Received: 3 May 2024 | Revised: 18 June 2024 | Accepted: 23 June 2024

DOI: 10.1002/pro.5104

RESEARCH ARTICLE

THE PROTEIN SOCIETY WILEY

Unveiling the stochastic nature of human heteropolymer ferritin self-assembly mechanism

Fadi Bou-Abdallah[1] | **Jeremie Fish**[2] | **Genki Terashi**[3] | **Yuanyuan Zhang**[3] | **Daisuke Kihara**[3] | **Paolo Arosio**[4]

[1]Department of Chemistry, State University of New York, Potsdam, New York, USA

[2]Department of Electrical & Computer Engineering, Coulter School of Engineering, Clarkson University, Potsdam, New York, USA

[3]Department of Biological Sciences and Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

[4]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

Correspondence
Fadi Bou-Abdallah, Department of Chemistry, State University of New York, Potsdam, NY 13676, USA.
Email: bouabdf@potsdam.edu

Funding information
National Science Foundation, Division of Molecular and Cellular Biosciences (MCB), Grant/Award Number: 1934666; National Institutes of Health, Grant/Award Number: R01GM133840; National Science Foundation, Grant/Award Numbers: DBI2003635, DBI2146026, IIS2211598, DMS2151678, CMMI1825941; Research Corporation for Science Advancement, Cottrell Instrumentation Supplements Award, Grant/Award Number: 27452

Review Editor: Aitziber L. Cortajarena.

Abstract

Despite ferritin's critical role in regulating cellular and systemic iron levels, our understanding of the structure and assembly mechanism of isoferritins, discovered over eight decades ago, remains limited. Unveiling how the composition and molecular architecture of hetero-oligomeric ferritins confer distinct functionality to isoferritins is essential to understanding how the structural intricacies of H and L subunits influence their interactions with cellular machinery. In this study, ferritin heteropolymers with specific H to L subunit ratios were synthesized using a uniquely engineered plasmid design, followed by high-resolution cryo-electron microscopy analysis and deep learning-based amino acid modeling. Our structural examination revealed unique architectural features during the self-assembly mechanism of heteropolymer ferritins and demonstrated a significant preference for H-L heterodimer formation over H-H or L-L homodimers. Unexpectedly, while dimers seem essential building blocks in the protein self-assembly process, the overall mechanism of ferritin self-assembly is observed to proceed randomly through diverse pathways. The physiological significance of these findings is discussed including how ferritin microheterogeneity could represent a tissue-specific adaptation process that imparts distinctive tissue-specific functions to isoferritins.

KEYWORDS

cryo-EM, ferritin subunits, ferritin microheterogeneity, human heteropolymer ferritin, isoferritins, self-assembly mechanism

1 | INTRODUCTION

Iron is an essential element for life, serving as a critical cofactor for numerous biological processes, including oxygen transport, electron transfer, and DNA synthesis. Although indispensable, labile iron ions can engage in harmful free radical reactions by generating highly reactive hydroxyl radicals capable of inflicting permanent damage to DNA and proteins. Most organisms have evolved complex biochemical mechanisms to maintain a balance between iron as an essential nutrient and iron as a potential cytotoxin. To safeguard against these toxic

Jeremie Fish and Genki Terashi contributed equally to this work.

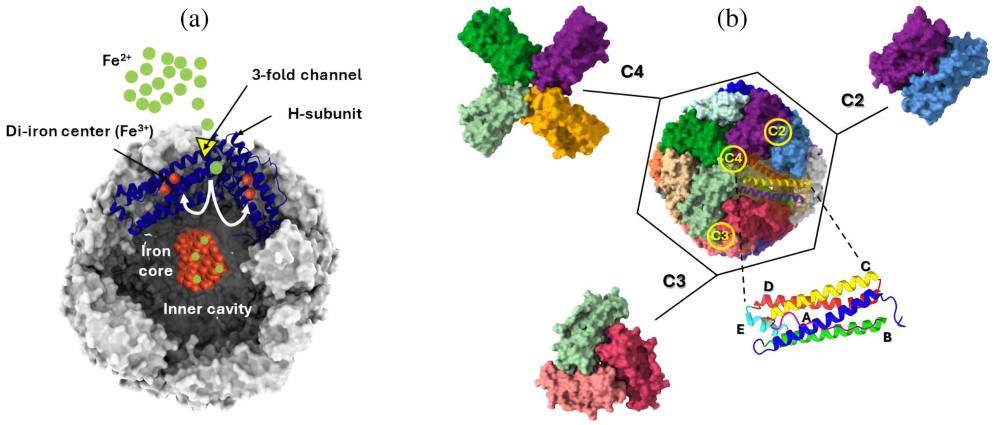

FIGURE 1 (a) Structure of a ferritin molecule showing (a) Fe^{2+} ions entrance through the threefold channel, oxidation at di-iron centers, and inorganic Fe^{3+} iron core deposition inside the ferritin inner cavity, (b) the five helices (A-E) of one bundle monomer, and the inter-subunit interactions of each subunit at the C2 (dimer), C3 (trimer), and C4 (tetramer) symmetry interfaces. The figure is generated using UCSF ChimeraX software (version 1.7.1) (Meng et al., 2023) and ferritin PDB:7jgk.

effects, cells produce ferritin, a ubiquitous multi-subunit iron storage and detoxification protein capable of sequestering thousands of iron atoms in the form of inorganic ferrihydrite core within its inner cavity (Arosio et al., 2009; Bou-Abdallah, 2010; Chasteen, 1998; Harrison & Arosio, 1996) (Figure 1a). Ferritin is a protein nanocage composed of 24 subunits, forming a hollow sphere with an outer diameter of ~120 Å and an inner core diameter of ~80 Å. Each ferritin subunit folds into four α-helix bundles (helices A–D) with helices B and C connected by a loop L that traverses the length of the bundle and a short fifth helix E that ends at the C-terminal (Figure 1b). Within the octahedral cage structure (with a 432-point group symmetry), every subunit engages with six neighboring monomers through three distinct symmetry-related interfaces (Figure 1b).

Eukaryotic ferritins are hetero-polymeric complexes (i.e., isoferritins) of two structurally similar, but functionally different subunit types, named **H** for **H**eavy and **L** for **L**ight. The H-subunit is responsible for the rapid oxidation of Fe^{2+} to Fe^{3+} at a dinuclear ferroxidase center (Figure 1a), whereas the L-subunit is believed to assist in iron nucleation and mineralization. While homopolymer H-ferritin is able to rapidly take up iron and form iron cores, homopolymer L-ferritin can still accumulate iron and produce iron cores, albeit at a much slower rates, probably because of acidic residues lining the surface of the inner cavity (Zhao et al., 2003).

Surprisingly, despite their discovery more than eight decades ago and the pivotal role that ferritin plays in iron homeostasis, notwithstanding the widespread occurrence of heteropolymer ferritins in tissues of vertebrates, virtually nothing is known about the structural and functional differences between isoferritin populations and the impact of varying proportions of H- and L-subunits on the biochemical and functional properties of these unique proteins. This is largely due to the lack of a good expression system that can produce these complex nanostructures in a controlled manner. In a recent study (Srivastava et al., 2021), our group was able to successfully engineer a novel plasmid design capable of generating ferritin heteropolymers with specific H to L subunit ratio, allowing the synthesis of a full spectrum of heteropolymer ferritins, from H-rich ferritin, such as those found in the heart and brain, to L-rich ferritins as found in the spleen, liver, and placenta, and any combinations in-between. Although mammalian ferritins are heteropolymeric, exceptions are human mitochondrial ferritin, a 24-subunit homopolymer composed of 100% H-like subunits (Bou-Abdallah et al., 2005), and human serum ferritin, a 24-subunit homopolymer composed mainly of L-subunits (Wang et al., 2010). Interestingly, the iron cores formed in H-rich and L-rich heteropolymer ferritins exhibit significant differences, both structurally and morphologically, pointing to a deeply intertwined and intricate role for the H and L subunits in the protein nanostructure (Longo et al., 2023), underscoring the important physiological and functional significance of isoferritins populations and their heterogeneous distribution in tissues (Srivastava et al., 2023).

Although our knowledge of ferritin self-assembly/disassembly mechanisms is not fully elucidated, most studies consider ferritin dimers as essential and stable intermediates in protein assembly (more on that topic in the Results section under the subtitle "Ferritin self-assembly: Proposed mechanism"). However, there is evidence that folded monomers exist as separate states, suggesting that self-assembly might also proceed through

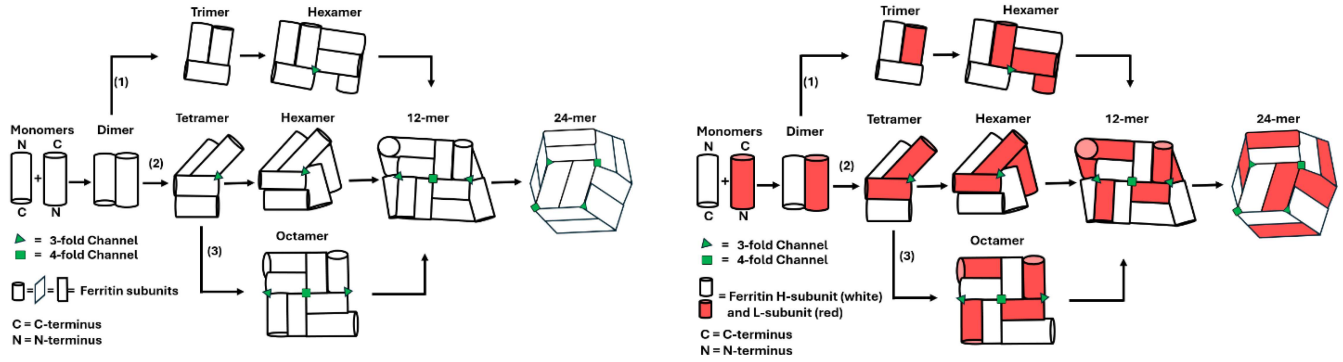

FIGURE 2 Different proposed mechanisms of ferritin self-assembly. The left schematic illustrates three distinct assembly pathways (1, 2, and 3) involving various configurations of ferritin subunits (i.e., dimers, trimers tetramers, hexamers, octamers, 12-mer), culminating in the formation of a 24-mer assembly. On the right, one configuration highlights the favored interaction between H and L heterodimers, resulting in the assembly of a 24-mer human heteropolymer ferritin structure. Due to the preference for H/L heterodimer formation and the specific number of threefold and fourfold channels formed (as shown in Table 1), only certain structures are permissible, despite the potential for different spatial arrangements of H and L subunits.

odd-numbered intermediates like trimers. A schematic (Figure 2) of all the possibilities leading to the 24-subunit ferritin complex shows that three valid pathways (e.g., 2→3→6→12→24, 2→4→6→12→24 or 2→4→8→12→24) can lead to a fully assembled ferritin nanocage (Gerl et al., 1988; Gerl & Jaenicke, 1987; Mohanty et al., 2021; Sato et al., 2016; Sato & Ikeguchi, 2019; Stefanini et al., 1987; Sudarev et al., 2023), where the numbers represent the progression of the self-assembly process (i.e., dimer, trimer, hexamer, etc., until the 24-mer structure is fully assembled).

