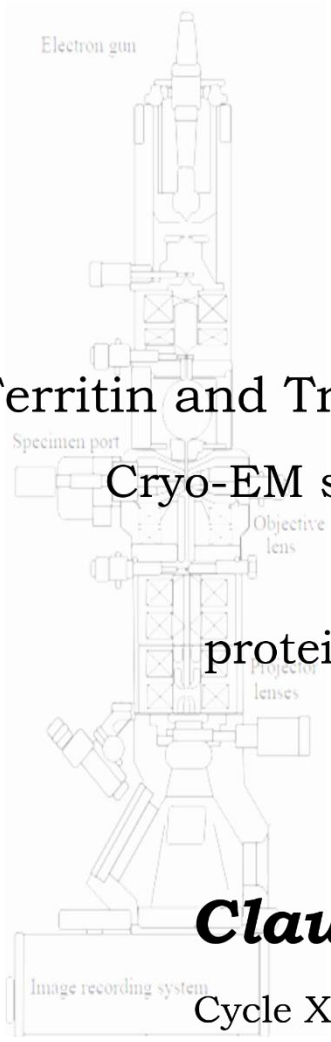
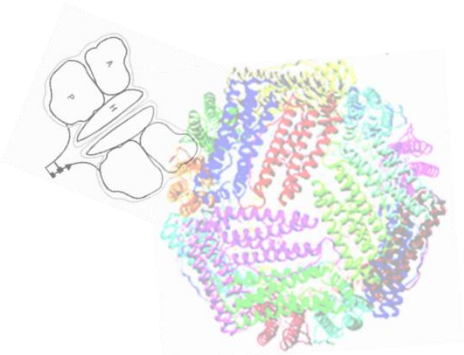
SAPIENZA
UNIVERSITÀ DI ROMA

PhD course in Biochemistry





Ferritin and Transferrin Receptor 1:

Cryo-EM structural studies

&

protein engineering

Claudia Testi

Cycle XXXI (2015-2018)

Tutor
Prof. A. Boffi

Coordinator
Prof. S. Gianni

Contents

List of abbreviations

- **Ft**: Ferritin
- **HFt:** Ferritin H-chain ("human" implied, otherwise the species is specified)
- **LFt:** Ferritin L-chain
- **TfR1**: Transferrin Receptor 1
- **hTfR1**, **CD71**: human Transferrin Receptor 1
- **mTfR1**: murine Transferrin Receptor 1
- **Tf**: Transferrin
- **HFE**: Hemochromatosis protein
- **AfFt**: *Archaeglobus Fulgidus* Ferritin
- **HumAfFt**: Humanized *Archaeglobus Fulgidus* Ferritin
- **HFt-LBT Tb(III):** Lanthanide (Tb(III)) Binding H-chain Ferritin
- **LBT**: Lanthanide binding tags
- **TIM-2**: T cell immunoglobulin and mucin domain protein-2

- **MACV**: Machupo Virus (New World hemorrhagic fever arenavirus)
- **JUNV**: Junin Virus (New World hemorrhagic fever arenavirus)
- **GTOV**: Guanarito Virus (New World hemorrhagic fever arenavirus)
- **SABV**: Sabìa Virus (New World hemorrhagic fever arenavirus)
- ***Pv*RBP2b**: malaria parasite *Plasmodium vivax* invasion protein

- **cryo-EM**: cryo-electron microscopy
- **DDD**: direct detector device
- **CCD**: charge coupled device
- **CTF**: contrast transfer function
- **FSC**: Fourier Shell Correlation
- **FACS**: Fluorescence Activated Cytometry Analysis
- **SPR**: Surface plasmon resonance

1. Introduction

1.1 Ferritin

Biochemical properties

Ferritin (Ft) is a globular protein characterized by a typical tetraeicosameric assembly. Each subunit is ≈20 kDa, while the assembled 24-mer has a total weight of ≈500 kDa; its hollow cage-like spherical structure has an outer diameter of roughly 12 nm and an inner diameter of 8 nm [1] (Figure 1).

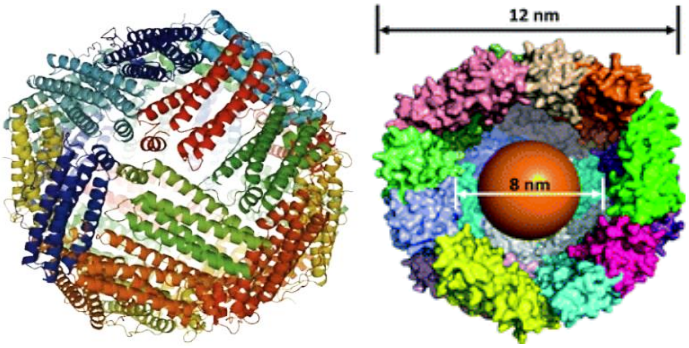


Figure 1: Left: Ferritin three-dimensional structure with 24 subunits highlighted by different colors. **Right**: Ferritin shell internal view, sliced at the protein center, shows the diameters of the internal and external surfaces [2].

