analyzed using
676 small-angle X-ray scattering (SAXS) experiments to identify crystallographic parameters (*e.g.*,
677 unit cell parameters and lattice symmetry) of a bulk sample containing hundreds of protein crystals

(in contrast with probing individual crystals for sc-XRD experiments). SAXS experiments can also be performed in a 96-well tray format which allows for screening protein-MOF growth conditions and crystal stability (*e.g.*, varying pH, temperature, and introduction of organic solvents) rapidly.

Materials

Reagents

CAUTION Many reagents used in this protocol are potentially harmful and toxic. Please follow the appropriate safety procedures, such as wearing goggles, gloves, and using a fume hood, as described in the protocol.

- 2,6-Pyridinedicarboxylic acid (DPA; Sigma Aldrich, cat. no. P63808)
- 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; Biopioneer Inc., C0113)
- Acetone (≥99.5%; Fisher Scientific, cat. no. A18-4)
- Chloroacetyl chloride (99.0% (GC); Sigma Aldrich, cat. no. 22880)
- Chloroform (≥99.8%; Fisher Scientific, cat. no. C298-4)
- Deuterated dimethylsulfoxide (DMSO-d_6, Cambridge Isotope Laboratories, Inc., DLM-10)
- Dichloromethane (≥99.5%; Fisher Scientific, cat. no. D37-4)
- Dimethyl isophthalate (Sigma Aldrich, cat. no. 194239)
- Dimethyl terephthalate (Sigma Aldrich, cat. no. 185124)
- Distilled water
- Dithiothreitol (DTT; Fisher BioReagents, cat. no. BP172)
- Ethyl acetate (≥99.5%; Fisher Scientific, cat. no. E145-4)
- Ethylenediaminetetraacetic acid (EDTA; Fisher BioReagents, BP118)
- Ferric (III) chloride hexahydrate ($FeCl_3 \cdot 6\ H_2O$; Fisher Scientific, cat. no. 50146613)

- Formvar/carbon-coated Cu TEM grids (Ted Pella, Inc., cat. no. 01754-F)
- Hydrazine hydrate (80%, Hydrazine, 51%; Acros Organics, cat. no. 209592500)
- Hydrochloric acid (HCl, Fisher Scientific, cat. no. A144S)
- Hydroxylamine hydrochloride (Fisher Scientific, cat. no. MK-5258-125)
- Iron (II) sulfate ($FeSO_4$; Fisher Scientific, I146)
- Iron (III) acetylacetonate ($Fe(acac)_3$; Sigma Aldrich, cat. no. 517003)
- Methanol (≥99.8%, Fisher Scientific, cat. no. A412-4)
- N,N--diisopropylethylamine (99.5%; Acros Organics, cat. no. AC367841000)
- N,N-Dimethylformamide (DMF, Millipore Sigma, cat. no. DX1726)
- N-Cyclohexyl-2-aminoethanesulfonic acid (CHES; Grainger, manufacturer model C40020)
- n-Hexanes (≥98.5%; Fisher Scientific, cat. no. H292-4)
- N-Hydroxyphthalimide (98%; Acros Organics, cat. no. 329875000)
- Pentaerythritol propoxylate 5/4 PO/OH (PEP; Hampton Research, cat. no. HR2-739)
- Silica gel (Fisher Scientific, cat. no. S161-500)
- Sodium chloride (NaCl, Fisher BioReagents, cat. no. BP358-10))
- Sodium hydroxide (NaOH, Fisher Scientific, cat. no. S318-10)
- Sodium iodide (NaI; EMD, cat. no. SX0625-1)
- Sodium sulfate anhydrous (Na_2SO_4, ≥99.0%; Fisher Scientific, cat. no. S421-1)
- Trifluoroacetic acid (Oakwood Chemicals, cat. no. 001271)
- Tris(hydroxymethyl)aminomethane (Tris; Sigma Aldrich, cat. no. T5941)
- Trityl chloride (Triphenylmethyl chloride, 99.43%; Chem Impex Intl Inc., cat. no. 00974)
- Uranyl acetate (Electron Microscopy Sciences, cat. no. 22400)
- Xylitol (≥99.0%; cat. no. X3375)

- Zinc chloride ($ZnCl_2$, Alfa Aesar, A16281)

Equipment

- 0.22-µm filter (Acrodisc 25 mm; Pall Corporation, supplier no. 4612)
- −20 °C freezer
- 5 mL PEEK Sample Loop (BioRad, cat. no. 7500497)
- Aluminum foil
- Amicon membrane (Millipore Sigma, 3 kDa, cat. no. PLBC07610, 10 kDa, cat. no. PLGC07610)
- Amicon Stirred Cell (Millipore Sigma, cat. no. UFSC40001)
- Amicon Ultra spin filters (Millipore Sigma, 3 kDa, cat. no. UFC500324, 10 kDa, cat. no. UFC501024)
- Balance
- Beakers
- BioLogic DuoFlow 10 system (BioRad)
- Biological pipettes (2 µL, 10 µL, 200 µL, 1000 µL)
- Buchner funnel
- Cary 60 UV-Vis spectrometer (Agilent)
- Cell culture plate (ThermoFisher, cat. no. 150628)
- Clear heavy duty Scotch packaging tape
- Cryschem crystallization tray (Hampton Research, cat. no. HR3-160)
- CrystalWand Magnetic (Hampton Research, cat. no. HR4-729)
- Disposable graduated syringes (1 mL, 10 mL)

- DynaLoop 90 (BioRad, part no. 750-0450)
- Econo-Pac 10DG pre-packed desalting column (Biorad, cat. no. 7322010)
- Eppendorf tube rack
- Eppendorf tubes (0.65 mL, 1.5 mL)
- Erlenmeyer flask (250 mL, 500 mL)
- Falcon tube 4-way rack
- Falcon tubes (15 mL, 50 mL)
- FEI Tecnai G2 Sphera
- Graduated cylinders (25, 100 mL)
- Light microscope
- Macroprep High Q-cartridge column (BioRad, cat. no. 7324124)
- Magnetic stir plate with heating capabilities
- Magnetic CryoVial (MiTeGen, cat. no. CV-1-50)
- Micromass Quattro Ultima Triple Quadrupole mass spectrometer
- Mounted CryoLoop (20 micron; Hampton Research, cat. no. HR4-970)
- Needles (BD Precision Glide, cat. no. 305176)
- NMR spectrometers (1H and 13C, 400 MHz or 500 MHz)
- Pasteur pipettes
- pH indicator strips
- Pipette bulbs
- Pyrex crystallizing dish
- Quattro Ultima Triple Quadrupole ESI-MS
- Reflux condenser

- Rotary evaporator (Buchi)
- Round-bottom (RB) flasks (50, 100, 250, 500, 1000 mL)
- Separatory funnel (100 mL, 1 L)
- Side-arm Erlenmeyer flask (250, 500 mL)
- Silica gel column (57 × 508 mm, 1000 mL capacity)
- Silicone oil (Sigma Aldrich, cat. no. 85409)
- Spatula
- Teflon-coated magnetic stir bar
- Thermometer
- TLC Silica gel 60 F254 plate (Merck, cat. no. 105554)
- Tweezer (PELCO Biology by Dumont, cat. no. 510)
- Vacuum pump
- Vinyl anaerobic chamber (Coy Laboratory Products)
- Vivaspin 6 centrifugal concentrator (Viva Products, 10 kDa, cat. no. VS0601)
- Whatman filter paper (1001-185)
- XL-1 analytical centrifuge (Beckman Coulter)

Reagent setup

0.5 M DPA/EDTA stock solution

Dissolve 8.356 g DPA and 14.612 g EDTA in 90 mL ddH_2O. Adjust the pH to 8.0 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

50 mM CHES (pH 8.5), 150 mM NaCl stock solution

Dissolve 1.04 g CHES and 0.88 g NaCl in 90 mL ddH_2O. Adjust the pH to 8.5 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

50 mM CHES (pH 9.5), 150 mM NaCl stock solution

Dissolve 1.04 g CHES and 0.88 g NaCl in 90 mL ddH_2O. Adjust the pH to 9.5 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

20 mM HEPES (pH 7.5) stock solution

Dissolve 0.477 g of HEPES into 90 mL ddH_2O. Adjust the pH to 7.5 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

50 mM $Fe(acac)_3$ stock solution

Dissolve 17.7 mg of $Fe(acac)_3$ into 1 mL ddH_2O. The solution can be stored for 6 h at room temperature.

50 mM $FeSO_4$ stock solution

Dissolve 7.6 mg of $FeSO_4$ into 1 mL ddH_2O. The solution can be stored for 6 h at room temperature.

20 mM Tris (pH 7.5) stock solution

Dissolve 0.242 g of Tris into 90 mL ddH_2O. Adjust the pH to 7.5 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

50 mM Tris (pH 8.5) stock solution

Dissolve 0.606 g of Tris into 90 mL ddH_2O. Adjust the pH to 8.5 and stir until the salts have completely dissolved. Fill to 100 mL and syringe filter through a 0.22-µm membrane. The solution can be stored for 6 months at room temperature.