Some studies demonstrated the association of different ferritin subunits from different species (Otsuka et al., 1980), or different subunit types from the same species (Carmona et al., 2017; Yao et al., 2022), illustrating a conserved nature of inter-subunit interaction domains and recognition mechanisms despite differences in amino acid composition. These results suggest a flexibility in subunit exchange that preserves the integrity of the protein nanocage structure. In human ferritin, despite the H and L subunits sharing only about 55% sequence homology, the assembled 24-subunit ferritins are tightly packed together leaving eight narrow threefold hydrophilic channels, and six fourfold hydrophobic channels of ~4 Å in diameter (Bou-Abdallah, 2010). Studies using Förster Resonance Energy Transfer revealed that the formation of H/L-heterodimers in human ferritin is both kinetically quicker and thermodynamically more favorable than H/H-homopolymers, indicating a preference for the heteropolymeric nature of structured ferritin over homopolymers (Carmona et al., 2017), consistent with the findings of our investigation (more below).

The first and only solved x-ray structure of a 12H:12L heteropolymer ferritin is that of a secreted ferritin from the cabbage looper/tiger moth *Trichoplusia ni* (Hamburger et al., 2005). The structure revealed a tetrahedral symmetry that conferred crystallization advantages over random subunit arrangements, substantiating the location of ferritin storage granules in insect fat bodies (Hamburger et al., 2005). Moreover, a tetrahedral arrangement that allowed the H and L subunits to form inter-subunit disulfide bonds provided further stability to the ferritin shell. From an evolutionary perspective, the fact that homo-polymeric ferritins are found in bacteria suggests that the earliest ferritins may have been homomeric, and that the emergence of heteromeric ferritins in animals may have been an adaptive strategy, a hypothesis reinforced by the structural resemblance between H and L chains in eukaryotes. Whether the H and L subunits in vertebrates mix randomly or assemble in a specific arrangement reflective of cellular localization and functions is currently unknown and necessitates further studies.

Here, we have employed an integrated molecular and structural analysis approach, combining protein engineering, molecular biology, statistical analysis, cryo-electron microscopy (cryo-EM), and Deep learning-based Amino acid-wise model Quality (DAQ) score to investigate the molecular architecture that imparts heteropolymer ferritins their distinctive H and L subunit arrangement, and explored how these structural intricacies contribute to the vital roles of isoferritins in cellular and systemic iron regulation. Our study presents a significant advancement in our understanding of ferritin structure and assembly mechanisms and the distinct functionality conferred by hetero-oligomeric ferritins. Unlike previous studies, which have primarily focused on homopolymer ferritins, or lacked controlled synthesis of heteropolymer ferritins, our approach involved synthesizing ferritin heteropolymers with specific H to L subunit

ratios using a novel plasmid design (Srivastava et al., 2021). Although subunit dimers appear to be crucial building blocks in the protein self-assembly process, our study uncovered a preferred association of H-L heterodimers over H-H or L-L homodimers and an unexpected random ferritin self-assembly mechanism that utilizes various pathways to build the 24mer protein structure. These new insights underscore the importance of considering the heterogeneity of ferritin populations and the impact of varying H and L subunit proportions on the biochemical, biological, and functional properties of isoferritins.

2 | MATERIALS AND METHODS

2.1 | Protein expression and purification

Recombinant human heteropolymer ferritins with different H to L subunit ratios were produced in *E. coli* Rosetta-gami B strain our engineered pWUR-FtH-TetO-FtL plasmid and different concentrations of inducers, as described in detail elsewhere (Srivastava et al., 2021; Srivastava et al., 2022; Srivastava et al., 2023). Briefly, transformed cells were induced at 37°C for 4–6 h using 10–1000 μM isopropyl β-D-1-thiogalactopyranoside (IPTG from Sigma Aldrich), 25–1600 ng/mL of anhydrotetracycline (IBA solutions). The exact concentrations of these inducers depend on the desired H to L ferritin subunit ratio (Srivastava et al., 2021). Protein purification and quantification were performed according to established protocols using size exclusion chromatography (Äkta Go, GE Healthcare), native and SDS-PAGE, and capillary electrophoresis (7100 model from Agilent Technologies) (Carmona et al., 2017; Srivastava et al., 2021; Srivastava et al., 2022). Protein concentrations were determined using the Micro BCA (bicinchoninic acid) Protein Assay Kit and a Varioskan LUX microplate reader from Thermo Fisher Scientific. Purified ferritin samples contained a small iron core (i.e., $<200 \pm 50$ Fe(III)/ferritin molecule), as determined by an iron reductive mobilization assay. We opted to use all purified ferritin samples in this study as expressed in *E. coli*, and without the removal of this small iron core to preserve the integrity of the ferritin shell and avoid oxidative damage due to otherwise potential harsh chemical treatments.

2.2 | Cryo-electron microscopy

2.2.1 | Cryo-EM sample preparation

To determine the structure of H-rich (70%H:30%L, corresponding to 17H and 7L subunit composition) and L-rich (38%H:62%L, corresponding to 9H and 15L subunit composition) heteropolymeric ferritins, single-particle cryo-EM was used. Purified ferritin samples in 50 mM MOPS, 100 mM NaCl, pH 7.4 were diluted down to 0.25–1.1 mg/mL, with or without 10 mM DTT. Holey carbon grids (R1.2/1.3 Cu 200 or 300 mesh, Quantifoil) were glow-discharged for 30–45 s at 10–20 mA using an easi-Glow (PELCO). 2–3 μL sample was then applied on each grid and plunge-frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher), with 4°C temperature, 95%–100% humidity, 6–8 s blot time, and 0 to +20 blot force. At times, grids were double-blotted to increase particle density.

2.2.2 | Cryo-EM data collection

Grids (Sample #1: 70%H:30%L; Sample #2: 38%H:62%L) were loaded either onto a Thermo Fisher instrument, a Glacios cryo transmission electron microscope equipped with Falcon 4 direct detector (Sample #1) or onto a Titan Krios G3i 300 kV electron microscope (Sample #2) equipped with a K3 Summit direct electron detector (Gatan). 4055 (Sample #1) and 10,105 (Sample #2) EER movie-mode images were collected using EPU, with super-resolution pixel size 0.95 Å/pixel, dose rate of 8.2 $e^-/Å^2/s$ and total dose of 30 $e^-/Å^2$ (Sample #1) and 0.55 Å/pixel, dose rate of 5.8 $e^-/Å^2/s$ and total dose of 42 $e^-/Å^2$ (Sample #2). Figure 3 shows the electron micrographs of the two ferritin samples employed in our study.

2.2.3 | Cryo-EM data processing

Single particle analysis was performed using the cryoSPARC software (Punjani et al., 2017) package or Relion 4.0 (Kimanius et al., 2021). All movies were motion corrected and aligned using cryoSPARC patch motion or the Relion implementation of MotionCor2. Contrast Transfer Function (CTF) estimation for the micrographs was performed using CTFFind4 (Rohou & Grigorieff, 2015).

For sample #1 (H-rich ferritin, 70%H:30%L), particles from 100 micrographs were picked using the Blob picker algorithm and 2D classified. 165 k particles from the whole data set were picked with select 2D class averages as templates and extracted as 2× binned (pixel size: 1.9 Å, box size: 120) to speed up processing and increase contrast. 2D classification and ab initio classification were used to separate high-resolution particles (Figure 4). 115 k particles with no dense iron density were selected. Selected particles were re-extracted with a pixel size of 0.95 Å and a box size of 300 pixels. The initial model was built by ab initio reconstruction. Particles were reconstructed with and without symmetry imposed using

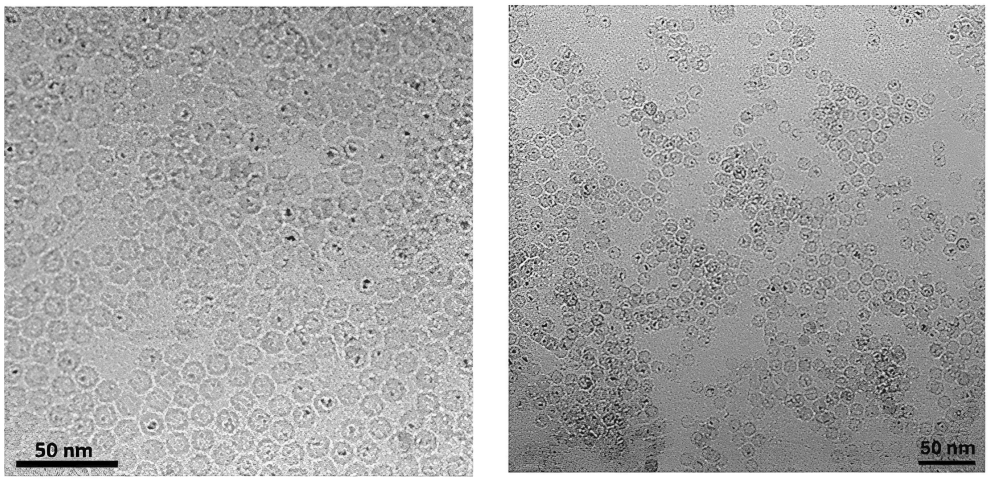

FIGURE 3 Electron micrographs of H-rich (70% H:30%L, left) and L-rich (38% H:64%L, right) ferritins.

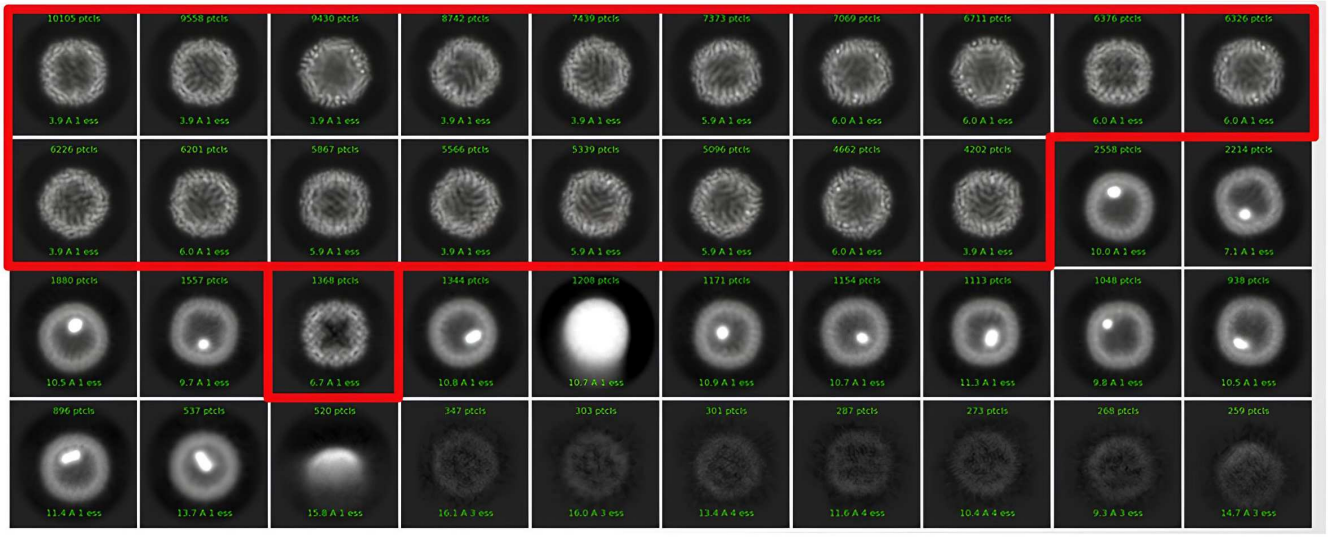

FIGURE 4 Top 40 out of 100 classes are shown as an example for the H-rich ferritin sample (70%H:30%L). Particles with no dense iron density were selected based on their 2D class averages (red box).

homogeneous refinement and generated a map at 2.1 Å resolution (with O symmetry) and 2.56 Å (without symmetry).