Ft physiological function is centered around intracellular iron storage and transport[3]. A tightly controlled iron metabolism is in fact crucial for physiological processes, since this element possesses both toxic and beneficial properties for cellular health. Iron is in fact a cofactor for a vast array of biochemical reactions, like oxygen transport, energy metabolism and DNA synthesis, but unbound Fe(II) is toxic[4,5]: upon interaction with free oxygen, at neutral pH it produces reactive oxygen species (ROS) by the Haber-Weiss process *via* a Fenton reaction[6,7]:

$$Fe^{2+} + O_2 \longrightarrow Fe^{3+} + O_2^-$$
$$2O_2^- + 2H^+ \longrightarrow H_2O_2 + O_2$$
$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + OH^- + \cdot OH$$

Free ROS are deeply destructive for cells: because of their extreme reactivity, they may damage enzymes, membrane lipids, proteins and nucleic acids. In addition, under physiological conditions Fe(III) is highly insoluble (solubility 10^{-18} M at neutral pH[7]) and precipitates as Fe(III) oxide species. To avoid these lethal side effects, Fe(II) is always bound to a protein ligand, such as Transferrin (Tf) or Ferritin[3].

Ferritin binds and oxidizes Fe(II) in Fe(III) through the ferroxidase reaction, a mechanism that involves several steps and leads to the production of different intermediate species, whose general overview is this [7]:

$$2Fe^{2+} + O_2 + 4H_2O \rightarrow [Fe_2O_2]^{2+} \rightarrow [Fe_2O(OH)_2]^{2+} \rightarrow 2FeOOH_{core} + H_2O_2 + 4H^+$$

The resultant, inert, Fe(III) product is stored inside Ft internal cavity in the form of the crystalline mineral ferrihydrite: the shell encapsulates up to 4500 Fe^{3+} atoms[1].

There is a universal need for removing and storing iron in an inert form: accordingly, Fts are ubiquitously distributed in vertebrates, invertebrates,

plants and fungi; even bacterial species have their own type of Ferritin, although it has evolved separately[1]. In mammalians, it is prevalently found inside the cytosol of the cells and modestly in the serum. The mammalian Ferritin is the product of self-assembly of 24 subunits of two different types[8,9]: LFt ("light" Ferritin type, 20 kDa) and HFt ("heavy" Ferritin type, 22 kDa). The proportions of H and L subunits in a whole Ferritin molecule vary across different tissues: H subunits is mainly expressed in the heart, at early stages of development and in cancer tissues, while the L subunit is most represented in the liver and spleen[9]. Importantly, only the H subunits are able to catalytically oxidize Fe(II)[2].

The conformation of each subunit consists of a 4-helix bundle of alpha helices: A (in humans, residues 10-39), B (45-72), C (92-120) and D (124-155); a fifth helix, E (res 160-169), lies at an acute angle to the bundle, pointing towards the internal cavity of the three-dimensional protein. A long loop L (res 73-91) connects B and C helices[1].

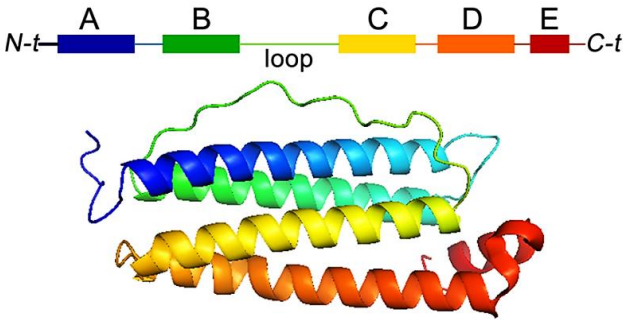


Figure 2: Ferritin tertiary structure of one subunit: a monomer, composed by a 4 helix-boundle (A, B, C, D helices), a loop L and a fifth E helix [10].

The quaternary structure is arranged in octahedral 432 symmetry (point group 'O', Figure 3). Pairs of helix bundles lie antiparallel, related by 2-fold symmetry;

the loops of these pairs form a region of anti-parallel sheet lying on the outside surface of the shell.

The symmetrical junction of three or four monomers in the protein shell results in the formation of several channels (6 hydrophobic channels at the 4-fold symmetry axis and 8 hydrophilic canals at the 3-fold axis intersections) connecting Ft inner cavity to the outside (Figure 4). These channels bind Fe^{2+} and other cations (e.g., Tb^{3+}, Cd^{2+}, Zn^{2+} in the 3-fold channels[11]) with a broad selectivity. HFt has specific active sites for catalyzing the Fe(II) oxidation in Fe(III), known as the ferroxidase centers, located in every monomer between A and B helices. Very high sequence homology of their residues is found throughout the three kingdoms of nature[2]. Despite the available information regarding the amino acids involved in iron binding, it is not clear how the ferroxidase center operates, nor is still understood the pathway for Fe(II) translocation from the 3-fold channels to the ferroxidase site[2].