2% (w/v) Uranyl acetate solution

Dissolve 200 mg of uranyl acetate into 10 mL ddH_2O. Stir in the absence of light for 12 h. Filter the solution through a 0.22-µm membrane. The solution can be stored for 6 months at 4 °C in the absence of light. Periodically filter the solution through a 0.22-µm membrane to remove precipitated uranyl salts.

50 mM $ZnCl_2$ stock solution

Dissolve 6.8 mg of $ZnCl_2$ into 1 mL ddH_2O. The solution can be stored for 6 months at room temperature.

$FeCl_3$ stain (1% (w/v) solution of 1% ferric (III) chloride hexahydrate in 50% aqueous methanol)

Add 1.0 g $FeCl_3$ into a solution containing 50 mL methanol and 50 mL distilled water.

FPLC Buffer A

Dissolve 4.15 g of CHES and 0.309 g of DTT into 950 mL ddH_2O. Adjust the pH to 9.5 and stir until the salts have completely dissolved. Fill to 1000 mL and filter through a 0.22-µm membrane. The solution can be stored for 2-3 d at room temperature.

FPLC Buffer B

Dissolve 4.15 g of CHES, 0.309 g of DTT, and 58.44 g of NaCl into 900 mL ddH_2O. Adjust the pH to 9.5 and stir until the salts have completely dissolved. Fill to 1000 mL and filter through a 0.22-µm membrane. The solution can be stored for 2-3 d at room temperature.

Coy chamber for anaerobic protein self-assembly setup

The anaerobic chamber should be maintained in an oxygen-free (or very low oxygen) atmosphere (*e.g.*, a mix of 10% H_2/90% Ar). Self-assembly buffers are prepared in an anaerobic environment and degassed prior to storage in an anaerobic Coy chamber. Transition metal salts used for self-assembly are taken into the Coy chamber as solid salts in Eppendorf tubes and dissolved into degassed ddH_2O. To set up self-assembly conditions, a stock solution of protein (10-20 µL) was taken into the anaerobic chamber and diluted with degassed buffers.

ESI-MS analysis

Small molecule samples are dissolved in methanol and diluted to a concentration of 0.1-1.0 mg/mL using a solution of 50% methanol in water. Protein samples are buffer exchanged into water using Amicon spin filters and diluted to a concentration of 0.1-1.0 mg/mL using a solution of 0.1% TFA and 50% methanol in water.

FPLC for protein purification

Equilibrate a Q-cartridge at 3 mL/min in FPLC Buffer A for ca. 10 column volumes prior to loading any protein onto the column. Proteins are loaded using either a 5 mL sample loop or a 90 mL DynaLoop and eluted using a linear gradient of NaCl at 3 mL/min. Clean the Q column with ca. 10 column volumes of FPLC Buffer B and equilibrate in FPLC Buffer A prior to the application of additional protein solutions.

NMR analysis

^{1}H and ^{13}C spectra are collected at ~25 °C in DMSO-d_6. NMR chemical shifts (relative to tetramethylsilane) are 2.49 (^{1}H) and 39.5 (^{13}C) for DMSO-d_6. MestReNova software (Mestrelab Research) is used for spectral analysis.

[ALERT: The protocol has two Procedures.]

Procedure 1: <u>HA-mediated protein cages</u>

Preparation of *O*-tritylhydroxylamine (**TIMING**: 10-12 h)

1. Prepare a 250 mL RB flask with a magnetic stir bar and 50 mL of DMF as the solvent.

   CAUTION DMF is a skin irritant, carcinogenic and toxic. Wear goggles and gloves, and work inside a fume hood to avoid breathing in vapors.

2. Add 10.0 g (61.3 mmol, 1.0 equiv.) of *N*-hydroxyphthalimide and 11.78 mL (67.4 mmol, 1.1 equiv.) of *N,N*-diisopropylethylamine to the flask, and stir the solution on a magnetic stir plate.

   CRITICAL STEP The solution should turn bright red (**Fig. 7a**).

3. Add 17.01 g (61.3 mmol, 1.0 equiv.) of trityl chloride to the stirring solution (300 rpm), and allow the mixture to stir for 2 h at room temperature (25 °C).

4. Pour the reaction mixture into a 500-mL beaker with 200 mL of distilled water. Vacuum-filter the precipitate with a Buchner funnel lined with filter paper placed on a side-arm Erlenmeyer flask.

5. Wash the precipitate with an additional 100 mL of distilled water. Allow the precipitate to dry under vacuum for at least 1 h. The resulting crude product, (*N*-(trityloxy)phthalimide), should be a white powder (**Fig. 7b**).

   CRITICAL STEP The white powder should be dried sufficiently to eliminate excess water as the crude product is used directly in the next step without purification. The powder does not have to be completely dry, and minimal water should not affect the next step.

   PAUSE POINT The crude *N*-(trityloxy)phthalimide can be stored at room temperature for at least 3 d.

6. Dissolve the crude product in 600 mL of chloroform in a 1-L RB flask. Add a magnetic stir bar to the flask.

CAUTION Chloroform is a skin irritant and toxic substance with high volatility. Wear goggles and gloves, and work under a fume hood to avoid breathing in vapors.

7. Add 15.0 mL (244 mmol, 3.98 equiv.) of hydrazine hydrate (ca. 51% hydrazine) to 100 mL of methanol. Add the diluted hydrazine solution slowly over 20 min into the stirring solution (300 rpm) and allow the mixture to stir for 6 h at room temperature (25 °C).

   CAUTION Hydrazine hydrate is a skin irritant, carcinogenic, and toxic. Wear goggles and gloves, and work inside a fume hood to avoid breathing in vapors.

8. Pour the reaction mixture into a separatory funnel and mix with 300 mL of distilled water. Separate the organic layer and wash with distilled water (2 × 300 mL).

9. Collect the organic layer and dry with the addition of anhydrous Na_2SO_4 until you see white clumps. Decant the solution, and remove the solvent via rotary evaporation at 40 °C.

   CRITICAL STEP The crude product should be an off-white oil.

10. Purify the crude product (*O*-tritylhydroxylamine) via silica gel column chromatography (ca. 100 g Silica gel) using a gradient of 0-10% ethyl acetate in hexanes as the eluent. Remove the solvent via rotary evaporation at 40 °C and dry *in vacuo* to give a white solid (**Fig. 7c**).Yield: 10.9 g (39.6 mmol, 65% yield) Theoretical Yield: 16.9 g (61.3 mmol)

    CRITICAL STEP The crude product from step 9 can be dissolved in ethyl acetate. The non-dissolvable white precipitate might be on the top of the column and it is not product. Expect to use ~1 L of eluent.

CAUTION Silica gel may cause an allergic skin reaction and asthma symptoms. Work under a fume hood to avoid breathing the dust.

CAUTION Hexanes and ethyl acetate are skin irritants with high volatility. Wear goggles and gloves, and work inside a fume hood to avoid breathing in vapors.

TROUBLESHOOTING

PAUSE POINT *O*-tritylhydroxylamine can be stored at room temperature for at least 12 months.

Preparation of 2-iodo-*N*-hydroxyacetamide (IHA) (**TIMING:** 6-7 h)

11. Prepare a water-ice slurry in a Pyrex crystallizing dish. The temperature should be <4 °C.

12. Add 2.0 g (7.3 mmol, 1.0 equiv.) of *O*-tritylhydroxylamine and 2.5 mL (14.5 mmol, 2 equiv.) of *N,N*-Diisopropylethylamine to 15 mL of dichloromethane in a 50 mL RB flask with a magnetic stir bar. Place the RB flask into the ice bath such that the entire solution is submerged (**Fig. 7d**).

CAUTION *N,N*-Diisopropylethylamine is highly flammable and toxic. Wear goggles and gloves and avoid any contact with skin or eyes. Keep away from heat and flames.

13. Add 0.58 mL (7.3 mmol, 1.0 equiv.) of chloroacetyl chloride to 2.0 mL of dichloromethane. Add the diluted chloroacetyl chloride solution dropwise into the stirring suspension (300 rpm) over 5 min. **CRITICAL STEP** The reaction mixture will turn cloudy.

CAUTION Chloroacetyl chloride is toxic. Wear goggles and gloves. Avoid any contact with skin or eyes.

14. Remove the ice bath and allow the mixture to slowly warm to room temperature (25 °C). Stir for 1 h.

15. Add 15 mL of dichloromethane and pour the mixture into a separatory funnel. Wash with distilled water (3 × 30 mL).

16. Collect the organic layer and dry it with the addition of anhydrous Na_2SO_4 until you see white clumps. Decant the solution, and remove the solvent via rotary evaporation at 40 °C.