Similarly, for Sample #2 (L-rich ferritin, 38%H:62%L), Laplacian-of-Gaussian autopicking was carried out on a subset of 55 CTF-corrected micrographs with a range of defocus values, and the resulting 8 k particles were subjected to 2D classification. Selected 2D class averages were used as templates for reference-based picking of the entire data set. The resulting 1.6 M particles were extracted with binning four (box size 224, 4.4 Å/pix) and 2D classified. 600 k particles were chosen from 2D classes with high resolution and no iron density, and used to build an ab initio model, which was used as a reference for 3D classification. Based on multiple rounds of 3D classification, 74 k particles were chosen and re-extracted without binning (box size 256 1.1 Å/pix) for 3D refinement. A final map was obtained at 2.9 Å resolution without imposed symmetry.

2.3 | Atomic model refinement with the EM map

PDB structures of human ferritins H-chain (PDB 7jgk) and L-chain (PDB 6tsj) were used as initial atomic models. These models were then fitted to the EM map and subjected to the Rosetta Relax protocol (Wang et al., 2016) using the cryo-EM density map.

2.4 | DAQ score computation

To evaluate the local quality of atomic models from the EM maps, we used the Deep learning-based Amino acid-

wise model Quality (DAQ) score (Nakamura et al., 2023; Terashi et al., 2022; Terashi et al., 2023). DAQ assesses the likelihood of an amino acid type at a specific location in an EM map using deep learning (Maddhuri Venkata Subramaniya et al., 2019; Wang et al., 2021). The input EM map is scanned with a box size of 11 Å along the three coordinate axes with an interval of 1 Å. From the box, the trained deep neural network outputs probabilities for 20 amino acid types at the center position of the box. Then, DAQ compares the computed probability to the amino acid in the model. For an amino acid aa_i and the Cα atom position p_i of the ith amino acid residue in the model, the DAQ score is computed as:

$$\mathrm{DAQ}(i) = \log\left(\frac{P_{aa_i}(p_i)}{\sum_k P_{aa_i}(p_k)/N_{\mathrm{atom}}}\right),$$

where $P_{aa_i}(p_i)$ is the computed probability of amino-acid type aa_i at the position p_i by deep learning. The denominator normalizes the score by the average probability value among all atom positions p_k in the model. N_{atom} represents the total number of atoms in the model. If the predicted probability for p_i is above average, DAQ(i) is positive, indicating higher agreement between the amino-acid type aa_i and its local density than the average.

2.5 | Training datasets of experimental maps for the amino acid type detection

In the original DAQ score, the neural network was trained on cryo-EM maps with a resolution between 2.5 and 5.0 Å. For the H-rich and the L-rich human ferritin cryo-EM maps with 2.56 Å and 2.96 Å resolutions without symmetry, respectively, we updated the DAQ score by retraining the neural network with high-resolution maps. Initially, we meticulously curated cryo-EM maps within a resolution range of 1.0 to 3.0 Å, all of which have corresponding entries in PDB. Maps associated with PDB entries containing unresolved residues were excluded from consideration. Rigorous quality control measures were enacted, requiring a minimum cross-correlation of 0.65 and an overlap between the experimental map and a simulated map derived from the corresponding PDB entry with the same resolution.

Furthermore, a meticulous manual assessment was conducted to validate the alignment between each map and its corresponding PDB entry. In order to ensure a non-redundant data set, instances where two maps shared a protein chain pair with a sequence similarity exceeding 0.25 were deemed redundant, and one of the maps was omitted. Adhering to these exacting criteria yielded a final selection of 128 cryo-EM maps, of which 102 were allocated for training (80%) and validation (20%), while the remaining 26 were reserved for testing.

2.6 | Training procedure

The architectural framework of the model's network mirrors that of the DAQ model. Among the 102 maps in the training set, ∼0.8 million boxes were sampled from 82 maps for training, whereas an additional 0.1 million boxes were systematically gathered from 20 maps to serve as a validation resource. During each iteration of the training process, a set of 256 randomly chosen boxes were employed as input data. The training regimen encompassed a total of 30 epochs. The training was optimized by the Adam optimizer, maintaining the same learning rate of 0.002 and L2 regularization with a coefficient of $1e^{-5}$ as the DAQ model.

2.7 | Identification and subunit localization on the protein shell

While Cryo-EM represents a significant advancement in imaging science, it is not currently optimized for detecting heterogeneity in biological samples. This is because Cryo-EM involves capturing many blurred two-dimensional images of individual particles, such as proteins, from random planes. These images are then categorized into classes, with numerous particles grouped into a few classes before the three-dimensional image is reconstructed. This process produces an averaged image, posing challenges for reconstructing a three-dimensional image of particles with heterogeneous subunits, especially when the subunits appear very similar when folded, as with our heteropolymer ferritin samples.

In order to estimate the placement of the subunits, we relied on statistical analysis to determine which subunits were most likely to be of H or L type. The procedure for this estimation involved several steps. First, individual residues along each of the 24 subunits of a homogeneous ferritin (in this case, all H subunits) were assigned a DAQ score and served as a reference model. Similarly, all residues of each of the heterogeneous ferritin were assigned DAQ scores. Then, each of the 24 subunit positions was compared to the reference model (i.e., H-subunit) using two different methods: the KS_H model which utilizes the Kolmogorov–Smirnov two sample test (Smirnov, 1939) to compare empirical cumulative distribution functions of the reference and of the

heterogeneous model in question, and the TT_H model which uses a student *t* test approach (Pearson & Adyanthāya, 1929). The slight discrepancy between the two models lies in their distinct methodologies for assessing statistical significance, yielding slightly different (though insignificant) rankings of positions as likely H or L subunits, as reported in Table 1. Ultimately, L subunits were assigned within a 95% confidence interval, while the remaining subunits were assumed to be H subunits.

2.8 | Electrostatic potential calculations

For each of the H-rich and L-rich EM maps, monomer models for both H and L subunits, dimer models (HH, HL, and LL), and a comprehensive series of trimer models (H-HH, L-HH, H-HL, L-HL, HL-H, HL-L, H-LL, L-LL, and LL-L) have been constructed using PDB entries 7jgk for the H-subunit and 6tsj for the L-subunit. For each model-map pairing, the Rosetta FastRelax protocol (Alford et al., 2017; Conway et al., 2014; Nivón et al., 2013) with the EM map was applied to produce 100 relaxed models. Subsequently, an analysis focused on the change of the electrostatic interactions (Δfa_elec) among the Rosetta Energy terms was conducted. For the dimer and trimer models, the Δfa_elec at the dimer interface and at the interface between the dimer and the monomer was computed with Figures S1-S5 illustrating boxplots of the Δfa_elec distribution across the 100 constructed models. All statistical graphics and plots were generated using Python and Seaborn.

There are eight different H and L permutations that are possible in dimers, trimers, tetramers, and so forth (Figure S1), all of which led to similar overall electrostatic potentials (Figure S2). Whether H-rich or L-rich samples, there was a lack of discernible trend or notable distinctions among the electrostatic potentials of different dimers (HH, HL, and LL), indicating that all potential configurations occur with almost equal probability (Figure S2). However, clusters of 3 L-chains are least likely to form due to their low electrostatic potentials (Figure S3), consistent with our DAQ scores, computational analysis, and PyMOL models.

3 | RESULTS

3.1 | Cryo-EM structure

The cryo-EM maps of our two heteropolymer ferritin samples (Figure 5) were of sufficient resolution to allow recognition of backbones and some side chains of ferritin molecules. However, light chains and heavy chains of ferritin molecules, despite only 55% sequence identity, are very similar in secondary structure (Figure 5b). Therefore, the alignment of the EM data was dominated by backbone densities. Side-chain densities of light chains and heavy chains were not separable and were averaged out in the EM map, and thus it was not possible to study the distribution of the two chains directly from the EM structure. Figure 5 displays the EM maps of H-rich ferritin and L-rich ferritin samples with and without symmetry, used to build the ferritin models.

3.2 | DAQ scores

Figure 6 displays the raw DAQ scores of two H/L homopolymer (100%H:0%L; and 0%H:100%L) and three H/L heteropolymer EM-maps (38%H:62%L; 70%H:30%L; and 90%H:10%L) in comparison to the published H-homopolymer EM-map. In this diagram, blue represents a high DAQ score, indicating that the model aligns with the map, while red represents a low DAQ score, indicating a disagreement with the map. Compared to the H-model, the 38%H:62%L map exhibits more red positions than the 90%H:10%L map, indicating weak alignment or high disagreement between the DAQ analysis and the H-model, as expected. Notably, the average DAQ score among all residues for each ferritin model strongly correlates with the proportion of H subunits present in each of the EM maps.

3.3 | Generation of ferritin models

The DAQ scoring system is trained through deep learning techniques (Nakamura et al., 2023) and measures how well the electron density of an amino acid matches local features of the protein cryo-EM structure, using the H-homopolymer as our base. Models of assembled H and L subunits were then generated using these DAQ scores and the protein electron density maps to predict whether an amino acid belongs to an H or an L subunit, and the likelihood that any given subunit is an H or an L subunit. In the case of homopolymer ferritin containing only H or L subunits, the DAQ scoring system was applied after reducing the resolution through Gaussian blurring, to align better with the resolution utilized during the training of the DAQ scoring system. For heteropolymers, the estimation of L subunit positions involved comparing the DAQ scores of each chain using a Kolmogorov–Smirnov test, a 2-sample distribution test, against the corresponding positions of the H homopolymer model.