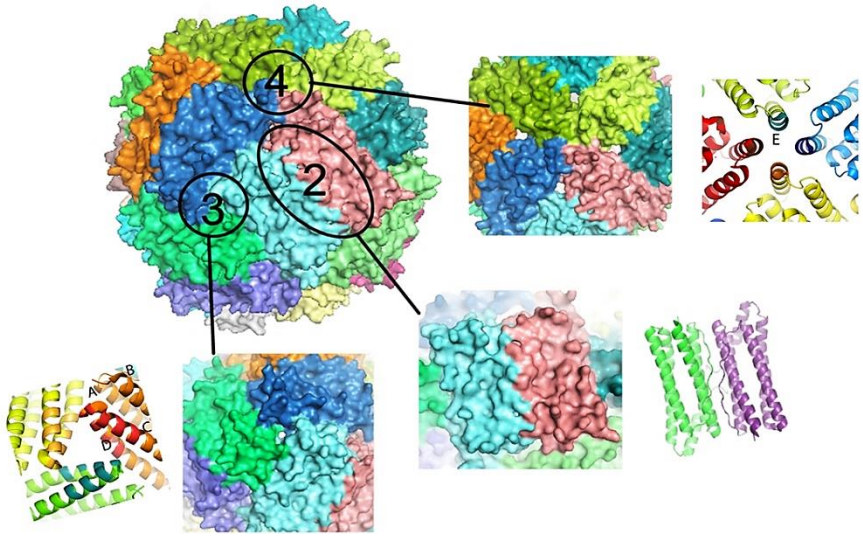


Figure 3 Ferritin quaternary structure, arranged with 24 subunits in an octahedral 4-3-2 symmetry: surfaces and crystallography overview. (Image modified [10])

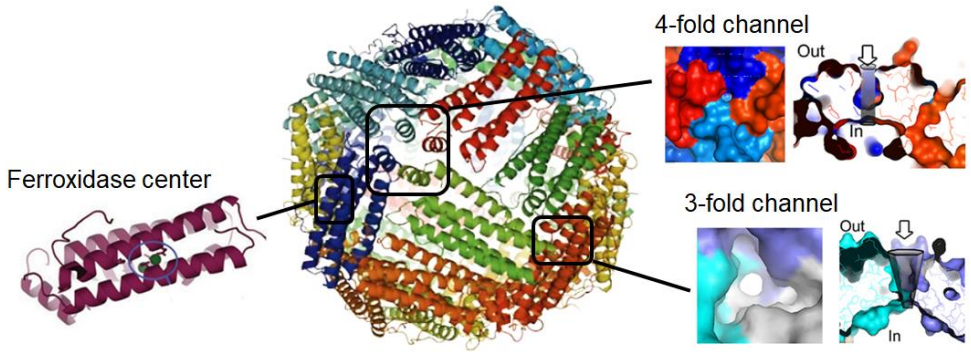


Figure 4 Ferritin inner cavity is connected with the outside through several channels, known as 4-fold and 3-fold channels found in the respective symmetry axes. Ferroxidase centers oxidize Fe(II) in Fe(III) and are located at the center of the helix bundle. (Image modified [2, 10])

Hence, Ft is made by flexible structures that allow the entry and exit of different ions and molecules through the protein shell. Despite this flexibility, its quaternary structure is incredibly stable, thanks to the numerous intersubunit interactions (salt bridges, H-bonds, apolar interactions) between the monomers[1]: the cage is resistant to denaturants and to high temperatures (even higher than 80 °C)[12]. For this reason, subunits can be dissociated from the tetraicosameric structure only by reaching drastic pH values (i.e. pH <3 or pH >10): when the pH returns to neutrality, monomers spontaneously assemble in a shape memory fashion[13].

In the self-assembly of the 24-meric shell, the first stable intermediate appears to be a dimer[14] through the interaction of the respective BC loops (deletion of two residues in the loop has been proved to abolish the formation of the whole 24-mer[15]). The kinetics pathway from dimers to a complete 24-mer shell remains elusive, even though this pH-dependent assembly is currently widely studied[16]. The disassembly in dimers is partially reversible and, as mentioned before,

occurs spontaneously at very acidic or basic pHs.

In summary, the mechanism of functioning of mammalians Ft is divided into six steps[2,10] (Figure 5):

(1) Self-assembly of dimers to form the 24-mer shell at neutral pH, with unknown kinetics[14];

(2) entry of Fe(II) in Ft through the 3-fold channels;

(3) Fe(II) translocation to the catalytic sites in the ferroxidase center;

(4) ferroxidase reaction: oxidation of Fe(II) to Fe(III) in the catalytic center, according to the Haber-Weiss process[6,7]. Only HFts are catalytically active;

(5) storage of Fe(III) product inside the cavity in form of ferrihydrite;

(6) when needed, release in a controlled fashion of Fe(III) from the mineral core to the external through the 3-fold channels.