17. Prepare a solution of 10% (v/v) trifluoroacetic acid in 15 mL of dichloromethane. Add to the residue and stir for 30 min at room temperature (25 °C).

    CAUTION Trifluoroacetic acid is toxic and corrosive. Wear goggles and gloves. Avoid any contact with skin or eyes.

18. Add 5 mL of methanol to the reaction mixture to get rid of the excess trifluoroacetic acid. Remove the solvent via rotary evaporation at 40 °C.

19. Add 10 mL ethyl acetate to dissolve the crude product. A white precipitate should form. Filter the white precipitate and retain the filtrate. Remove the solvent from the filtrate via rotary evaporation at 40 °C and dry in vacuo.

20. Purify the crude product (2-chloro-*N*-hydroxyacetamide) via silica gel column chromatography using a gradient of 0-100% ethyl acetate in hexanes as the eluent.
    - Load a 1000 mL capacity silica column (57 × 508 mm) with 25% ethyl acetate in hexanes.
    - Dissolve the crude reaction mixture in a small volume of ethyl acetate.
    - Load the sample and run 200 mL of 25% ethyl acetate in hexanes.
    - Run 400 mL of 50% ethyl acetate in hexanes. A yellow solution will elute (by-product).
    - Run 400 mL of 75% ethyl acetate in hexanes followed by 400 mL of 100% ethyl acetate. The desired product should elute around 80-100% ethyl acetate in hexanes. The product should be yellow-orange in color.
    - Follow the elution of the product via thin-layer chromatography (TLC) using a $FeCl_3$ stain.
    - Remove solvent by rotary evaporation at 40 °C and dry *in vacuo*. Yield: 450 mg (4.1 mmol, 57% yield) Theoretical Yield: 800 mg (7.3 mmol)

    CRITICAL STEP Depending on the reaction yield and the size of the column, the product may continue to elute at 100% ethyl acetate. If the column is packed shorter (*e.g..,* a 1-2 inch tall silica bed), the product will elute in fewer fractions.

    TROUBLESHOOTING

    PAUSE POINT The pure product can be stored at room temperature for at least 6 months.

21. Analyze the structure and purity of the product by NMR spectral analysis. The product can be dissolved in DMSO-d_6.

22. Heat an oil bath in a Pyrex crystallizing dish to at least 65 °C.

23. Add 400 mg (3.7 mmol, 1.0 equiv.) of 2-chloro-*N*-hydroxyacetamide and 2.7 g (18.3 mmol, 5.0 equiv.) of NaI to 30 mL of acetone in a 50 mL RB flask fitted with a magnetic stir bar. Wrap the flask with aluminum foil to perform the reaction in the dark. Attach a reflux condenser to the flask, and reflux for 1 h.

CRITICAL STEP The product formed in the reaction (2-iodo-*N*-hydroxyacetamide) is light sensitive. Conduct this step in the dark.

24. Allow the mixture to cool and remove the solvent via rotary evaporation.

25. Purify the crude product (2-iodo-*N*-hydroxyacetamide) using a small silica plug with 100% ethyl acetate as the eluent. Lightly wrap the column in aluminum foil and perform the purification with minimal ambient light.

- Add a small volume (10-15 mL) of ethyl acetate to the crude product.
- Run a silica plug with 100% ethyl acetate to remove precipitated salts. The eluent should be yellow-orange in color
- Remove the solvent to yield a solid. Repeat the silica plug 1-2 more times to remove any residual salts.
- Remove the solvent for the final time and dry *in vacuo* overnight. The pure product should be an orange solid.

Yield: ca. 700 mg (ca. 3.5 mmol, >90% yield) Theoretical Yield: 730 mg (3.7 mmol)

CRITICAL STEP A small volume of ethyl acetate is necessary to dissolve the crude product because NaI is partially soluble in ethyl acetate. If some residual salt remains in the sample after purification, it should not hinder eventual protein conjugation.

TROUBLESHOOTING

PAUSE POINT 2-iodo-*N*-hydroxyacetamide can be stored at -20 °C in the dark for at least 6 months.

26. Analyze the structure and purity of the product by NMR spectral analysis. The product can be dissolved in DMSO-d_6.

IHA labeling onto Cys-bearing proteins and post-labeling purification (TIMING: 18-24 h)

CRITICAL The self-assembly of cages using protein-HA conjugates has been reported using engineered variants of cytochrome cb_{562}[10]. The expression and purification of cytochrome cb_{562} has been previously described[81]. While protein conjugation and cage formation can be performed as described for other proteins, notes will be placed throughout the protocol specific to the hemoprotein.

CRITICAL IHA labeling and protein self-assembly involves multiple centrifugation steps for protein concentration, which may not be tolerated by some proteins. If your protein is more sensitive, buffer exchanging via dialysis to avoid repeated centrifugation is a potential alternative, but the steps described below presume the use of a protein building block that is amenable to repeated centrifugation procedures while maintaining stability in solution.

27. Prepare a stock solution of a Cys-bearing protein in a 15-mL Falcon tube. The following protocol will be using 2.7 mL of 100 µM protein. For a medium-scale preparation, 2-3 mL of 100 µM protein is advised.

CRITICAL STEP Using a buffered solution of 20 mM Tris (pH 7.5) works well when tested with cytochrome cb_{562} variants. In our hands, a buffered solution at pH 7-8 is appropriate at this step.

28. Dissolve 4.16 mg (100 equiv.) of DTT in 300 µL of the same buffer used in Step 27.

29. Add the DTT solution to the protein solution and gently mix to homogeneity to give a final volume of 3 mL. Place in an anaerobic chamber uncovered (uncap the Falcon tube and place on a 4-way tube rack) so that there is a chance for any dissolved O_2 to be removed.

CRITICAL STEP If using cb_{562} proteins, there will be a noticeable colorimetric change from red to pink due to a spectroscopic shift in the Soret maximum from 415 nm to 421 nm. This can be confirmed by measuring a small sample of protein on a UV-Vis spectrometer.

30. Equilibrate a 10DG desalting column with a degassed, buffered solution containing 20 mM HEPES (pH 7.5). A 10DG column with a 10 mL bed volume should be equilibrated with at least 20 mL of buffer prior to use.

31. Apply up to 3 mL of the protein solution to the column and elute with 4 mL of the degassed solution containing 20 mM HEPES (pH 7.5).

32. Dissolve 0.9 mg (15 equiv.) of IHA in degassed 100 µL DMF in a 1.5-mL Eppendorf tube. Add the IHA solution to the protein, gently mix to homogeneity, and allow to react overnight.

CRITICAL STEP Protect the IHA and protein solutions from light to prevent degradation of IHA prior to protein conjugation. We opt to cover the tubes in aluminum foil for this step.

33. Remove the protein solution from the anaerobic chamber. Analyze the crude conjugated product by ESI-MS to confirm the formation of the protein-HA adduct.

CRITICAL STEP It is important to verify that the IHA conjugation step was successful before proceeding to the purification procedure. Otherwise, the protein will needlessly be subject to column chromatography without yielding any HA-conjugated samples.

TROUBLESHOOTING

PAUSE POINT Crude protein-IHA conjugate can be stored at 4 °C for 1 week or flash frozen and stored at -80 °C for 6 months.

34. Concentrate the protein solution to 3 mL using a Vivaspin 6 concentrator at 8,000 x g for 5 min. If the volume after concentration is < 3 mL, add a buffered solution containing 20 mM CHES (pH 9.5) and 2 mM DTT (FPLC Buffer A).

35. Equilibrate a 10DG desalting column with at least 20 mL of a buffered solution containing 20 mM CHES (pH 9.5).

36. Apply 3 mL of the protein solution to the column and elute with 4 mL of FPLC Buffer A.

37. Load the solution using a 5-mL injection loop onto a Duoflow workstation equipped with a Macroprep High Q-cartridge at 1 ml/min. The column should be equilibrated in FPLC Buffer A (see Equipment setup).

38. Purify the crude protein-IHA conjugate using a linear gradient over 0 – 0.5 M NaCl using FPLC Buffers A and B.
    - Apply a linear gradient of 0 – 0.3 M NaCl over a 200 mL volume.
    - At approximately 0.15 M NaCl, stop the linear gradient and hold at that [NaCl] until protein begins to elute.
    - Once protein begins to elute, proceed with the linear gradient up to 0.3 M NaCl.
    - Ramp from 0.3 – 0.5 M NaCl over a 60 mL volume. Any remaining unconjugated protein should elute during this step.

CRITICAL STEP For our cytochrome proteins, we monitor the protein on the Q-cartridge and stop the linear gradient after observing protein movement. We then maintain this salt concentration (ca. 0.1 – 0.15 M NaCl) until the band is approximately 50% down the column.

TROUBLESHOOTING

39. Assess sample purity of the FPLC fractions by ESI-MS and pool to combine. Fractions near elution peaks and troughs can be tested first to reduce the total number of samples that need to be assessed by mass spectrometry.