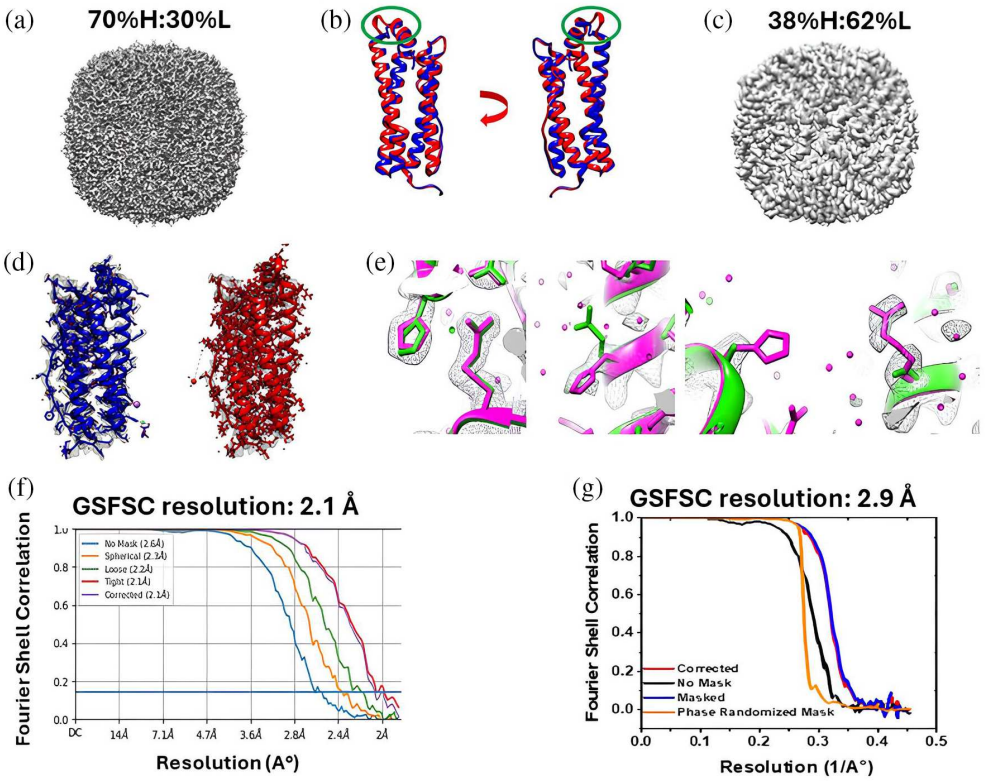

FIGURE 5 Cryo-EM reconstruction of (a) H-rich ferritin (70%H:30%L) with symmetry imposed, and of (c) L-rich ferritin (38%H:62%L) without symmetry for modeling. (b) Overlay of published monomeric structures of ferritin heavy H-chain (blue, PDB 7jgk) and light L-chain (red, PDB 6tsj) with the main location of structural divergence at the (D–E) loop circled in green. (d) Cryo-EM map of one subunit, fitted with H-chain model (blue) or L-chain model (red). (e) Example of side-chain densities at 2.1 Å resolution of the H-rich ferritin sample showing hybrid densities of both heavy chain (magenta) and light chain (green). (f, g) Gold standard fourier shell correlation (GSFSC) curves for H-rich (f) and L-rich (g) samples.

To create distribution curves, we assumed that the composition of the heteropolymer ferritin sample is centered around its average H and L composition, with up to 20% variation in subunit composition, as determined by size exclusion column chromatography following sample fractionation (Srivastava et al., 2023), and that the distribution curves are randomly distributed according to the discrete binomial distribution. Each of the 24 subunits heteropolymer is regarded as a random trial, with the parameter (n) set to 24, and the probability of success (p) as the fractionation average of the H subunit. For instance, if the sample averaged 70% H, then (p) was set to 0.7. To create a smoothed curve rather than a simple histogram, the kernel density estimation technique was employed.

Figure 7 displays the distribution curves for two heteropolymer ferritin samples, an H-rich sample at 70%H:30%L (or 17H:7 L, on average) and an L-rich sample at 38%H:62%L (or 9H:15 L, on average).

Using the H-model and the computed DAQ scores (red and blue, Figure 6), we estimated the positions of the H and L-chains in the two heteropolymer ferritin samples (the H-rich ferritin at 70%H:30%L or 17H:7L, and the L-rich ferritin at 38%H:62%L or 9H:15L), using the Kolmogorov–Smirnov two-sample test, as depicted in Figure 8.

3.4 | Model validation

To compute the probabilities of encountering HH, HL, or LL pairings assuming they occur entirely at random, a probability tree depicting all potential outcomes is initially constructed and is depicted in Figure 9. We chose four representative samples, two L-rich (11H:13L and 7H:17L) and two H-rich (19H:5L and 15H:9L) samples, representing the 20% variation around the average H/L subunit ratio, determined experimentally (Srivastava et al., 2023). Starting with 24 subunits at the root of the tree, Figure 9 displays possible paths when two subunits are randomly removed with each lower level, and the number of possibilities following the removal of each pair of subunits, leading to a path where all H subunits are

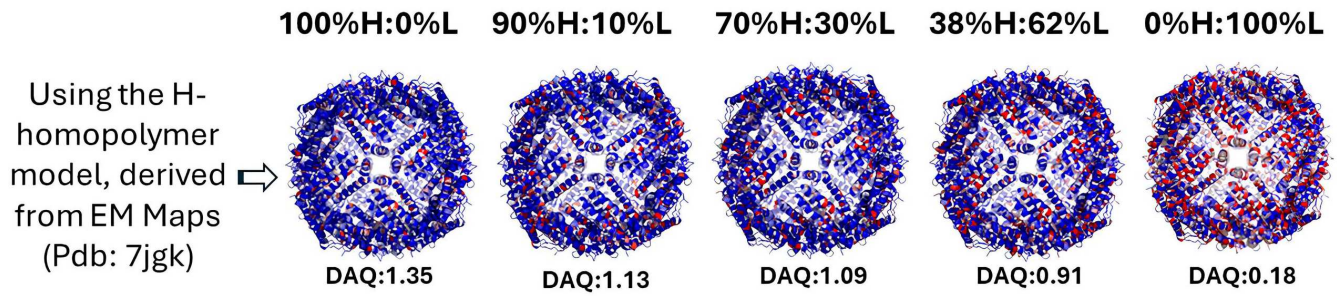

FIGURE 6 Comparative analysis of raw DAQ scores for various ferritin EM-maps with different H and L compositions. DAQ scores for each residue are colored red when DAQ score < -1.0 and blue for DAQ score >1.0. Blue indicates high DAQ scores reflecting model-map alignment, while red denotes low DAQ scores indicating disagreement. The DAQ value below each EM map represents the average DAQ score.

eliminated, thus restricting the total number of potential paths.

At each step moving down the tree, there is a probability of removing a pair of H subunits (denoted as p_{HH}), a pair of L subunits (p_{LL}) or a mixed H and L pair (p_{HL}). Assuming n_H represents the number of H subunits, and n_L the number of L subunits, the p_{HH} and p_{HL} probabilities are given by:

$$p_{HH} = \frac{\frac{n_H(n_H-1)}{2}}{n_H n_L + \left(\frac{n_H(n_H-1)}{2}\right) + \left(\frac{n_L(n_L-1)}{2}\right)},$$

$$p_{LL} = \frac{\frac{n_L(n_L-1)}{2}}{n_H n_L + \left(\frac{n_H(n_H-1)}{2}\right) + \left(\frac{n_L(n_L-1)}{2}\right)},$$

$$p_{HL} = \frac{n_H n_L}{n_H n_L + \left(\frac{n_H(n_H-1)}{2}\right) + \left(\frac{n_L(n_L-1)}{2}\right)}.$$

Using the example shown in Figure 9, where $n_H = 11$ and $n_L = 13$, the probability of removing a pair containing H and L is,

$$p_{HL} = \frac{143}{143+55+78} \approx 0.52$$

or about a 52% chance, if the pairs are removed at random.

To compute the probability of finding an exact number of H/L pairs (including H/H or L/L pairs), one can determine the total number of potential paths by following a probability tree from its origin to its end, ensuring a consistent reduction of two subunits (either 2H, 2L, or one of each) at each step until reaching the final layer of the tree. For example, one valid path for the 11H:13L sample in Figure 9 is highlighted in red until 1H:1L is reached. At the conclusion of this process, there are roughly 3^{12} distinct paths that could result in one of three final pairings, a 1H:1L heterodimer, a 1H:1H homodimer, or a 1L:1L homodimer pair. In all cases, our models suggest a preference for H-L heterodimer formation over either H-H or L-L homodimers. In terms of model validation, an important question arises: what is the likelihood of the model being generated by random chance? To address this, we used the aforementioned framework to ascertain the probability of encountering the same number of H-L heterodimers observed in the model. Interestingly, for many of the scenarios, the probability of randomly obtaining the observed number of H-L heterodimers was significantly lower than that of flipping a coin, and both the KS_H and the TT_H models corroborate this preference (Table 2). Considering the concurrence between both models and the unlikelihood of such models arising by random chance in many instances, we believe that our models likely represent genuine structures.

3.5 | Probabilities of H/L heterodimer formation at random

In all ferritin models, a high preference for H/L heterodimer formation is observed, regardless of the likelihood of such heterodimers forming through random selection from the pool of subunits. The findings presented in Table 2 provide evidence that the KS and TT models did not merely allocate subunits randomly, as might be anticipated, especially in scenarios where the ratio of H to L approaches 1. In a model where dimers are randomly assigned, fewer H/L heterodimers would be expected; however, the data in Table 2 indicate otherwise, suggesting a true preference for H/L heterodimer formation in these ferritin models.

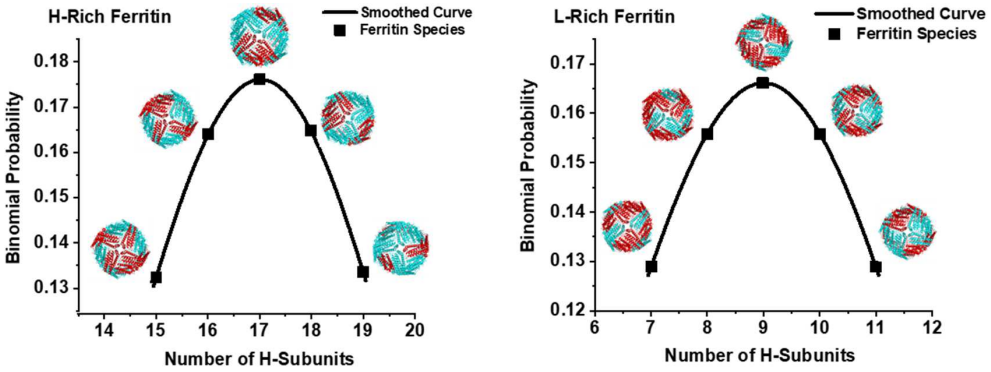

FIGURE 7 Distribution curves of ferritin species with their average H and L subunits composition displayed at the peak of the curve, extending to +/− 20% variation on each side of the curve (i.e., +/− up to 2H subunits/protein). (Left) H-rich ferritin (70%H:30%L or 17H:7L), (Right) L-rich ferritin (38%H:62%L or 9H:15L). The smoothed curves were generated using the kernel density estimation technique. The H-monomers are depicted in cyan while the L-monomers in red.