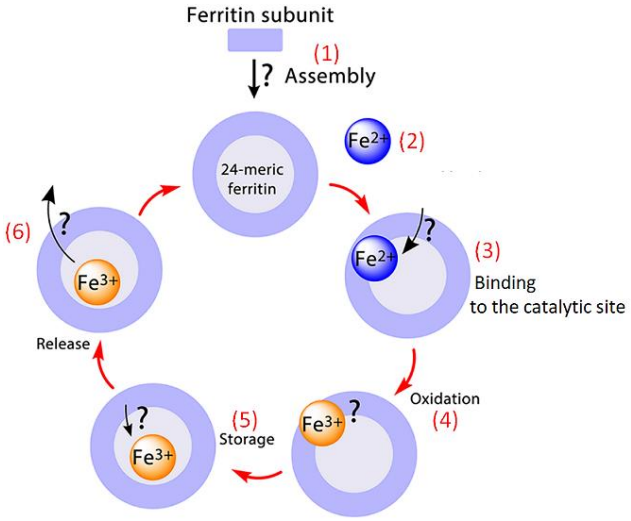


Figure 5 Mechanism of assembly and iron homeostasis of mammalian HFt. (Image modified[10])

A peculiar, non-pH-dependent association is noted in an archaeal protein, the

Archaeglobus fulgidus ferritin (AfFt), that self-assembles from dimers to a complete 24-mer structure only at high salt concentrations (>0.3 M NaCl)[17]. Its oligomerization mechanism is thus easily controlled just by altering cations concentration in the buffer and does not require the harsh conditions of a pH jump needed for mammalian Fts.

The secondary and tertiary structures of AfFt and human HFt closely resemble, even in the metal ion binding mode, despite their low sequence similarity. Nevertheless, they deeply differ in the quaternary structure: AfFt has tetrahedral (2-3) symmetry, while HFt displays the canonical octahedral (4-3-2); instead of HFt 4-fold symmetry channels, four large (45 Å) pores open in the shell of AfFt as a result of the packing between two hexamers[17,18].

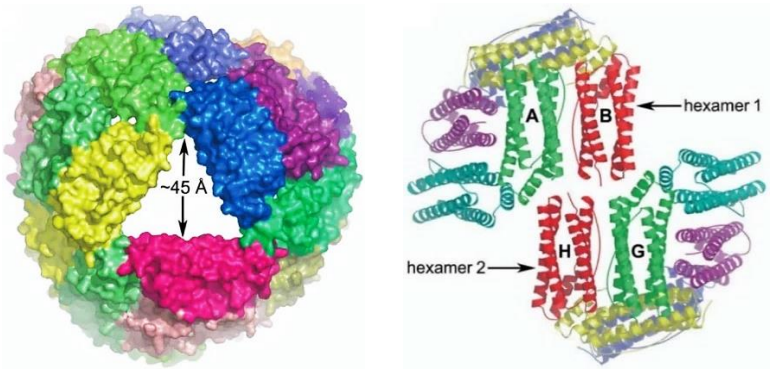


Figure 6 Left: Surface representation of AfFt tetraeicosameric assembly. The molecule is viewed from one of the 4 large pores at a 3-fold symmetry axis. **Right**: ribbon model of the AfFt asymmetric unit[18].

Its salt-dependent reversible association might be used for nanotechnological purposes; we will deal with AfFt nanoparticle in section 4.2.

Apart from the cytosol, HFt is also modestly found in the serum, where it acts as an iron carrier, transporting Fe(III) to areas where it is required: HFt is in fact able to bind to a great variety of cells in a saturable fashion[19]. It was not until 2010 that HFt cellular receptor has been identified as the human

transferrin receptor 1 (hTfR1 or CD71)[20], a transmembrane protein able to bind other iron-regulating ligands as Transferrin (Tf)[21,22] or hemochromatosis-associated protein (HFE)[23]. It is worth noting that CD71 recognizes only HFt, while it does not recognize LFt[20]; HFt binding epitopes with CD71 are to date unidentified.

After binding of HFt to CD71 on the surface, HFt enters in the cell through a hTfR1-mediated endocytosis, a complex process mediated by a clathrin-dependent pathway. Once the iron-loaded Ft is transported to endosomes and lysosomes, the receptor is recycled to the cell surface[20].

Curiously, this interaction seems to be specie-related: for example, in mice TIM-2 (T cell immunoglobulin and mucin domain protein-2) serves as an endocytosing receptor for mouse HFt (that shares 92% identity with human HFt[20]), while the murine LFt recognizes Scara5 receptor; however, TIM-2 is not expressed in humans[24,25].

We will deal with Tf, HFE, clathrin-dependent pathways and CD71 biochemical properties in session 1.2.