40. Combine pure fractions and concentrate to < 3 mL using an Amicon equipped with a 3 kDa membrane.

41. From a 100 mM stock solution, add a DPA/EDTA mixture to the protein solution to a final concentration of 5 mM DPA/EDTA and incubate for 1-2 h.

42. Equilibrate a 10DG desalting column with at least 20 mL of a buffered solution containing 20 mM Tris (pH 7.5). Apply 3 mL of the protein solution to the column and elute with 4 mL of the buffered solution containing 20 mM Tris (pH 7.5).

43. Assess protein concentration on a UV-vis spectrometer.

44. (optional) Concentrate the protein using a 10 kDa Amicon spin filter (12,000 × g, 10 min) to a final concentration of at least 1 mM. This step is not necessary for self-assembly and was performed in our lab to make sample preparation easier. If the protein is unstable at high concentrations, the dilute protein solution must be degassed in the assembly buffer outside the anaerobic chamber prior to the preparation of protein cages in the following section.

PAUSE POINT The pure protein-HA conjugate can be stored at 4 °C for 1 month or flash frozen and stored at -80 °C for 6 months.

Preparation of protein cages (**TIMING:** 1-7 d)

CRITICAL The protocol outlined below details the formation of bimetallic hexameric or dodecameric cytochrome *cb*$_{562}$ cages with Zn^{2+} and Fe^{3+} coordination. HA motifs will selectively bind Fe^{3+}, so the addition of Zn^{2+} is not necessary for any designed systems solely dependent on Fe^{3+}:$(HA)_3$ complex formation for self-assembly.

45. Bring a stock solution (> 500 µM) of HA-conjugated protein in a 0.65-mL Eppendorf tube into an anaerobic chamber. A 15-µL aliquot of 500 µM protein (per Cys-HA) will be required for one cage sample. If a more dilute protein solution is preferred, samples must be degassed outside the anaerobic chamber prior to the next steps.

CRITICAL STEP Cage preparation must be performed anaerobically. Please refer to Equipment Setup to ensure the anaerobic chamber contains the necessary reagents and equipment.

46. Buffer exchange the protein using a 10 kDa Amicon spin filter (12,000 × g, 10 min) into a degassed, buffered solution containing 20 mM Tris (pH 8.5). Perform the step five times to ensure that the protein is thoroughly exchanged into the degassed buffer.

47. Remove a small aliquot of protein from the anaerobic chamber and measure its concentration on a UV-vis spectrometer.

CRITICAL STEP A small amount of protein will inevitably be lost during the buffer exchange process so a volumetric conversion based on the initial protein concentration will likely be inaccurate.

48. In a 1.5-mL Eppendorf tube, prepare a 50 mM stock solution of $FeSO_4$ or $Fe(acac)_3$ in degassed water. Perform a serial dilution into a second 1.5-mL Eppendorf tube to a final concentration of 5 mM of the metal salt.

CRITICAL STEP Prepare these solutions immediately prior to setup of the self-assembly solutions. The iron salts form yellow precipitates within a few hours.[82]

49. Prepare the cage self-assembly solution in a 1.5-mL Eppendorf tube, as detailed in the table below. Add the components in the order listed. The setup should result in [Protein (per Cys-HA)] = 20 µM at a Protein:Fe:Zn ratio of 1:1:3.

TROUBLESHOOTING

Component	Ratio	Stock concentration	Final concentration	Volume
Protein	1	500 µM	20 µM	15 µL
Buffer (50 mM Tris pH 8.5)	N/A	50 mM	20 mM	150 µL
Water	N/A	N/A	N/A	204 µL
$FeSO_4$ or $Fe(acac)_3$	1	5 mM	20 µM	1.5 µL
$ZnCl_2$	3	5 mM	60 µM	4.5 µL

50. After 3-4 h, concentrate the self-assembly solution using a 10 kDa Amicon spin filter (12,000 × g, 10 min) to a final volume of 50 µL. If the solution volume is too low after concentration, dilute to 50 µL using the eluent. Cage formation can be monitored over several days.

CRITICAL STEP While the sample preparation as listed above is sufficient to form protein cages, we have found that the additional concentration step after a few hours of incubation improves cage yield.

Characterization of self-assembled cages (**TIMING:** 2-24 h)

51. There are various methods to assess protein cage formation and estimate cage yield. Analytical ultracentrifugation (AUC) enables the quantification of oligomeric species and determine the overall yield of the cages relative to monomers or smaller oligomers in solution.

In our lab, sedimentation velocity measurements are performed on a XL-1 analytical centrifuge (Beckman Coulter) and scans are analyzed using SEDFIT. Additional details on AUC procedures can be found here.[83,84]

Transmission electron microscopy (TEM) can be used to screen multiple conditions rapidly to identify conditions that give rise to self-assembled cages. Self-assembled protein cages can be observed via negative-stain TEM.

In our lab, we perform negative-stain TEM experiments using a FEI Tecnai G2 Sphera operating at 200 keV and collected micrographs are analyzed using Fiji (http://fiji.sc/Fiji).

Protocols for AUC or TEM characterization are described in options A and B respectively .

(A) AUC characterization of protein cages (**TIMING:** 16-24 h)

i. **Experimental setup** Load 350 µL of the protein sample using a gel-loading tip into a two-sector cell with a 30-50 µL excess of an appropriate buffer blank (*i.e.*, the buffer used for sample preparation from Step 45).

ii. Place the sample cell into the rotor and a weighted blank cell (or secondary sample cell) as a counterbalance at the opposite location in the rotor. Secure the rotor into the centrifuge, being sure that the laser attachment is fastened correctly.

iii. Perform a test scan at the wavelength of choice (*e.g.*, 415 nm at the Soret maximum for cytochrome cb_{562} proteins) at 3,000 rpm and 25 °C. This initial measurement is used to ensure that the sample cell is not leaking and that the absorbance values fall within a reasonable range (0.5 – 1).

iv. Sediment the sample at 135,000 × g (41,000 rpm) at 25 °C. Monitor continuously at the wavelength of choice for at least 500 scans for 16-20 h. Once absorbance readings are nearly 0, the sample has fully sedimented.

v. **Sample analysis** Load the sedimentation velocity scans (400-450 scans) into SEDFIT.

vi. Manually set cell and data-fitting limits on the scans. These positions will remain fixed during the fitting procedure.

vii. Estimate the partial specific volume (mL/g) by taking the quotient of protein volume and the molecular weight. For cb_{562} samples, we use 0.7313 mL/g.

viii. Estimate the buffer viscosity and buffer density of the sample using SEDNTERP.

ix. Enter the estimated partial specific volume, buffer viscosity, and buffer density and fit the data to a continuous molecular weight (c(M)) or sedimentation coefficient (c(S))distribution. Use an initial confidence of 0.95.

x. Use the "Run" command in SEDFIT to set the baseline and time-invariant noise of the scans.

xi. After an initial run, fit the weight-averaged frictional coefficient (f/f_0) of the protein using the "Fit" command. The value should be between 1.1 – 1.4 for symmetric structures. At this stage, use an initial confidence of 0.0.

xii. After fitting, use the "Run" command at a confidence of 0.95 to yield the final distribution profile.

The final distribution profile can be copied into a spreadsheet and plotted to afford molecular weight distributions (c(M)) or sedimentation distributions (c(S)) and determine the percentage of each oligomeric species present in the sample.

(B) TEM characterization of protein cages (**TIMING:** 1-2 h)

i. **Sample preparation** Using an Emitech K100X Glow Discharge machine, negatively glow-discharge formvar/carbon-coated Cu grids (Ted Pella, Inc.) at ~25 mA for 45 s.

ii. Using a reverse tweezer, pick up the grid. Pipette 3.5-4 µL of the protein solution from Step 46 onto the glow-discharged side of the grid and allow to bind for 5 min.

iii. Prepare 3 × 20 µL water droplets on parafilm. Gently wash the grids with MilliQ water by dipping the grid into a water droplet and blotting using Whatman filter paper. Repeat this process for all three water droplets.

CRITICAL STEP Be sure not to completely dry the grid during the blotting steps. There should be a small amount of moisture remaining on the grid prior to the addition of uranyl acetate.

iv. Pipette 3.5 µL of a 2% uranyl acetate solution onto the grid and allow to bind for 1 min.

v. Blot dry using a Whatman filter paper and return the grid into its storage container.

vi. **TEM imaging** Insert the grid into the sample holder. For imaging, use objective-lens underfocus settings ranging from 500 nm to 1.5 µm.

vii. After data collection, micrographs are loaded into Fiji for further analysis.

TROUBLESHOOTING

PROCEDURE 2: HA-mediated protein-MOFs

Preparation of bidentate linkers.

1. In this protocol, we describe the detailed synthesis of *p*-H_2bdh and *m*-H_2bdh (options A and B). The synthesis of *m*-H_2bdh is nearly identical to that of *p*-H_2bdh, differing primarily in the addition of dimethyl isophthalate instead of dimethyl terephthalate.