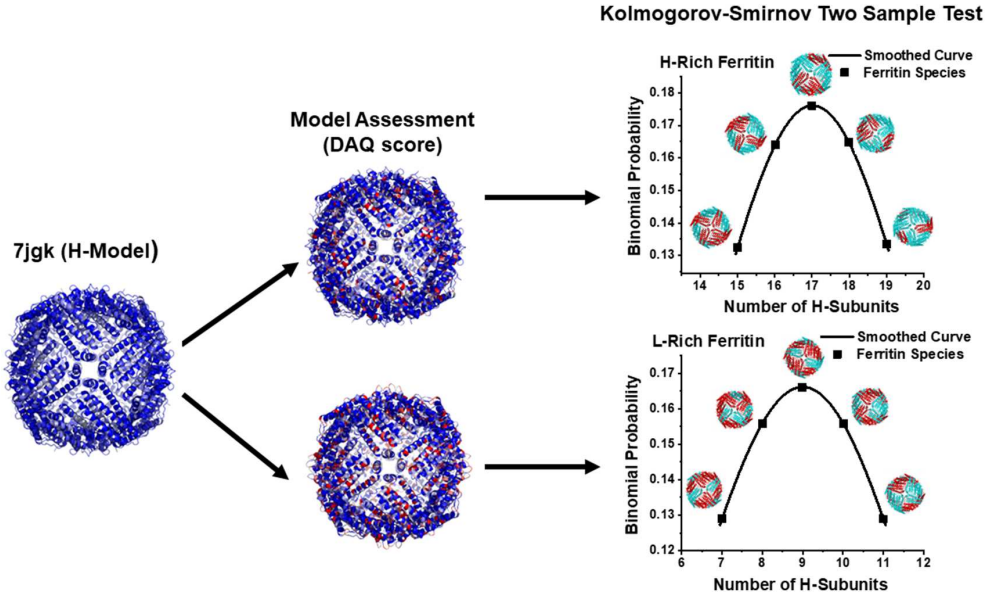

FIGURE 8 Depiction of the predicted H- and L-chain distribution in two heteropolymer ferritin samples, an H-rich ferritin (70%H:30%L or 17H:7L) and an L-rich ferritin (38%H:62%L or 9H:15L) using the DAQ score and the Kolmogorov–Smirnov two-sample test.

3.6 | Ferritin self-assembly: Proposed mechanism

While some studies have identified ferritin dimers as crucial intermediates in the protein assembly process, supported by their stability and prevalence across ferritins from diverse origins (Andrews et al., 1993; Fan et al., 2009; Gerl & Jaenicke, 1987; Kilic et al., 2003; Kumar et al., 2021; Stefanini et al., 1987), the existence of folded monomers prompts inquiry into whether ferritin self-assembly is initiated solely by dimers, or if it might also involve odd-numbered intermediates such as trimers. Our results strongly suggest a pronounced tendency for H and L chains to interact preferentially, although this interaction is not exclusive and does not preclude the involvement of monomers in the assembly process, with both H-rich and L-rich ferritin self-assembly mechanisms, demonstrating a notable preference for H-L heterodimer formation over H-H or L-L homodimers. Figure 10 illustrates one ferritin self-assembly mechanism, although alternative mechanisms with different H and L conformations exist, leading to the same final 24-mer structures. Although H-L heterodimers are predominantly formed, H-H and L-L homodimers do also occur, as indicated by DAQ scores and the information presented in Table 1, leading to the formation of heterogeneous ferritin structures with various compositions of H and L subunits. To better visualize the self-assembled H/L heteropolymer ferritins (both H-rich and L-rich) and the arrangement of the H and L subunits around the ferritin 24-mer nanostructure, 10 different (but not exclusive) models have been constructed using PyMOL (The PyMOL Molecular Graphics System, https://www.pymol.org) and the data in Table 1 (see Data S1).

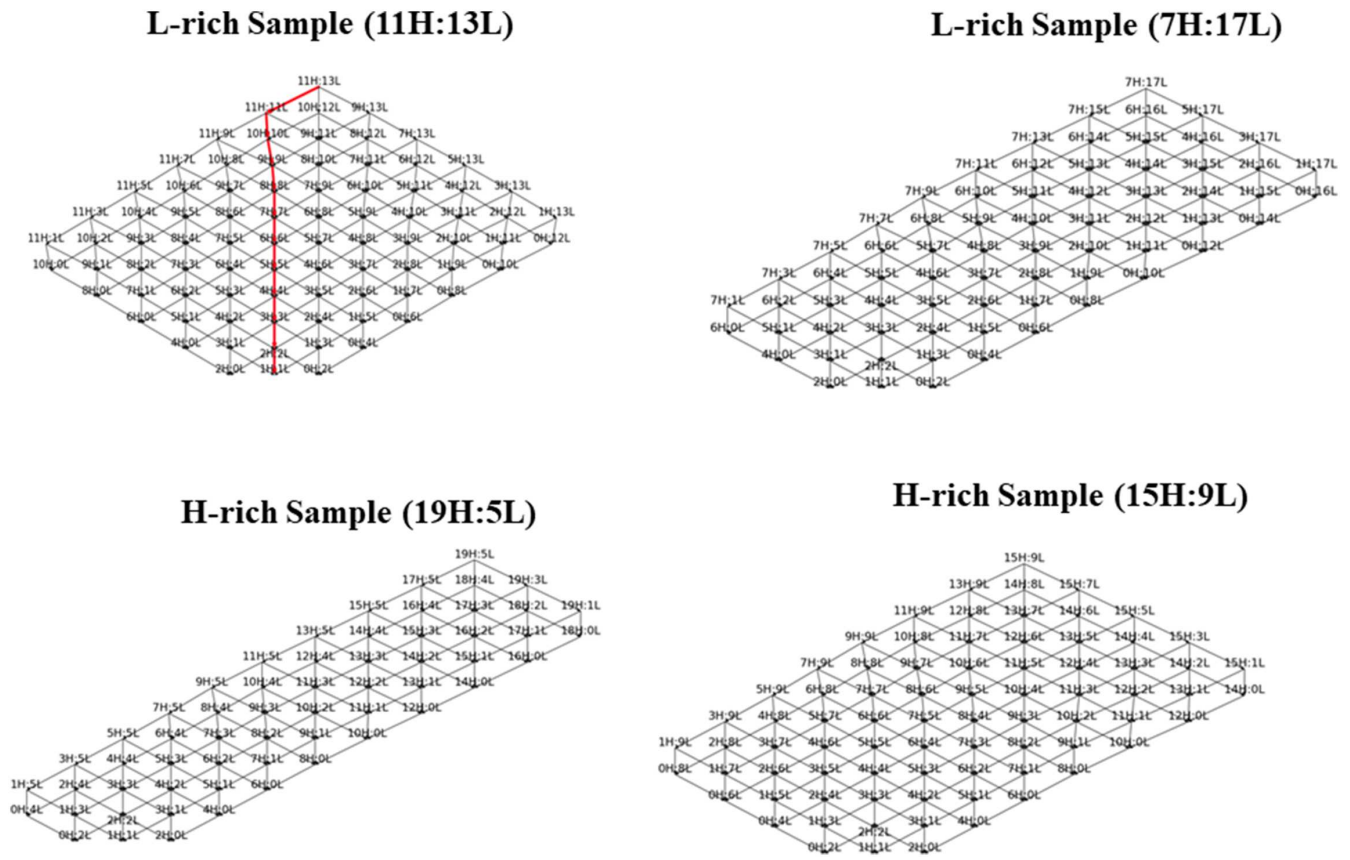

FIGURE 9 Probability tree illustrating every possible outcome originating from the specified H:L ratio representing +/− 20% variation from the parent H-rich (17L:7H) and the L-rich (9H:15L) heteropolymer ferritin samples. One valid path for the L-rich Sample (11H:13L) is highlighted in red until 1H:1L is reached (See text for more details).

Table 1 shows the H and L subunit configuration based on DAQ scores and the KS_H and TT_H statistical models. The predicted ferritin models were then visualized by PyMOL to determine their 3D structures. In the H-rich sample, none of the 5 H/L heteropolymers examined (i.e., 15H:9L, 16H:8L, 17H:7L, 18H:6L, and 19H:5L) showed threefold or fourfold channels composed entirely of L subunits, mirroring the findings in the 5 H/L heteropolymers identified in the L-rich sample (i.e., 7H:17L, 8H:16L, 9H:15L, 10H:14L, and 11H:13L) which also lacked three or fourfold H channels, suggesting the preferential formation of channels with mixed H and L chains (Figure S1). An exception was noted in one of the 10 analyzed samples (i.e., L-rich sample, 11H:13L) where a fourfold H channel was observed, but *only* in the TT_H model.

4 | DISCUSSION

4.1 | Ferritin assembly pathways

Over the past two decades, structural biology has experienced remarkable advancements in two key areas: cryo-electron microscopy (cryo-EM) and predictive modeling. While cryo-EM has progressed from low-resolution images to near-atomic resolution views of diverse macromolecular complexes, predictive modeling has emerged as a powerful tool, capable of accurately forecasting protein structures directly from their amino acid sequences, spearheaded by developments in machine learning and tools like AlphaFold (Corum et al., 2024). When combined, these technologies exhibit a synergistic potential that facilitates the creation of intricate models of macromolecular structures and a richer understanding of their functions and dynamics. In this study, we employ an integrated molecular and structural biology approach, encompassing cryo-EM, DAQ score, and statistical analysis to explore the molecular architecture and subunit distribution of heteropolymer ferritin samples with distinct H and L subunit ratios.