H-Ferritin as nanovector for biomedicine

Among thousands of nanovectors currently under study, Ferritin H-chain homopolymer is one of the most versatile for biomedical purposes. In fact, thanks to the above-mentioned biochemical properties, this nanoparticle is endowed with special features for targeted delivery of molecules [26,27,28,29]:

i) biocompatibility: biological-derived nanoparticles are exploited in biomedicine for their low toxicity and ability to elude the immune system. HFt is indeed a physiological non-toxic protein, already present in human organism;

ii) size: this parameter is critical for directing vectors distribution *in vivo*[30]. Nanoparticles should be not too small (<5 nm) nor too large (>150 nm) for avoiding kidneys filtering in the former case[31] or liver/spleen entrapment in the latter[32]. Thus, 10-100 nm diameter generally proves long-lasting times in the circulation[30]: indeed, Ferritin is 12 nm large. Its purification protocols give high levels of homogeneity and monodispersity[33], important factors that increase the efficiency of a nanovector [34];

iii) stability: upon *in vivo* administration, particles should remain stable while in circulation in order to arrive intact at the intended site. HFt intersubunit interactions give to it an exceptional stability: as discussed before, the cage is resistant both to denaturants and to high temperatures[33];

iv) specificity towards a target: one of the major challenges in drug-delivery is to release the drug in the right organ: unspecific uptake by healthy cells leads to severe cytotoxic side effects and to low therapeutic levels at disease sites[34,35]. The delivery properties of nanoparticle-associated molecules are also much more effective if compared to conventional free drug formulations, that tend to distribute randomly throughout the body[30,34,35,36,37]. Consequently, nanoparticles need to be selective for specific tumor environments or, possibly, tumor cell types: in the case of HFt, it binds hTfR1 with nanomolar affinity[20,38]. As explained in section 1.2, this receptor is highly expressed in iron avid cancer cells and represents the ideal target which a nanoparticle should bind to;

v) internalization in cells: after reaching the right organ, nanoparticles should be internalized in cells to release the cargo[30]. hTfR1-mediated endocytosis[39] assures HFt entry in cells;

vi) surface properties: HFt internal and external surfaces can be genetically and chemically modifiable. Thus, a plethora of molecules can be attached to the outside (Figure 7) or encapsulated in the internal cavity (Figure 8): metals[40], drugs[41], antibodies[42], fluorescent molecules[43], contrast agents[44]...

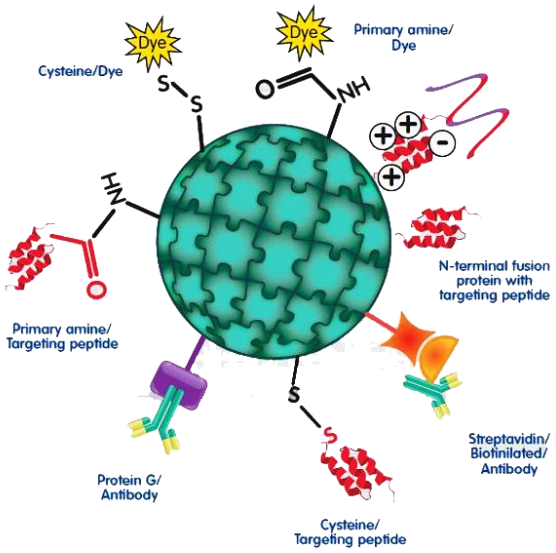


Figure 7 Ferritin external surface can be functionalized by attaching different molecules [12].

Commonly employed cargo encapsulation techniques involve the HFt quaternary structure disassembly at drastic pH values followed by re-assembly at neutral pH[45]. The drawbacks of this rough procedure lie in the possibility of a partial and incomplete re-assembly of the cage and in an inevitable damage to the loaded compound[46].

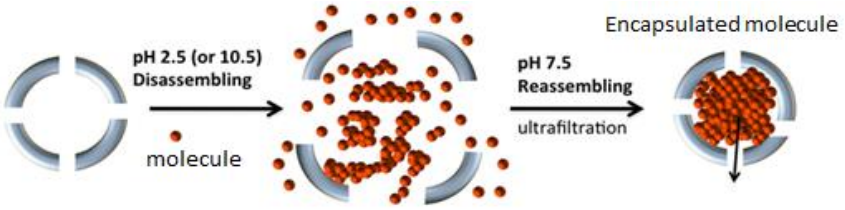


Figure 8 Loading of molecules in Ferritin internal cavity through pH jump.