Other bidentate linkers can be used for the formation of protein-MOFs and can be synthesized as previously reported[20,21].

A Preparation of N^1,N^4-dihydroxyterephthalamide (*p*-H_2bdh) (**TIMING:** 18-24 h)

i. Pour 20 mL of methanol as the solvent into a 50 mL Falcon tube.

ii. Add 1.06 g (15.45 mmol, 1 equiv.) of hydroxylamine hydrochloride and 1.24 g (30.9 mmol, 2 equiv.) of NaOH to the Falcon tube. Shake vigorously to mix the solution thoroughly and pour it into a 100-mL RB flask containing a magnetic stir bar.

CAUTION Sodium hydroxide is corrosive. Wear gloves and goggles. Avoid inhalation and any contact with skin or eyes.

iii. Place the RB flask in an ice bath such that the entire solution is submerged for at least 10 min. A solid precipitate (NaCl) should form in the solution.

iv. Vacuum-filter the precipitate with a Buchner funnel lined with filter paper placed on a side-arm Erlenmeyer flask.

v. Add 1 g (5.15 mmol, 0.33 equiv.) of dimethyl terephthalate to 30 mL of methanol and combine with the filtrate. Stir the solution overnight at room temperature.

vi. Remove the solvent via rotary evaporation at 40 °C.

vii. Dissolve the remaining solid material in 20 mL H_2O. Add 5% HCl to acidify the solution to a pH of 5.5. Check the pH periodically while adding HCl using pH strips.

CRITICAL STEP A white precipitate should form.

CAUTION Hydrochloric acid is corrosive. Wear gloves and goggles. Avoid inhalation and any contact with skin or eyes.

viii. Vacuum-filter the precipitate with a Buchner funnel lined with filter paper placed on a side-arm Erlenmeyer flask. Remove the solvent via rotary evaporation at 40 °C and dry *in vacuo*. Yield: 0.66 g (3.35 mmol, 65% yield), Theoretical yield: 1.01 g (5.15 mmol).

PAUSE POINT The pure product can be stored at room temperature for at least one year.

B Preparation of N^2,N^3-dihydroxyisophthalamide (*m*-H_2bdh) **(TIMING:** 18-24 h)

i. Pour 20 mL of methanol as the solvent into a 50 mL Falcon tube.

ii. Add 1.06 g (15.45 mmol, 1 equiv.) of hydroxylamine hydrochloride and 1.24 g (30.9 mmol, 2 equiv.) of NaOH to the Falcon tube. Shake vigorously to mix the solution thoroughly and pour it into a 100-mL RB flask containing a magnetic stir bar.

CAUTION Sodium hydroxide is corrosive. Wear gloves and goggles. Avoid inhalation and any contact with skin or eyes

iii. Place the RB flask in an ice bath such that the entire solution is submerged for at least 10 min. A solid precipitate (NaCl) should form in the solution.

iv. Vacuum-filter the precipitate with a Buchner funnel lined with filter paper placed on a side-arm Erlenmeyer flask.

v. Add 1 g (5.15 mmol, 0.33 equiv.) of dimethyl isophthalate to 30 mL of methanol and combine with the filtrate. Stir the solution overnight at room temperature.

vi. Remove the solvent via rotary evaporation at 40 °C.

vii. Dissolve the remaining solid material in 20 mL H_2O. Add 5% HCl to acidify the solution to a pH of 5.5. Check the pH periodically while adding HCl using pH strips.

CRITICAL STEP A white precipitate should form.

CAUTION Hydrochloric acid is corrosive. Wear gloves and goggles. Avoid inhalation and any contact with skin or eyes.

viii. Vacuum-filter the precipitate with a Buchner funnel lined with filter paper placed on a side-arm Erlenmeyer flask. Remove the solvent via rotary evaporation at 40 °C and dry *in vacuo*. Yield: 0.7 g (3.55 mmol, 69% yield), Theoretical yield: 1.01 g (5.15 mmol).

PAUSE POINT The pure product can be stored at room temperature for at least one year.

Preparation of ferritin-MOFs (TIMING: 24-72 h)

<CRITICAL> In this protocol, we use a variant of human heavy-chain ferritin, ^{H122}HuHF, for the formation of ferritin-MOFs. The protein can be expressed and purified, as described previously[11]. After purification, the protein concentrated to 25 µM (24meric cage), filtered through a 0.22-µm filter, and stored at 4 °C for long term storage in a buffered solution containing 50 mM CHES (pH 8.5), 150 mM NaCl. Avoid using any buffers that would strongly chelate metal ions and inhibit protein-HA interactions. When considering alternative protein scaffolds, ensure that the protein is stable at pH 8-10 at 200 µM for protein-MOF self-assembly conditions (described below).

2. Warm a stock solution of ^{H122}HuHF (25 µM protein in a buffered solution containing 50 mM CHES (pH 8.5), 150 mM NaCl) to room temperature prior to self-assembly experiments. Prepare a 10 mM solution of the bidentate bridging linker (*p*-H_2bdh or *m*-H_2bdh) in 50 mM CHES (pH 9.5), 150 mM NaCl.

3. To determine the optimal conditions for growing high-quality crystals, the concentration of each component can be varied. As an example, here are variations presented for the components used to generate ferritin-MOFs: 1-12.5 µM of ferritin cage in a buffered solution

containing 50 mM CHES (pH 8.5), 150 mM NaCl, 0.5-2 mM of the bidentate bridging linker, and 50-150 equiv. of a transition metal salt ($CoCl_2$, $NiCl_2$, or $ZnCl_2$) per ferritin cage.

Prepare the sitting drop and a reservoir solutions separately and pipette into a 24-well Cryschem crystallization tray. See the table below for an example set of crystallization conditions to use for the formation of ferritin-MOFs.

CRITICAL STEP A thorough screen of commonly used precipitating and crowding agents is recommended to identify optimal conditions for the formation of diffraction-quality crystals; in our experience, low molecular weight PEGs (PEG 300, PEG 350 MME, and PEG 400) or pentaerythritol propoxylate (5/4 PO/OH) were the most successful. For initial screens, the reservoir solution can be supplemented with 0-20% of a crystallization precipitant.

Component	Stock concentration	Final concentration	Volume
Reservoir (500 µL)			
NaCl	5 M	150 mM	15 µL
CHES (pH 8.5)	500 mM	50 mM	50 µL
$ZnCl_2$	10 mM	0.47 mM	23.7 µL
H_2O	N/A	N/A	411.3 µL
Sitting drop (12 µL)			
Protein	25 µM	4 µM	2 µL

Linker	10 mM	2 mM	2.4 µL
Reservoir	N/A	N/A	7.6 µL

4. After preparing 24 screening solutions on a Cryschem Plate, cover the tray with clear packaging tape. Be sure to press down and flatten the tape across the plate such that the solutions cannot evaporate or mix with neighboring wells.

5. Inspect the crystal trays under a light microscope. After 24 h, crystals should appear in the wells. Crystals are suitable for harvesting for structural analysis after 1-2 d.

Characterization of ferritin-MOFs (TIMING: 2-14 d)

6. Analyze protein-MOF crystals using sc-XRD and/or SAXS measurements following the steps in options A and B respectively. For our work, sc-XRD data are collected at a synchrotron source (*e.g.*, Stanford Synchrotron Radiation Laboratory or the Advanced Light Source at Lawrence Berkeley National Laboratory) and analyzed using a suite of X-ray crystallography programs[85-88].

   SAXS data are collected at a synchrotron source (*e.g.*, Argonne National Laboratory-Advanced Photon Source) and analyzed using the powder diffraction processing software JADE (MDI). Minor adjustments for processing ferritin-MOFs can be found in Bailey *et. al.*[20]

(A) Structural analysis of ferritin-MOFs using sc-XRD (TIMING: 1-7 d)

i. Briefly soak a single crystal in a cryoprotectant solution for 5-10 s using a mounted CryoLoop. For ferritin-MOFs, we have found success using a 50% (w/v) solution of xylitol or PEP in 50 mM CHES (pH 8.5 – 9.5), 150 mM NaCl.

ii. Rapidly plunge the crystals into liquid N_2. Transfer the CryoLoop into a magnetic CryoVial using a magnetic CrystalWand.

CRITICAL STEP Once the crystals are frozen, they should be handled at liquid N_2 temperatures. Do not allow the crystals to warm up to maintain suitable conditions for sc-XRD data collection.

iii. Collect data at a synchrotron facility. For ferritin-MOFs, data were collected at 100 K using 0.98 Å radiation.

iv. After data collection, process the collected images using a standard protein structural determination workflow. Briefly, integrate the collected images using iMosflm and scale and merge the data using Aimless. Perform molecular replacement with Phaser using a previously solved structure as a search model. Perform rigid-body and further refinements in Phenix.