Ferritins exhibit a unique ability to spontaneously form globular structures via interactions between subunits. Although different intermediates have been suggested, the self-assembly process always results in a 24-subunit nanocage structure (Sudarev et al., 2023), or a 12-mer complex in the case of Dps (DNA binding proteins from starved cells) (Ilari et al., 2000), that is highly

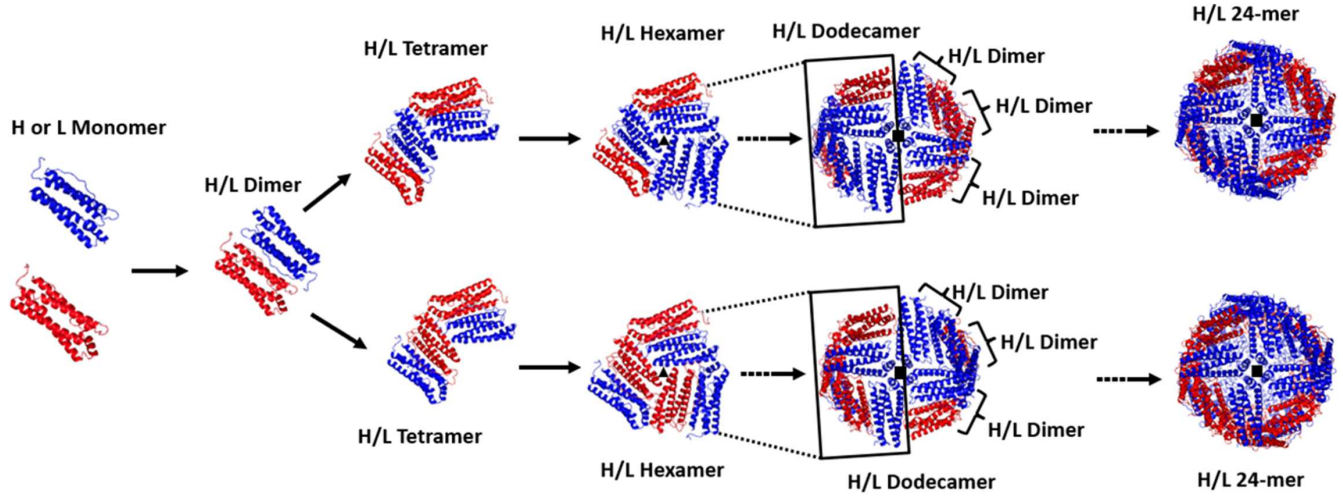

FIGURE 10 Example illustration of an H-rich heteropolymer ferritin (15H:9L) self-assembly mechanism. This ferritin type is similar to that found in the heart or the brain. Alternative mechanisms and conformations exist, all leading to the same final 24-mer structures, in accordance with the data in Table 1. In this diagram, H-monomers are depicted in blue, while L-monomers are in red. Each progression in this self-assembly process builds upon the previous step by incorporating new H-L dimers, maintaining the same number of threefold or fourfold channels identified in Table 1. Dotted arrows in steps four and five represent multiple consecutive additions of HH, LL, or HL dimers to achieve the depicted structures.

stable and resilient even under adverse conditions (Bou-Abdallah, 2010). Despite the significance of the ferritin self-assembly process and its impact on protein functionality, the precise molecular mechanisms governing self-assembly remain elusive. Conflicting models for self-assembly have been proposed, with differing views on the intermediates involved and their structural arrangements (Ilari et al., 2000; Sato et al., 2016; Stefanini et al., 1987; Sudarev et al., 2023), and the role of dimers as a starting point in ferritin self-assembly remains debated. Some studies suggested dimers as stable intermediates while others question their necessity, proposing the potential involvement of monomers at various stages of the process (Andrews et al., 1993; Fan et al., 2009; Gerl & Jaenicke, 1987; Ilari et al., 2000; Kilic et al., 2003; Kumar et al., 2021; Pulsipher et al., 2017; Sana et al., 2013; Stefanini et al., 1987). Other studies showed a random reconstitution process between ferritin subunits, whether from the same species or from distinct species, highlighting the consistent nature of inter-subunit interaction domains that shape the highly specific quaternary structure of ferritin despite variations in amino acid composition of the subunits (Otsuka et al., 1980). Notably, members of the sub-family of ferritins, encapsulated ferritins (EncFtn), showed dynamic structures of multiple even-numbered oligomerization states (dimer, tetramer, hexamer, and decamer), implying that higher order oligomers are formed by the association of dimer units (He et al., 2019; Ross et al., 2020). High-speed atomic force microscopy and molecular dynamic simulations along with optical nanotweezer experiments have elucidated the real-time disassembly mechanism of ferritin and its fragmentation into dimers. The results revealed a stepwise breakdown mechanism into major fragments corresponding to 22-mer, 12-mer, tetramer (4-mer), and dimer (2-mer) subunits, thereby corroborating the crucial involvement of dimers in ferritin's self-assembly and disassembly processes (Yousefi et al., n.d.; Maity et al., 2020). The observation of heteropolymer ferritin microheterogeneity in various animal tissues, including recent findings in recombinantly expressed human heteropolymer ferritin (Srivastava et al., 2023) indicates that ferritin microheterogeneity is a widespread phenomenon across species and suggests that the mechanisms driving ferritin assembly and heterogeneity are intrinsic to the protein itself (and not the source) and its intricate assembly process. Furthermore, the concept of heterodimer over homodimer formation has been explored in various other proteins and receptors (Ayoub et al., 2004; Beck et al., 2024; Lv et al., 2021; Yin et al., 2023), suggesting a broader biological principle that would also likely apply to the assembly and functionality of ferritin heterodimers. Once synthesized, ferritin subunits move randomly in solution due to Brownian motion, and their frequency of collisions is proportional to their concentrations and their ability to diffuse. The preferential formation of H-L heterodimers over H-H or L-L homodimers in ferritin, and the different pathways through which the 24-mer ferritin assembly mechanism can proceed (Figure 10) underscores a complex interplay between structural

TABLE 1 Subunit configurations of different heteropolymer samples, determined by DAQ scores and statistical analysis. Most if not all of the L chains in the H-rich sample, and the H chains in the L-rich sample form heterodimers.

H-rich ferritin 70% H:30%L (17H:7L)							
KS_H model	**L-H dimers**	**H-H dimers**	**L-L dimers**	**Threefold H-channels**	**Threefold L-channels**	**Fourfold H-channels**	**Fourfold L-channels**
15H:9L	9	3	0	2	0	2	0
16H:8L	8	4	0	2	0	2	0
17H:7L	7	5	0	3	0	2	0
18H:6L	6	6	0	3	0	2	0
19H:5L	5	7	0	4	0	2	0
TT_H model							
15H:9L	7	4	1	3	0	1	0
16H:8L	6	5	1	3	0	1	0
17H:7L	5	6	1	3	0	2	0
18H:6L	4	7	1	4	0	2	0
19H:5L	3	8	1	4	0	2	0
L-rich ferritin 38% H:62%L (9H:15L)							
KS_H model	**H-L dimers**	**H-H dimers**	**L-L dimers**	**Threefold L-channels**	**Threefold H-channels**	**Fourfold L-channels**	**Fourfold H-channels**
7H:17L	5	1	6	2	0	1	0
8H:16L	6	1	5	1	0	1	0
9H:15L	7	1	4	1	0	1	0
10H:14L	8	1	3	1	0	1	0
11H:13L	9	1	2	1	0	1	0
TT_H model							
7H:17L	7	0	5	2	0	0	0
8H:16L	8	0	4	2	0	0	0
9H:15L	9	0	3	2	0	0	0
10H:14L	10	0	2	2	0	0	0
11H:13L	11	0	1	2	0	0	1

features, functional requirements, and regulatory mechanisms that govern ferritin assembly and function. This can be understood through several molecular and biochemical principles, including charge distribution, hydrophobic patches, thermodynamic stability, and functional requirements (Butler & Parker, 1995; Keppler et al., 2015; Srivastava et al., 2022; Srivastava et al., 2023). These factors collectively enhance the affinity of H and L subunits for each other, optimizing ferritin's ability to store and release iron efficiently. For instance, the distinct isoelectric points (pIs) between H and L subunits (5.8 and 6.8, respectively) can lead to electrostatic attractions between the subunits, enhancing heterodimer formation (Srivastava et al., 2023). Electrostatic interactions were also the likely reason for the assembly of hybrid frog ferritin variants due to surface charge modification and thus different PI values (Subhadarshanee et al., 2017).

A recent investigation from our lab found a remarkable hyperthermostability of heteropolymer ferritin, over homopolymers, that was attributable to an extensive cooperative network of inter- and intra-subunit interactions (i.e., hydrophobic, electrostatic, hydrogen bonds, and salt bridges) and to secondary structural composition (Srivastava et al., 2022). However, despite advances in imaging and structural biology techniques, and our general understanding of self-assembly processes (Arosio et al., 2009; Li et al., 2022; O'Shea et al., 1989), uncertainties persist regarding the specific pathways and intermediates involved in ferritin self-assembly and nanocage formation, largely because of the inherent heterogeneity,

TABLE 2 The likelihood that H/L heterodimer formation occurs randomly. Entries correspond to the probabilities of obtaining exactly the number of H/L heterodimers found in each of the ferritin heteropolymer models derived from the Kolmogorov–Smirnov (KS) test and the student *t* test (TT). The computed probabilities assume that all pairs of subunits are selected uniformly at random from the available pool of 24 subunits, ranging from a coin flip (~ 50%) to highly improbable (<1%).

Ferritin models	Number of computed H/L heterodimers	Probability (%) of finding H/L heterodimers at random
H-rich ferritin 70% H:30%L (17H:7L)		
KS_H model		
19H:5L	5	59
18H:6L	6	44
17H:7L	7	29
16H:8L	8	17
15H:9L	9	9
TT_H model		
19H:5L	3	37
18H:6L	4	47
17H:7L	5	51
16H:8L	6	48
15H:9L	7	39
L-rich ferritin 38%H:62%L (9H:15L)		
KS_H model		
11H:13L	9	14
10H:14L	8	26
9H:15L	7	39
8H:16L	6	48
7H:17L	5	51
TT_H model		
11H:13L	11	1
10H:14L	10	3
9H:15L	9	9
8H:16L	8	17
7H:17L	7	29

complexity, and randomness of the assembly pathways (Figure 10). Nonetheless, our findings indicate that interactions at dimer-monomer and dimer-dimer interfaces (but not at C2 interfaces) may play crucial roles in ferritin heteropolymer assembly (Figure S4). This is consistent with recent research highlighting the importance of hydrogen bond formations at various dimer interfaces for ferritin stability and structural integrity, with the N- and C-termini of each subunit being pivotal in influencing conformational changes in the threefold and fourfold channels (Bradley et al., 2016). The significance and roles of the C3 and C4 interfaces (Figure 1), especially in heteropolymer ferritins, remain largely unexplored, although early mutagenesis studies in homopolymer H-ferritin (Santambrogio et al., 1997) revealed that the reassembly pathway involves the coalescence of H-homodimers around the three and fourfold axes, followed by stepwise addition of dimers until the 24-mer ferritin nanostructure is assembled, a result consistent with the data of Table 1.

It is important to mention that the differential synthesis rates of H and L ferritin chains can thermodynamically direct the formation of heteropolymers by influencing the availability and interactions of these chains, thereby determining the structural and functional composition of the ferritin assembled and ensuring the most suitable ferritin structures for the physiological needs of the cell. If one type of chain is synthesized more rapidly or in greater quantities, it becomes statistically more likely to be incorporated into forming ferritins. For instance, a higher synthesis rate of H chains would lead to a relative abundance of H chains available for ferritin assembly, promoting the formation of more H-containing heteropolymers. Additionally, the preferential formation and incorporation of H-L dimers into ferritin structures can be explained through simple physical arguments related to the dimensions and mass of H and L monomers. Both monomers are cylindrical and similar in size, but L monomers are about 10% lighter than H monomers, potentially moving about 5% faster (kinetic energy = $\frac{1}{2}mv^2$). This difference in velocity could enhance the likelihood of H-L dimer formation, particularly in H-rich environments, due to more frequent H-L interactions compared to L-L interactions. This pattern supports the formation of more H-L dimers even in L-rich conditions, aligning with the models and the data in Table 1. Postdimer formation, the slower-moving H-rich dimers (H-H or H-L) are more likely to interact, suggesting fewer L-L dimers in the final ferritin structure, especially in H-rich scenarios, a trend also reflected in the data from Table 1.