Consequently, HFts are implemented in nanotechnology as biological nano-cages for applications in targeted drug delivery[38,41,47,48], in nanoelectronic devices[49], bioimaging[40], vaccine development[50] and cancer theranostics[43,51,52]; in the latter case, HFt-CD71 interaction is particularly exploited. In a recent study, over 400 clinical specimens from patients with nine types of cancer were examined: because of hTfR1 overexpression, HFts nano particles could distinguish cancerous cells from normal cells with a sensitivity of 98% and specificity of 95%[51]. Ft-based nanoparticles might even be used to treat hypoxic solid tumor sites, where nanoparticles usually undergo impaired diffusion and are not able to penetrate. In fact, these areas, located far away from blood vessels, attract macrophages (able to internalize HFt[53]): recent findings show that macrophages are also able to transfer drug-loaded Fts to these hypoxic areas. Thus, macrophages might be used to deliver ferritin-encapsulated compounds directly to the solid tumor zones[54].

HFt presents also several advantages even over targeted antibodies and CD71 main ligand, *i.e.* transferrin (Tf): its 250 nm^3 hollow cage can host up to hundreds small molecules[41], while Tf or antibodies can be engineered to bind only a few molecules in positions far from the receptor recognition epitopes[55]. Thus, HFt has more a convenient "drug/protein ratio" than the classical "drug/antibody ratio". Furthermore, HFt has 24 symmetrical recognition epitopes for CD71, instead of just 2 of antibodies or Tf: this results in high

affinity towards the target receptor (a property often referred to as "multivalence effect"[33], that, in the case of CD71, rivals with typical antibodies affinities[38]). Currently, Doxorubicin encapsulation in Ft nanocages represents the most extensively investigated system for delivery of anticancer drugs[56].

Despite the large use in literature of HFt-CD71 interaction so far discussed, the structural atomic details are to date unknown. Therefore, finding the residues involved in this contact is of major importance for providing the structural basis that govern HFt-CD71 interaction, beyond nanotechnological applications.

1.2 Transferrin Receptor 1

Transferrin receptor 1 (TfR1; in humans hTfR1 or CD71) is a universally expressed cell entry carrier whose primary function is to import Transferrin-bound iron in response to variations in intracellular concentration of this essential element[57].

It's a homodimeric type II transmembrane protein composed of a small cytoplasmic domain (in humans, residues 1-60), a single-pass transmembrane region (residues 61-88), and a dimeric extracellular domain (residues 89-760). Each monomer of the ectodomain, whose total molecular weight is 150 kDa, is in turn subdivided in a protease-like domain (P) in contact with the cell membrane, a helical domain (H) comprising the dimer contact regions, and an apical (A) domain[58] (Figure 9).

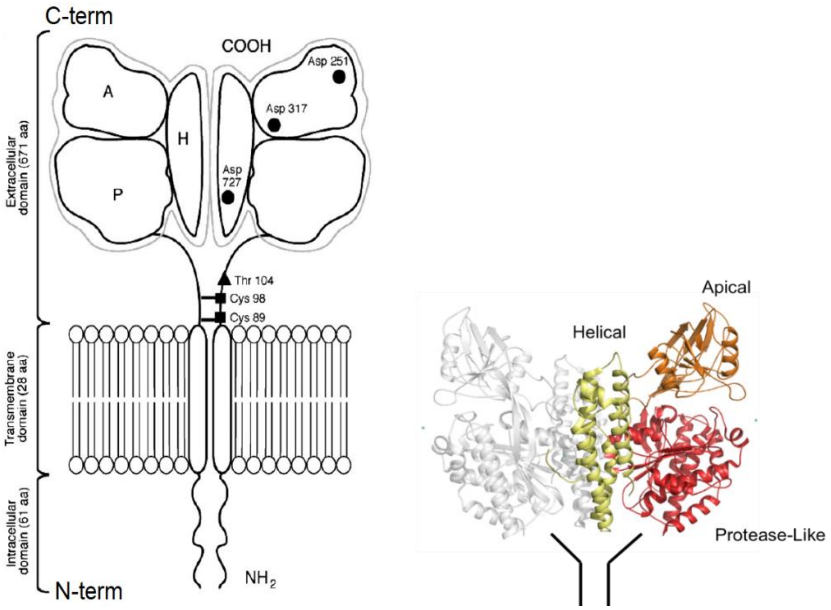


Figure 9 Human Transferrin Receptor 1 (hTfR1 or CD71). **Left**: cartoon structure of the whole protein, divided in intracellular, transmembrane and extracellular domains [55]. **Right**: crystallographic structure (pdb 3KAS, visualized with Chimera[59]) of the sole ectodomain, subdivided in protease-like, helical and apical domains.

The ectodomain displays binding sites for diverse ligands (Figure 10).

Its basal portion (formed by P and H domains) binds transferrin (Tf), while the H domain forms a complex with the Hereditary Hemochromatosis factor (HFE): these proteins are both involved in iron homeostasis. HFE negatively regulates iron uptake by competing with Tf for CD71: its association lowers the affinity of hTfR1 for Tf up to 50-fold[23, 60].