(B) Structural analysis of ferritin-MOFs using SAXS (TIMING: 1-7 d)

i. Prepare crystals for SAXS in 12-well cell culture plates. To gather enough crystals for SAXS measurements, each plate contains a single metal/linker combination. An example set of

crystallization conditions to generate ferritin-MOFs for SAXS analysis is shown in the table below.

Component	Stock concentration	Final concentration	Volume
Metal stock solution (5000 µL)			
NaCl	5 M	150 mM	150 µL
CHES (pH 9.5)	500 mM	50 mM	500 µL
$ZnCl_2$	10 mM	0.789 mM	394.7 µL
H_2O	N/A	N/A	3955.3 µL
Culture plate well (200 µL)			
Protein	25 µM	4 µM	33.3 µL
Linker	10 mM	2 mM	40 µL
Metal stock solution	N/A	N/A	126.7 µL

ii. Crystals should form in 12-24 h. Harvest crystals after 3 days, combining the crystals from all 24 wells into a 1.5-mL Eppendorf tube.

iii. After the crystals are settled into the bottom of the tube, carefully pipette them into a 1.5-mm quartz capillary tube with 50 µL of the reservoir solution from the tube. Seal the end of the capillary with modeling clay. Samples can be shipped to synchrotron facilities and stored under ambient temperature.

iv. Collect data at a synchrotron facility. For ferritin-MOFs, data were collected with collimated X-rays (0.7293 Å, 17 keV) with dimensions of 250 x 250 μm and exposure times between 0.5 – 2 s. Scattered radiation was collected with a CDD area detector, and 1D scattering data were obtained through an azimuthal averaging of 2D data to obtain plots of scattering intensity as a function of the scattering vector q: $q = 4\pi \sin(\theta) / \lambda$, where θ is ½ of the scattering angle, and λ is X-ray wavelength.

v. Analyze data using the JADE processing software. Simulated powder diffraction modeling of the SAXS profiles can be generated in Mercury[89].

Timing

To successfully complete the steps outlined in the protocol, researchers must have expertise in the following areas: basic organic synthesis, recombinant protein expression and protein purification, protein bioconjugation, biochemical analysis of proteins (*e.g.*, UV-Vis, circular dichroism, PAGE), and macromolecular structural biology (*e.g.*, single-crystal XRD, TEM, cryo-EM, SAXS).

Steps 1-10, preparation of O-tritylhydroxylamine: 10-12 h

Steps 11-26, preparation of 2-iodo-N-hydroxyacetamide (IHA): 6-7 h

Steps 27-40, IHA labeling onto Cys-bearing proteins and post-labeling purification: 18-24 h

Steps 41-46, preparation of protein cages: 1-7 d

Steps 47-58, AUC characterization of protein cages: 16-24 h

Steps 59-65, TEM characterization of protein cages: 1-2 h

Steps 66-73, preparation of N^1,N^4-dihydroxyterephthalamide: 18-24 h

Steps 74-81, preparation of N^2,N^3-dihydroxyterephthalamide: 18-24 h

Steps 82-85, preparation of ferritin-MOFs: 24-72 h

Steps 86-89, structural analysis of ferritin-MOFs using sc-XRD: 1-7 d

Steps 90-94, structural analysis of ferritin-MOFs using SAXS: 1-7 d

TROUBLESHOOTING

Troubleshooting guidelines can be found in Table 1.

Table 1: Troubleshooting

Step	Problem	Possible Reason	Solution
10	The purified product is an oil and challenging to work with.	There is residual solvent trapped in the oil.	(i) Add minimal isopropanol to the oil, sonicate, and dry *in vacuo*. This should turn the oil into a white powder (ii) After collecting the product fractions, allow the solvent to slowly evaporate in a fume

			hood uncovered. This should result in large crystals along the sides of the glass tubes, which can be washed briefly with hexanes to get rid of impurities.
	The product elutes with the by-products of the reaction.	The by-product has a similar polarity to the product.	Adjust the gradient from 0-20% ethyl acetate in hexanes to get better separation on the column.
20	The purification of 2-chloro-*N*-hydroxyacetamide cannot be followed by UV light.	The product is not UV active.	Apply a $FeCl_3$ stain to visualize the product.
25	Synthesis of IHA produces a very low yield	The reaction was not carried out in the dark.	Cover the reaction flask with aluminum foil and turn the fume hood lights off
31	I do not see any protein-IHA conjugate by ESI-MS	The reaction did not proceed in high yield.	Repeat steps 27-30. Any Cys residues that were not modified initially

			can be re-reduced and conjugated to IHA.
		The protein is not easily ionized in the mass spectrometer.	Use MALDI instead of ESI-MS to measure the protein mass and determine whether the protein-IHA conjugate was successfully formed.
35	There is only one protein peak observed in the FPLC chromatogram.	IHA conjugation to protein proceeded in high yields.	If the conjugation efficiency was nearly quantitative, only one protein species would be observed in the purification procedure.
		Both protein-HA conjugate and unmodified protein eluted at similar times.	(i) Combine fractions and repeat the purification using a slower linear gradient. (ii) Hold [NaCl] once you see movement of the protein bands on the column (for colored

			proteins) or hold at a concentration previously identified to be sufficient for protein elution. Hold until the protein elutes or a clear separation of colored bands is visible on the column.
45	There is precipitation in my protein solution after combining all of the components.	The [metal]:[protein] ratio is too high, leading to nonspecific metal-mediated aggregation.	Screen lower [metal]:[protein] ratios and check whether that results in less aggregation. A lower starting protein concentration may also be necessary if aggregation persists.
		Precipitation of Fe salts	Make sure to use a freshly prepared Fe stock solution in the anaerobic chamber (Coy Laboratory Products) to

minimize precipitation of Fe salts.

Anticipated results

*HA-mediated protein cage*s

Following the procedures detailed in this Protocol, IHA can be synthesized and yield pure product at an overall yield of 30-35% (**Supplementary Figures 1-5**). Most major impurities should be removed after the purification of 2-chloro-*N*-hydroxyacetamide. The final conversion of 2-chloro-*N*-hydroxyacetamide to 2-iodo-*N*-hydroxyacetamide can be performed in high yield with nearly quantitative conversion so additional purification steps are unnecessary and should be avoided as they will risk degradation of the 2-iodo-*N*-hydroxyacetamide product.

Site-selective conjugation of IHA to Cys-bearing proteins and subsequent purification by FPLC will yield pure protein conjugate, separated from any unmodified proteins during the FPLC NaCl gradient (**Fig. 5b, c**). Based on the quantities used in this Protocol, we routinely obtain 30-40% yield for protein conjugates bearing two Cys residues and 50-60% yield for protein conjugates bearing a single Cys residue. The conjugation procedure described in this Protocol can be repeated on a previously modified batch of proteins if overall yields are poor.

Incubation of both Fe^{2+} and Zn^{2+} ions with our cytochrome cb_{562} yields discrete dodecameric and hexameric protein cages (**Fig. 8a-c**). AUC experiments are useful for solution characterization of the self-assembled particles and additionally serve to help identify conditions under which self-assembly occurs poorly (*e.g.*, absence of both metal ions and using an impure HA-conjugated

protein solution). Detailed procedures and characterization of HA-mediated protein cages can be found in Golub *et. al*[10].

HA-mediated protein-MOFs

The synthesis of *p*-H_2bdh and *m*-H_2bdh should yield pure product at 60-70% yields. The procedure should result in minimal impurities, obviating a need for column chromatography to isolate the pure ditopic HA linkers (**Supplementary Figures 6-9**). **Fig. 8d** represents a prototypical image of ferritin-MOF crystals formed after incubation of ^{H122}HuHF with Zn^{2+} and *p*-H_2bdh. If large (> 100 μm) crystals do not form after 12-24 h, screen precipitants at varying concentrations. Ferritin-MOF crystals can be readily observed in SAXS experiments, with distinct SAXS profiles for the HA linker-mediated assembly of body-centered lattices (**Fig. 8e**). The peaks in the SAXS profile are unique to the molecular arrangement of the lattice and small shifts in these peaks can reflect changes in lattice symmetry or dimension. Crystals can be grown in a bulk solution to generate the large volume of sample required for SAXS analysis, whereas diffraction quality crystals for sc-XRD analysis should be performed in Cryschem crystallization trays. Detailed procedures and characterization of ferritin-MOFs are described in previous publications[11,20,21].

Analytical data for synthesized molecules

***O*-tritylhydroxylamine**

^{1}H NMR: (400 MHz, DMSO-d6) 7.47 – 7.43 (m, 6H, aromatic H) , 7.37 – 7.26 (m, 9H, aromatic H), δ 4.95 (br s, 2H). ^{13C}NMR: (100 MHz, DMSO-d6) δ 143.2, δ 128.8, δ 127.8, δ 127.2,δ 90.8. Measured molecular weight (m/z): 242.99 [M − H^+ − ONH_2]; calculated: 275.35 [M − H^+].