4.2 | Physiological implication

Discussion on the observed heterogeneity of ferritin isoforms in tissues prompted speculation on the underlying physiological implications, although such conjecture is

based partly on in vitro ferritin subunit assembly results (Srivastava et al., 2023). The preferential formation of H/L hybrid ferritin heteropolymers from denatured H- and L-homopolymers explains why H- and L-homopolymers are poorly populated in mammalian tissues, suggesting a recognition specificity between H and L subunits over H-H or L-L ferritin homopolymers (Carmona et al., 2017; Santambrogio et al., 1993). Because of the narrow distribution and similarity in the average composition of H and L subunits in tissue ferritin (Srivastava et al., 2023), a completely random interaction between H and L subunits is improbable, otherwise, it would have resulted in a wider array of isoferritin species, spanning from 24H:0L to 0H:24L and various compositions in between. Although the precise biological roles of heteropolymer ferritins remain unclear, their tissue distribution seems to correspond with their iron storage capacity and antioxidant properties by producing ferritins with specific subunit compositions. For example, tissues like the liver and spleen, with high iron storage needs, may produce ferritins with more L subunits for safe iron storage. Conversely, tissues with lower iron storage requirements, such as the heart and brain, may produce ferritins with more H subunits for rapid iron uptake to protect against oxidative damage. The coexistence of H- and L-rich ferritins in cells (i.e., ferritin microheterogeneity) allows them to adapt to their environments, maintain iron balance, protect against oxidative damage, and support tissue-specific functions (Srivastava et al., 2023). It has been shown that H-rich ferritin regenerated full ferroxidase activity more rapidly than H-homopolymer did and additionally formed larger iron cores, indicating dual roles for the L subunit in facilitating iron turnover at the dinuclear center of the H chain and in mineralization of the iron core. The L-rich ferritin, while also facilitating iron oxidation at the ferroxidase centers of the H chains present on the protein, additionally promoted oxidation at the mineral surface once the iron binding capacity of the ferroxidase center was exceeded, reinforcing the synergistic and complementary roles that H and L chains play in ferritin function and distribution in cells (Mehlenbacher et al., 2017).

Despite the understanding that ferritin lysosomal degradation involves the exclusive interaction of the nuclear receptor coactivator-4 (NCOA4) with the H subunit through a process called ferritinophagy (Dowdle et al., 2014; Galy et al., 2024; Mancias et al., 2014), the influence of ferritin's subunit composition on ferritin turnover during iron deficiency and the mechanism behind NCOA4-ferritin condensate formation are yet to be fully elucidated. Additionally, the role of the ferritin L subunit in recognition events and in NCOA4-mediated ferritin incorporation into cytosolic condensates remains unclear. However, it is possible that either the ferritin subunit composition, the size of the ferritin iron core, and/or the chemical nature and structure of the NCOA4 protein (i.e., Fe-S cluster-bound NCOA4 vs. Fe-S cluster-free NCOA4, and soluble vs. insoluble NCOA4 condensate vs. monomeric/dimeric NCOA4) (Kuno & Iwai, 2023; Lu et al., 2024) could influence the formation and fate of the ferritin-NCOA4 complex, but more research is needed to better understand its trafficking route, and the efficiency of condensate removal from lysosomes.

Lastly, the role of H and L subunits in ferritin is not limited to iron storage. H subunit-rich ferritins are suggested to permit more dynamic intracellular traffic and faster iron accumulation and release than L-subunit-rich ferritins (Harned et al., 2010; Koorts et al., 2012). It has been shown that cells-specific ferritin type and their H/L composition is critical for maintaining cellular homeostasis and appropriate level of the antioxidant GSH, and thus protection against iron-catalyzed free radical reactions. Alteration of ferritin subunit synthesis and ultimately the H/L ratio significantly changes intracellular iron dynamics and the availability of the labile iron pool (Koorts et al., 2011). Furthermore, the differential regulation of ferritin H and L subunit mRNA during inflammation and long-term iron overload suggests distinct roles for H and L subunits in response to physiological and pathological conditions, which could be exploited for therapeutic interventions targeting iron-related disorders (Leggett et al., 1993). Compared to healthy tissues, isoferritins from diseased tissues revealed a more acidic ferritin profile due to the presence of a higher number of H-subunits. This observed microheterogeneity could be due to various cellular stressors such as variations in iron levels and oxidative stress (Torti & Torti, 2002). Other factors such as hypoxic conditions regulate the synthesis of L-ferritin via the IRE/IRP signaling pathway, suggesting that H and L subunits can react independently to changes in cellular iron concentrations, further influencing how a cell might direct the assembly of a specific ferritin isoform (Sammarco et al., 2008). Thus, understanding the heteropolymer ferritin structure–function relationship, the multifaceted roles of H and L subunits in iron storage and cellular trafficking, their differential regulation in response to physiological and pathological conditions, and their assembly mechanism is pivotal for unraveling the complexities of iron homeostasis and exploiting ferritin in bio-nanotechnology.

AUTHOR CONTRIBUTIONS

Fadi Bou-Abdallah: Conceptualization; investigation; funding acquisition; writing – original draft; methodology; writing – review and editing; supervision; formal

analysis; data curation. **Jeremie Fish:** Data curation; methodology; validation; formal analysis; writing – review and editing; investigation. **Genki Terashi:** Formal analysis; investigation; methodology; writing – review and editing; data curation. **Yuanyuan Zhang:** Investigation; methodology; formal analysis. **Daisuke Kihara:** Funding acquisition; writing – review and editing; investigation; supervision; data curation; methodology; formal analysis. **Paolo Arosio:** Conceptualization; investigation; methodology; writing – review and editing; formal analysis.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation, Division of Molecular and Cellular Biosciences (MCB) award 1934666 (F. B.-A.), a Research Corporation for Science Advancement,Cottrell Instrumentation Supplements Award #27452 (F. B.-A.), the National Institutes of Health (R01GM133840) (D. K.) and the National Science Foundation (DBI2003635, DBI2146026, IIS2211598, DMS2151678, and CMMI1825941) (D. K.). The authors wish to thank Dr. Ma Boxue, an application scientist at Thermo Fisher Scientific, Dr. Vera Moiseenkova-Bell, faculty director of the Beckman Center for Cryo-Electron Microscopy, Dr. Stefan Steimle and Dr. Elaine Mihelc at the University of Pennsylvania for Cryo-EM data collection and technical support, and Dr. Ayush K. Srivastava for his assistance in the purification of the heteropolymer ferritin samples employed in this study.

ORCID

Fadi Bou-Abdallah https://orcid.org/0000-0002-8557-1827

REFERENCES

Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta all-atom energy function for macromolecular modeling and design. J Chem Theory Comput. 2017; 13(6):3031–48.

Andrews SC, Smith JM, Hawkins C, Williams JM, Harrison PM, Guest JR. Overproduction, purification, and characterization of the bacterioferritin of *Escherichia coli* and a C-terminally extended variant. Eur J Biochem. 1993;213:329–38.

Arosio P, Ingrassia R, Cavadini P. Ferritins: a family of molecules for iron storage, antioxidation and more. Biochim Biophys Acta, Gen Subj. 2009;1790(7):589–99.

Ayoub M, Levoye A, Delagrange P, Jockers R. Preferential formation of MT1/MT2 melatonin receptor heterodimers with distinct ligand interaction properties compared with MT2 homodimers. Mol Pharmacol. 2004;66(2):312–21.

Beck M, Covino R, Hänelt I, Müller-McNicoll M. Understanding the cell: future views of structural biology. Cell. 2024;187(3): 545–62.

Bou-Abdallah F. The iron redox and hydrolysis chemistry of the ferritins. Biochim Biophys Acta, Gen Subj. 2010;1800(8):719–31.

Bou-Abdallah F, Santambrogio P, Levi S, Arosio P, Chasteen ND. Unique iron binding and oxidation properties of human mitochondrial ferritin: a comparative analysis with human H-chain ferritin. J Mol Biol. 2005;347(3):543–54.

Bradley JM, Le Brun NE, Moore GR. Ferritins: furnishing proteins with iron. J Biol Inorg Chem. 2016;21(1):13–28.

Butler A, Parker M. COUP-TF II homodimers are formed in preference to heterodimers with RXRα or TRβ in intact cells. Nucleic Acids Res. 1995;23(20):4143–9.

Carmona F, Poli M, Bertuzzi M, Gianoncelli A, Gangemi F, Arosio P. Study of ferritin self-assembly and heteropolymer formation by the use of fluorescence resonance energy transfer (FRET) technology. Biochim Biophys Acta, Gen Subj. 2017; 1861(3):522–32.

Chasteen ND. Ferritin. Uptake, storage, and release of iron. Met Ions Biol Syst. 1998;35:479.

Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D. Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci. 2014;23:47–55.

Corum MR, Venkannagari H, Hryc CF, Baker ML. Predictive modeling and cryo-EM: a synergistic approach to modeling macromolecular structure. Biophys J. 2024;123:435–50.

Dowdle WE, Nyfeler B, Nagel J, Elling RA, Liu S, Triantafellow E, et al. Selective VPS34 inhibitor blocks autophagy and uncovers a role for NCOA4 in ferritin degradation and iron homeostasis in vivo. Nat Cell Biol. 2014;16:1069–79.

Fan R, Boyle AL, Vee VC, See LN, Orner BP. A helix swapping study of two protein cages. Biochemistry. 2009;48:5623–30.

Galy B, Conrad M, Muckenthaler M. Mechanisms controlling cellular and systemic iron homeostasis. Nat Rev Mol Cell Biol. 2024; 25:133–55.

Gerl M, Jaenicke R. Mechanism of the self-assembly of apoferritin from horse spleen - cross-linking and spectroscopic analysis. Eur Biophys J. 1987;15:103–9.

Gerl M, Jaenicke R, Smith JM, Harrison PM. Self-assembly of apoferritin from horse spleen after reversible chemical modification with 2,3-dimethylmaleic anhydride. Biochemistry. 1988;27: 4089–96.

Hamburger AE, West AP Jr, Hamburger ZA, Hamburger P, Bjorkman PJ. Crystal structure of a secreted insect ferritin reveals a symmetrical arrangement of heavy and light chains. J Mol Biol. 2005;349:558–69.

Harned J, Ferrell J, Lall MM, Fleisher LN, Nagar S, Goralska M, et al. Altered ferritin subunit composition: change in iron metabolism in lens epithelial cells and downstream effects on glutathione levels and VEGF secretion. Invest Ophthalmol Vis Sci. 2010;51(9):4437–46.

Harrison PM, Arosio P. The ferritins: molecular properties, iron storage function and cellular regulation. Biochim Biophys Acta, Bioenerg. 1996;1275(3):161–203.