Transferrin is a 80 kDa bilobed glycoprotein that transports one or two ferric ions in the serum. Two Fe(III)-bearing Tf molecules can bind the dimeric CD71 in the basal portion with nM affinity (iron-free transferrin is not recognized by

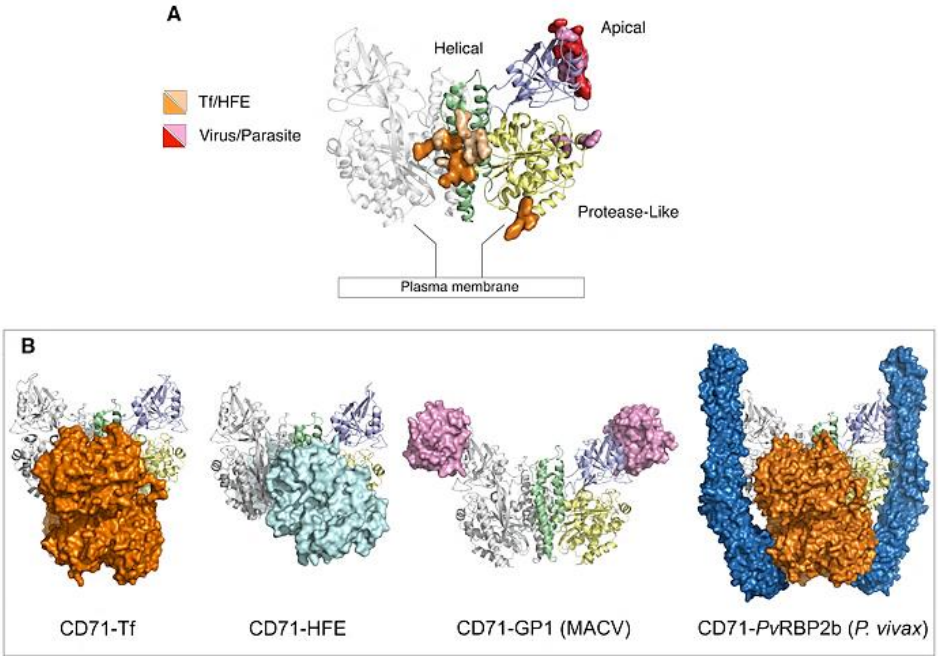


Figure 10 CD71 ligand recognition epitopes and binding modes that are to date known. **A.** CD71 residues identified as recognition epitopes for iron-regulating proteins Tf/HFE and viruses/parasite are represented as orange/wheat and red/pink surfaces, respectively. **B.** CD71 receptor is shown bound to Tf (orange surface, pdb 1SUV[22]), HFE (cyan surface, pdb 1DE4[23]), GP1 protein of MACV (pink surface, pdb 3KAS[61]), Tf and *Pv*RBP2b from *P. vivax* (orange and blue surfaces, pdb 6D04[62]). (Figure obtained with Chimera[59])

TfR at the extracellular pH of 7.4)[21,22]. Iron uptake is mediated by the internalization of the transferrins-iron complex through receptor-mediated endocytosis *via* a clathrin-dependent pathway[57] (Figure 11). Embedded by clathrin-coated pits, the complex is fused with endosomes, whose acidic pH induces a conformational change in Tf: iron is released from this protein while apoTf remains bound to CD71. Once the iron cargo is delivered in the cell, the apoTf-hTfR1 complex is recycled back to the membrane external surface; the extracellular pH leads to their dissociation and apo-transferrin is released into the bloodstream[63, 64].

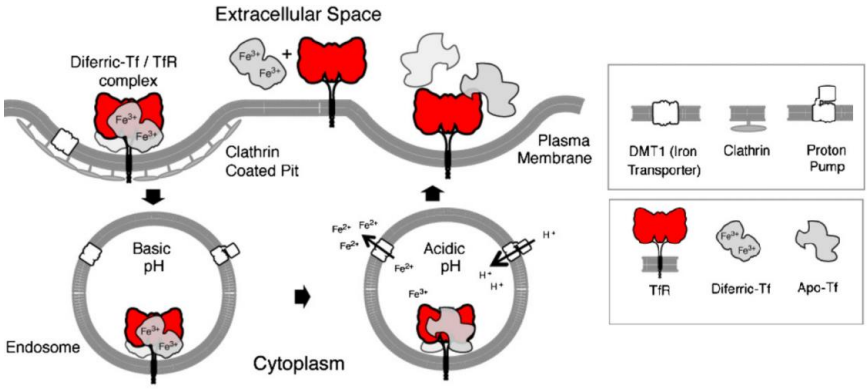


Figure 11 CD71 mediated endocytosis (clathrin-dependent pathway) of diferric-Tf [65].