2-chloro-*N*-hydroxyacetamide

^{1}H NMR: (300 MHz, DMSO-d6) δ 10.88 (s, 1H), δ 9.15 (s, 1H), δ 3.93 (s, 2H). ^{13C}NMR: (500 MHz, DMSO-d6) δ 162.88, δ 40.45. Measured molecular weight (m/z): 108.37 [M − H^+]; calculated: 107.99 [M − H^+].

2-iodo-*N*-hydroxyacetamide

^{1}H NMR: (300 MHz, DMSO-d6) δ 10.81 (s, 1H), δ 9.09 (s, 1H), δ 3.51 (s, 2H). ^{13C}NMR: (500 MHz, DMSO-d6) δ 164.83, δ −2.01. Measured molecular weight (m/z): 223.85 [M + Na^+]; calculated: 223.95 [M + Na^+].

***N*^{1},*N*^{4}-dihydroxyterephthalamide**

^{1}H NMR: (400 MHz, DMSO-d6) δ 11.35 (br s, 2H), δ 9.17 (br s, 2H), δ 7.80 (s, 4H). ^{13}C NMR: (500 Hz, DMSO-d6): δ 163.42, δ 135.04, δ 126.92. Measured molecular weight (m/z) = 196.97 [M + H^+]; calculated: 197.05) [M + H^+].

***N*^{1},*N*^{3}-dihydroxyisophthalamide**

^{1}H NMR: (400 MHz, DMSO-d6) δ 11.30 (br s, 2H), δ 9.14 (br s, 2H), δ 8.14 (s, 1H), δ 7.85 (dd, 2H), δ 7.53 (t, 1H). ^{13}C NMR (500 MHz, DMSO-d6): δ 163.65, δ 133.07, δ 129.29, δ 128.54, δ 125.85. Measured molecular weight (m/z) = 197.05 [M + H^+]; calculated: 197.05) [M + H^+].

Acknowledgments

We thank N. Avakyan and R. Alberstein for helpful discussions. This work was supported by the US Department of Energy (Division of Materials Sciences, Office of Basic Energy Sciences; DE-SC0003844; for protein-MOF data collection and analysis) and by the National Science Foundation (Division of Materials Research; DMR-1602537 and DMR-2004558; for protein cage data collection and analysis). R.H.S. was supported by the National Institute of Health Chemical Biology Interfaces Training Grant UC San Diego (T32GM112584).

Author contributions

E.G. and F.A.T. conceived the protein cage project. J.Z., J.A.C. and Y.L. synthesized HA ligands for the protein cage project. E.G. and R.H.S. performed protein cage experiments and data analysis. J.B.B. and F.A.T. conceived the protein-MOF project. J.B.B. synthesized ditopic HA ligands and performed protein-MOF experiments and data analysis. R.H.S. and F.A.T. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing financial interests.

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.

References

1 Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 551-575 (2015).

2 Barber, J. Photosystem II: the engine of life. *Q. Rev. Biophys.* **36**, 71-89 (2003).

3 Eitoku, M., Sato, L., Senda, T. & Horikoshi, M. Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life. Sci.* **65**, 414-444 (2008).

4 Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **46**, 23-42 (2017).

5 Bai, Y., Luo, Q. & Liu, J. Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **45**, 2756-2767 (2016).

6 Hamley, I. W. Protein Assemblies: Nature-Inspired and Designed Nanostructures. *Biomacromolecules* **20**, 1829-1848 (2019).

7 Churchfield, L. A. & Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **52**, 345-355 (2019).

8 Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).

9 Cannon, K. A., Ochoa, J. M. & Yeates, T. O. High-symmetry protein assemblies: patterns and emerging applications. *Curr. Opin. Struc. Biol.* **55**, 77-84 (2019).

10 Golub, E. *et al.* Constructing protein polyhedra via orthogonal chemical interactions. *Nature* **578**, 172-176 (2020).

11 Sontz, P. A., Bailey, J. B., Ahn, S. & Tezcan, F. A. A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. *J. Am. Chem. Soc.* **137**, 11598-11601 (2015).

12 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).

13 Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **346**, 1525-1528 (2014).

14 Churchfield, L. A., Medina-Morales, A., Brodin, J. D., Perez, A. & Tezcan, F. A. De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **138**, 13163-13166 (2016).

15 Radford, R. J., Nguyen, P. C., Ditri, T. B., Figueroa, J. S. & Tezcan, F. A. Controlled Protein Dimerization through Hybrid Coordination Motifs. *Inorg. Chem.* **49**, 4362-4369 (2010).

16 Radford, R. J., Nguyen, P. C. & Tezcan, F. A. Modular and Versatile Hybrid Coordination Motifs on Alpha-Helical Protein Surfaces. *Inorg. Chem.* **2010**, 7106-7115 (2010).

17 Yang, M. & Song, W. J. Diverse protein assembly driven by metal and chelating amino acids with selectivity and tunability. *Nat. Commun.* **10**, 5545 (2019).

18 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).

19 Crumbliss, A. L. Iron bioavailability and the coordination chemistry of hydroxamic acids. *Coord. Chem. Rev.* **105**, 155-179 (1990).

20 Bailey, J. B., Zhang, L., Chiong, J. A., Ahn, S. & Tezcan, F. A. Synthetic Modularity of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **139**, 8160-8166 (2017).

21 Bailey, J. B. & Tezcan, F. A. Tunable and Cooperative Thermomechanical Properties of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* (2020).

22 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).

23 Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **336**, 1129 (2012).

24 Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065-1071 (2014).

25 Cannon, K. A., Nguyen, V. N., Morgan, C. & Yeates, T. O. Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. *ACS Synth. Biol.* **9**, 517-524 (2020).

26 Cristie-David, A. S. *et al.* Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. *J. Am. Chem. Soc.* **141**, 9207-9216 (2019).

27 Sinclair, J. C., Davies, K. M., Venien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558-562 (2011).

28 Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein dodecahedron. *Nature* **535**, 136-139 (2016).

29 Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394 (2016).

30 Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365-1368 (2015).

31 Suzuki, Y. *et al.* Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369-373 (2016).

32 Alberstein, R., Suzuki, Y., Paesani, F. & Tezcan, F. A. Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly. *Nat. Chem.* **10**, 732-739 (2018).

33 Malay, A. D. *et al.* An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438-442 (2019).

34 Cristie-David, A. S. & Marsh, E. N. G. Metal-dependent assembly of a protein nano-cage. *Protein Sci.* **28**, 1620-1629 (2019).

35 Song, W. J., Sontz, P. A., Ambroggio, X. I. & Tezcan, F. A. Metals in Protein-Protein Interfaces. *Ann. Rev. Biophys.* **43**, 409-431 (2014).

36 Salgado, E. N. *et al.* Metal-Templated Design of Protein Interfaces. *Proc. Natl. Acad. Sci. USA* **107**, 1827-1832 (2010).

37 Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **43**, 661-672 (2010).

38 Salgado, E. N., Faraone-Mennella, J. & Tezcan, F. A. Controlling Protein-Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein. *J. Am. Chem. Soc.* **129**, 13374-13375 (2007).

39 Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L. & Tezcan, F. A. Control of Protein Oligomerization Symmetry by Metal Coordination: C_2 and C_3 Symmetrical Assemblies through Cu^{II} and Ni^{II} Coordination. *Inorg. Chem.* **48**, 2726-2728 (2009).

40 Brodin, J. D. *et al.* Evolution of Metal Selectivity in Templated Protein Interfaces. *J. Am. Chem. Soc.* **132**, 8610-8617 (2010).

41 Medina-Morales, A., Perez, A., Brodin, J. D. & Tezcan, F. A. In Vitro and Cellular Self-Assembly of a Zn-Binding Protein Cryptand via Templated Disulfide Bonds. *J. Am. Chem. Soc.* **135**, 12013-12022 (2013).

42 Brodin, J. D., Smith, S. J., Carr, J. R. & Tezcan, F. A. Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **137**, 10468-10471 (2015).

43 Song, W. J., Yu, J. & Tezcan, F. A. Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo-β-lactamases. *J. Am. Chem. Soc.* **139**, 16772-16779 (2017).

44 Churchfield, L. A., Alberstein, R. G., Williamson, L. M. & Tezcan, F. A. Determining the Structural and Energetic Basis of Allostery in a De Novo Designed Metalloprotein Assembly. *J. Am. Chem. Soc.* **140**, 10043-10053 (2018).