He D, Piergentili C, Ross J, Tarrant E, Tuck LR, Logan Mackay C, et al. Conservation of the structural and functional architecture of encapsulated ferritins in bacteria and archaea. Biochem J. 2019;476:975–89.

Ilari A, Stefanini S, Chiancone E, Tsernoglou D. The dodecameric ferritin from *listeria innocua* contains a novel intersubunit iron-binding site. Nat Struct Biol. 2000;7(1):38–43.

Keppler J, Martin D, Garamus V, Schwarz K. Differences in binding behavior of (−)-epigallocatechin Gallate to β-Lactoglobulin

heterodimers (AB) compared to homodimers (A) and (B). J Mol Recognit. 2015;28(11):667–75.
Kilic MA, Spiro S, Moore GR. Stability of a 24-meric homopolymer: comparative studies of assembly-defective mutants of *Rhodobacter capsulatus* bacterioferritin and the native protein. Protein Sci. 2003;12:1663–74.
Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW. New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem J. 2021;478(24):4169–85.
Koorts AM, Levay PF, Hall AN, van der Merwe CF, Becker PJ, Frantzen DJM, et al. Expression of the H- and L-subunits of ferritin in bone marrow macrophages of patients with osteoarthritis. Exp Biol Med. 2012;237(6):688–93.
Koorts AM, Levay PF, Hall AN, van der Merwe CF, Becker PJ, Viljoen M. Expression of the H-subunit and L-subunit of ferritin in bone marrow macrophages and cells of the erythron during cellular immune activation. Blood Cells Mol Dis. 2011;47(1):50–5.
Kumar M, Markiewicz-Mizera J, Janna Olmos JD, Wilk P, Grudnik P, Biela AP, et al. A single residue can modulate nanocage assembly in salt dependent ferritin. Nanoscale. 2021;13:11932–42.
Kuno S, Iwai K. Oxygen modulates iron homeostasis by switching iron sensing of NCOA4. J Biol Chem. 2023;299(5):104701.
Leggett BA, Fletcher LM, Ramm GA, Powell LW, Halliday JW. Differential regulation of ferritin H and L subunit mRNA during inflammation and long-term iron overload. J Gastroenterol Hepatol. 1993;8(1):21–7.
Li Z, Maity B, Hishikawa Y, Ueno T, Lu D. Importance of the subunit−subunit interface in ferritin disassembly: a molecular dynamics study. Langmuir. 2022;38:1106–13.
Longo T, Kim S, Srivastava AK, Hurley L, Ji K, Viescas AJ, et al. Micromagnetic and morphological characterization of heteropolymer human ferritin cores. Nanoscale Adv. 2023;5:208-219.
Lu Y, Zhang J, Sun Z, Cheng C, Liu Y, Wu L, et al. NCOA4 requires a [3Fe-4S] to sense and maintain the iron homeostasis. J Biol Chem. 2024;300(2):105612.
Lv C, Zhang X, Liu Y, Zhang T, Chen H, Zang J, et al. Redesign of protein nanocages: the way from 0D, 1D, 2D to 3D assembly. Chem Soc Rev. 2021;50:3957–89.
Maddhuri Venkata Subramaniya SR, Terashi G, Kihara D. Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning. Nat Methods. 2019;16:911–7.
Maity B, Li Z, Niwase K, Ganser C, Furuta T, Uchihashi T, et al. Single-molecule level dynamic observation of disassembly of the Apo-ferritin cage in solution. Phys Chem Chem Phys. 2020;22:18562–72.
Mancias JD, Wang X, Gygi SP, Harper JW, Kimmelman AC. Quantitative proteomics identifies NCOA4 as the cargo receptor mediating ferritinophagy. Nature. 2014;509:105–9.
Mehlenbacher M, Poli M, Arosio P, Santambrogio P, Levi S, Chasteen ND, et al. Iron oxidation and core formation in recombinant heteropolymer ferritins. Biochemistry. 2017;56(30):3900–12.
Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: tools for structure building and analysis. Protein Sci. 2023;32(11):e4792.
Mohanty A, Mithra K, Jena SS, Behera RK. Kinetics of ferritin self-assembly by laser light scattering: impact of subunit concentration, pH, and ionic strength. Biomacromolecules. 2021;22:1389–98.
Nakamura T, Wang X, Terashi G, Kihara D. DAQ-score database: assessment of map-model compatibility for protein structure models from cryo-EM maps. Nat Methods. 2023;20:775–6.
Nivón LG, Moretti R, Baker D. A Pareto-optimal refinement method for protein design scaffolds. PLoS One. 2013;8:e59004.
O'Shea E, Rutkowski R, Stafford W, Kim PS. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. Science. 1989;245(4920):646–8.
Otsuka S, Listowsky I, Niitsu Y, Urushizaki I. Assembly of intra- and interspecies hybrid apoferritins. J Biol Chem. 1980;255(13):6234–7.
Pearson ES, Adyanthāya NK. The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. Biometrika. 1929;21(1/4):259–86.
Pulsipher KW, Villegas JA, Roose BW, Hicks TL, Yoon J, Saven JG, et al. Thermophilic ferritin 24mer assembly and nanoparticle encapsulation modulated by interdimer electrostatic repulsion. Biochemistry. 2017;56:3596–606.
Punjani A, Rubinstein J, Fleet DJ, Brubaker MA. CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14:290–6.
Rohou A, Grigorieff N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015;192(2):216–21.
Ross J, Lambert T, Piergentili C, He D, Waldron KJ, Mackay L, et al. Mass spectrometry reveals the assembly pathway of encapsulated ferritins and highlights a dynamic ferroxidase interface. Chem Commun. 2020;56:3417–20.
Sammarco MC, Ditch S, Banerjee A, Grabczyk E. Ferritin L and H subunits are differentially regulated on a post-transcriptional level. J Biol Chem. 2008;283(8):4578–87.
Sana B, Johnson E, Le Magueres P, Criswell A, Cascio D, Lim S. The role of non-conserved residues of *archaeoglobus fulgidus* ferritin on its unique structure and biophysical properties. J Biol Chem. 2013;288:32663–72.
Santambrogio P, Levi S, Cozzi A, Rovida E, Albertini A, Arosio P. Production and characterization of recombinant heteropolymers of human ferritin H and L chains. J Biol Chem. 1993;268:12744–8.
Santambrogio P, Pinto P, Levi S, Cozzi A, Rovida E, Albertini A, et al. Effects of modifications near the 2-, 3- and 4-fold symmetry axes on human ferritin renaturation. Biochem J. 1997;322:461–8.
Sato D, Ikeguchi M. Mechanisms of ferritin assembly studied by time-resolved small-angle x-ray scattering. Biophys Rev. 2019;11(3):449–55.
Sato D, Ohtomo H, Yamada Y, Hikima T, Kurobe A, Fujiwara K, et al. Ferritin assembly revisited: a time-resolved small-angle x-ray scattering study. Biochemistry. 2016;55(2):287–93.
Smirnov NV. Estimate of deviation between empirical distribution functions in two independent samples. Bull Moscow Univ. 1939;2(2):3–16.
Srivastava AK, Poli M, Arosio P, Bou-Abdallah F. A novel approach for the synthesis of human heteropolymer ferritins of different H to L subunit ratios. J Mol Biol. 2021;433(19):167198.

Srivastava AK, Reutovich AA, Hunter NJ, Arosio P, Bou-Abdallah F. Ferritin microheterogeneity, subunit composition, functional, and physiological implications. Sci Rep. 2023;13:19862.

Srivastava AK, Scalcione LJ, Arosio P, Bou-Abdallah F. Hyperthermostable recombinant human heteropolymer ferritin derived from a novel plasmid design. Protein Sci. 2022;32(1):e4543.

Stefanini S, Vecchini P, Chiancone E. On the mechanism of horse spleen apoferritin assembly: a sedimentation velocity and circular dichroism study. Biochemistry. 1987;26:1831–7.

Subhadarshanee B, Mohanty A, Jagdev MK, Vasudevan D, Behera RK. Surface charge dependent separation of modified and hybrid ferritin in native PAGE: impact of lysine 104. Biochim Biophys Acta - Proteins Proteom. 2017;1865(10):1267–73.

Sudarev VV, Dolotova SM, Bukhalovich SM, Bazhenov SV, Ryzhykau YL, Uversky VN, et al. Ferritin self-assembly, structure, function, and biotechnological applications. Int J Biol Macromol. 2023;224:319–43.

Terashi G, Wang X, Kihara D. Protein model refinement for cryo-EM maps using AlphaFold2 and the DAQ score. Acta Crystallogr D Struct Biol. 2023;79:10–21.

Terashi G, Wang X, Maddhuri Venkata Subramaniya SR, Tesmer JJG, Kihara D. Residue-wise local quality estimation for protein models from cryo-EM maps. Nat Methods. 2022;19:1116–25.

The PyMOL Molecular Graphics System. PyMOL (v.2.5.4) by Schrodinger. Available from: https://www.pymol.org/

Torti FM, Torti SV. Regulation of ferritin genes and protein. Blood. 2002;99(10):3505–16.

Wang RY-R, Song Y, Barad BA, Cheng Y, Fraser JS, DiMaio F. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. elife. 2016;5:e17219.

Wang W, Knovich MA, Coffman LG, Torti FM, Torti SV. Serum ferritin: past, present and future. Biochim Biophys Acta, Gen Subj. 2010;1800(8):760–9.

Wang X, Alnabati E, Aderinwale TW, Maddhuri Venkata Subramaniya SR, Terashi G, Kihara D. Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning. Nat Commun. 2021;12(1):2302.

Yao H, Soldano A, Fontenot L, Donnarumma F, Lovell S, Chandler JR, et al. *Pseudomonas aeruginosa* Bacterioferritin is assembled from FtnA and BfrB subunits with the relative proportions dependent on the environmental oxygen availability. Biomolecules. 2022;12:366.

Yin V, Devine PWA, Saunders JC, Barendregt A, Cusdin F, Ristani A, et al. Stochastic assembly of biomacromolecular complexes: impact and implications on charge interpretation in native mass spectrometry. Chem Sci. 2023;14:9316.

Yousefi A, Zheng Z, Zargarbashi S, Assadipapari M, Hickman GJ, Parmenter CDJ, et al. Structural flexibility and disassembly kinetics of single ferritins using optical Nanotweezers. bioRxiv. https://doi.org/10.1101/2023.09.22.558948

Zhao G, Bou-Abdallah F, Arosio P, Levi S, Janus-Chandler C, Chasteen ND. Multiple pathways for mineral core formation in mammalian apoferritin. The role of hydrogen peroxide. Biochemistry. 2003;42:3142–50.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bou-Abdallah F, Fish J, Terashi G, Zhang Y, Kihara D, Arosio P. Unveiling the stochastic nature of human heteropolymer ferritin self-assembly mechanism. Protein Science. 2024;33(8):e5104. https://doi.org/10.1002/pro.5104