45 Edwardson, T. G. W. & Hilvert, D. Virus-Inspired Function in Engineered Protein Cages. *J. Am. Chem. Soc.* **141**, 9432-9443 (2019).
46 Heddle, J. G., Chakraborti, S. & Iwasaki, K. Natural and artificial protein cages: design, structure and therapeutic applications. *Curr. Opin. Struct. Biol.* **43**, 148-155 (2017).
47 Margolin, A. L. & Navia, M. A. Protein crystals as novel catalytic materials. *Angew. Chem. Int. Ed.* **40**, 2205-2222 (2001).
48 Abe, S. *et al.* Design of Enzyme-Encapsulated Protein Containers by In Vivo Crystal Engineering. *Adv. Mater.* **27**, 7951-7956 (2015).
49 McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 2-20 (2014).
50 Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: A paradigm for protein assemblies. *J. Mol. Biol.* **269**, 665-675 (1997).
51 Mateu, M. G. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys.* **531**, 65-79 (2013).
52 Pearson, R. G. Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **85**, 3533-3539 (1963).
53 Baldwin, A. D. & Kiick, K. L. Tunable Degradation of Maleimide–Thiol Adducts in Reducing Environments. *Bioconjug. Chem.* **22**, 1946-1953 (2011).
54 Toussaint, L., Bertrand, L., Hue, L., Crichton, R. R. & Declercq, J.-P. High-resolution X-ray Structures of Human Apoferritin H-chain Mutants Correlated with Their Activity and Metal-binding Sites. *J. Mol. Biol.* **365**, 440-452 (2007).
55 Kanekiyo, M., Ellis, D. & King, N. P. New Vaccine Design and Delivery Technologies. *J. Infect. Dis.* **219**, S88-S96 (2019).
56 Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415-420 (2017).
57 Uchida, M. *et al.* Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **12**, 942-953 (2018).
58 MaHam, A., Tang, Z. W., Wu, H., Wang, J. & Lin, Y. H. Protein-Based Nanomedicine Platforms for Drug Delivery. *Small* **5**, 1706-1721 (2009).
59 Dreschel, H. & Winkelmann, G. in *Transition metals in microbial metabolism* (eds. G Winkelmann & CJ Carrano) Ch. Iron chelation and siderophores, 1-51 (Amsterdam, The Netherlands, Hardwood Acad., 1997).
60 Kaes, C., Katz, A. & Hosseini, M. W. Bipyridine: The Most Widely Used Ligand. A Review of Molecules Comprising at Least Two 2,2‘-Bipyridine Units. *Chem. Rev.* **100**, 3553-3590 (2000).
61 Hofmeier, H. & Schubert, U. S. Recent developments in the supramolecular chemistry of terpyridine–metal complexes. *Chem. Soc. Rev.* **33**, 373-399 (2004).
62 Jiang, P. & Guo, Z. Fluorescent detection of zinc in biological systems: recent development on the design of chemosensors and biosensors. *Coord. Chem. Rev.* **248**, 205-229 (2004).
63 Xie, J., Liu, W. & Schultz, P. G. A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew. Chem. Int. Ed. Engl.* **46**, 9239-9242 (2007).
64 Lee, H. S., Spraggon, G., Schultz, P. G. & Wang, F. Genetic Incorporation of a Metal-Ion Chelating Amino Acid into Proteins as a Biophysical Probe. *J. Am. Chem. Soc.* **131**, 2481-2483 (2009).
65 Li, M. *et al.* Spectroscopic and Crystallographic Investigations of Novel BODIPY-Derived Metal–Organic Frameworks. *Inorg. Chem.* **54**, 1346-1353 (2015).
66 Bandara, H. M. D. & Burdette, S. C. Photoisomerization in different classes of azobenzene. *Chem. Soc. Rev.* **41**, 1809-1825 (2012).
67 King, N. P. *et al.* Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171-1174 (2012).
68 King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108 (2014).
69 Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205-210 (2019).

70 Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **115**, 1702-1724 (2015).
71 McPherson, A. & Cudney, B. Optimization of crystallization conditions for biological macromolecules. *Acta. Cryst. F. Struct. Biol. Commun.* **70**, 1445-1467 (2014).
72 Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* **16**, 1569-1576 (2007).
73 Banatao, D. R. *et al.* An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16230-16235 (2006).
74 Künzle, M., Eckert, T. & Beck, T. Binary Protein Crystals for the Assembly of Inorganic Nanoparticle Superlattices. *J. Am. Chem. Soc.* **138**, 12731-12734 (2016).
75 Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689-691 (2010).
76 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).
77 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).
78 Faraone-Mennella, J., Tezcan, F. A., Gray, H. B. & Winkler, J. R. Stability and folding kinetics of structurally characterized cytochrome cb_{562}. *Biochemistry* **45**, 10504-10511 (2006).
79 Carrano, C. J. & Raymond, K. N. Coordination chemistry of microbial iron transport compounds. 10. Characterization of the complexes of rhodotorulic acid, a dihydroxamate siderophore. *J. Am. Chem. Soc.* **100**, 5371-5374 (1978).
80 Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 622-642 (2015).
81 Bailey, J. B., Subramanian, R. H., Churchfield, L. A. & Tezcan, F. A. in *Methods in Enzymology* Vol. 580 (ed Vincent L. Pecoraro) 223-250 (Academic Press, 2016).
82 Hem, J. D. & Skougstad, M. W. Coprecipitation effects in solutions containing ferrous, ferric, and cupric ions. Report No. 1459E, (1960).
83 Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci.* **11**, 2067-2079 (2002).
84 Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **84**, 143-179 (2008).
85 Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033 (2012).
86 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133-144 (2010).
87 Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125-132 (2010).
88 Terwilliger, T. C. *et al.* Phenix - a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, Medium: ED (2009).
89 Macrae, C. F. *et al.* Mercury 4.0: from visualization to analysis, design and prediction. *J. Appl. Crystallogr.* **53**, 226-235 (2020).

Figure 1 | Hydroxamic acid-mediated protein self-assembly. a, Metal-binding modes for hydroxamic acids to form a discrete C_3 symmetric node or act as a bridging linker between two metal binding nodes. **b,** HA-mediated assembly of discrete protein cages via chemical conjugation of a HA motif onto a Cys residue. **c,** HA-mediated assembly of 3D protein-MOF lattices via ligand-mediated crystallization of symmetric HuHF.

Figure 2 | Design strategies for *de novo* protein self-assembly. a, Genetic fusion of natively symmetric protein oligomers can yield multidimensional assemblies upon protein expression. **b,** Computational redesign of protein-protein interfaces can be used to generate associative patches between symmetric building blocks to create larger assemblies. **c,** Installation of Cys residues on a symmetric protein oligomer can be used to trigger self-assembly via oxidation of Cys thiols. **d,** Installation of metal chelating motifs onto a protein building block can result in the formation of multidimensional protein assemblies.

Figure 3 | Experimental overview for the generation of HA-mediated protein cages. The protocol for HA-mediated formation of protein cages consists of four major parts: synthesis of a Cys-reactive HA molecule, protein conjugation and purification, preparation of protein cages, and characterization of the self-assembled structures.

Figure 4 | Experimental overview for the generation of HA-mediated protein-MOFs. The protocol for HA-mediated formation of protein-MOFs consists of three major parts: synthesis of ditopic HA bridging linkers, preparation of protein-MOFs with HuHF, and characterization of protein-MOFs.

Figure 5 | Selection of the protein building blocks for HA-mediated self-assembly. a, Overview of the cytochrome cb_{562} scaffold with potential binding sites for Zn^{2+} and Fe^{3+} shown as sticks. Mass spectra for **b,** BMC3 and **c,** BMC4 proteins conjugated to HA. BMC3 contains two Cys-HA motifs, BMC4 contains one Cys-HA motif. **d,** Structural overview of ^{H122}HuHF with insets showing tripodal metal coordination at the three-fold HuHF interface.

Figure 6 | Synthetic schemes for the generation of HA ligands. a, Chemical synthesis of IHA, broken down into three major steps. **b,** Chemical synthesis of *p*-H_2bdh. **c,** Chemical synthesis of *m*-H_2bdh. Yields are reported at each major step of the synthesis.

Figure 7 | Experimental setup and representative images of products in the synthesis of IHA. a, Setup for Step 2 in the synthesis of *O*-tritylhydroxylamine. The solution should turn red following the addition of DIPEA to *N*-hydroxyphthalimide. **b,** Image of crude *N*-(trityloxy)phthalimide (Step 5). **c,** Image of pure *O*-tritylhydroxylamine (Step 10). **d,** Ice-bath setup for Step 12 in the synthesis of 2-chloro-*N*-hydroxyacetamide.

Figure 8 | Anticipated results for HA-mediated protein self-assembly. a, AUC and **b,** TEM characterization of self-assembled dodecameric BMC3 cages upon addition of Fe^{2+} and Zn^{2+}. **c,** AUC characterization of self-assembled hexameric BMC4 cages upon addition of Fe^{2+} and Zn^{2+}. Both sets of AUC profiles reveal smaller, non-specific oligomers under improper self-assembly conditions. Adapted from reference 10. **d,** Representative light micrograph of *p*-bdh - Zn^{2+} - ferritin-MOFs. **e,** Experimental SAXS profile for body-centered cubic ferritin-MOF lattice. This figure is adapted in part from reference 10.