The apical domain, instead, interacts with ligands having completely different structures and functions respect to Tf or HFE. Its upper part binds New World hemorrhagic fever arenaviruses (namely Machupo, MACV, Junin, JUNV, Guanarito, GTOV, and Sabìa, SABV viruses)[61,66,67,68,69,70], feline and canine specific parvoviruses[71], hepatitis C virus[72] and the most common malaria parasite *Plasmodium vivax* invasion protein (*Pv*RBP2b)[62,73]. Viral systems recognize epitopes on the host-encoded CD71 receptor through their surface spike glycoproteins, triggering the internalization of the complex. Apart from specific ligand-dependent residues, all these pathogens recognize a specific portion of CD71 apical domain (Figure 12): residues 208-212 and 215 located on the βII-2 strand and residues 343, 344, 348 on the αII-2 helix; in particular, Tyr211 is a critical determinant for viruses cell entry[70].

As a result, it is clear that CD71 acts as a cellular entry carrier not only for iron-regulator proteins through its basal part, but also for several pathogens through the apical domain.

CD71 has been also shown to mediate the uptake of heavy chain Ferritin (HFt) from serum as an alternative or additional source of bioavailable iron[20], despite

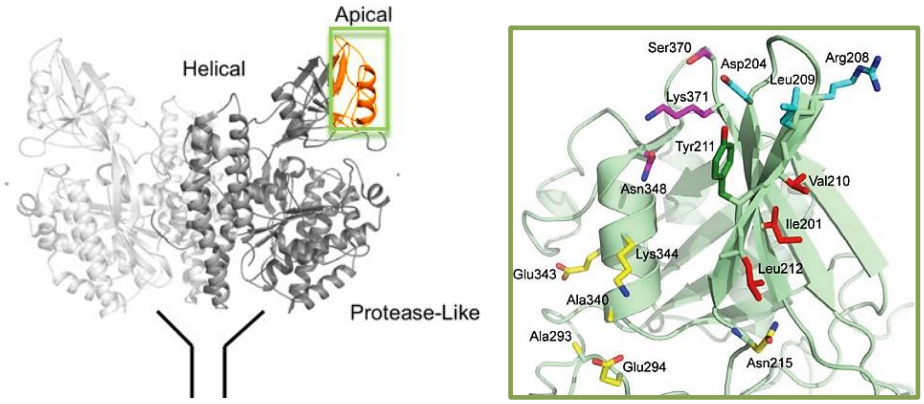


Figure 12 Left: Portion of CD71 apical domain recognized by several arenaviruses and *P. vivax* parasite for cellular invasion. **Right**: close-up view of the specific residues: 208-212, 215 on the βII-2 strand and 343-344, 348 on αII-2 helix [61].

Tf, HFE and HFt have completely different molecular structures and dimensions. The mechanism of how hTfR1 mediates the internalization of such different ligands remains unknown, nor are known the epitopes of interaction on both HFt or hTfR1 proteins sides. Surface plasmon resonance (SPR) data show that the two proteins interact with a nanomolar affinity[38].

This transmembrane protein constitutes also an attractive target in biomedicine: in a plethora of different types of cancer cells[74] (e.g. rectum, cardia, liver, stomach, breast, pancreas, ovary, prostate: Figure 13), its levels of expression are up to 100-fold higher than the average expression in normal cells[39,55,65,75,76,77]. It has been proposed that this overexpression may be attributed to the necessity of iron for cancer cell proliferation[78].

CD71 extracellular accessibility, its ability to internalize different proteins and its central role in the cellular pathology of human cancer further highlight its

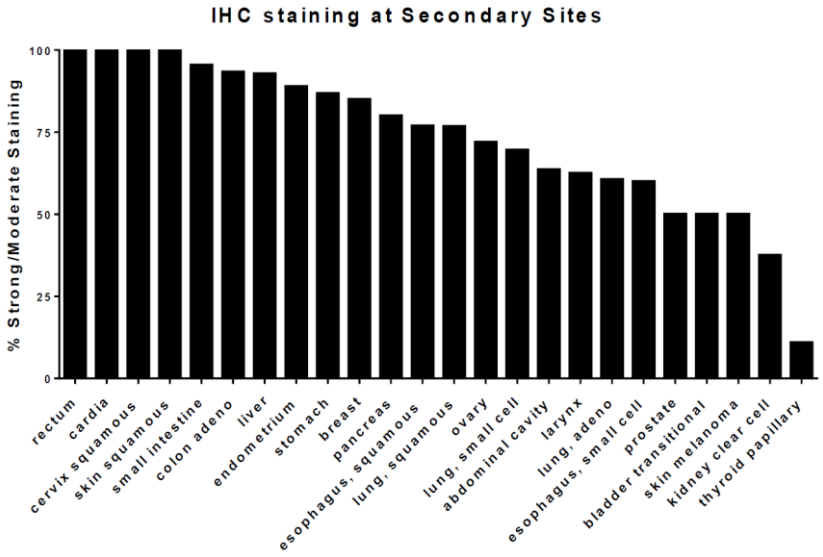


Figure 13 CD71 immunohistochemistry (IHC) staining in different tissues show that it is highly overexpressed in many metastatic cancers [74].

potential use as a privileged target for the selective delivery of cytotoxic drugs coupled to its ligands (Tfs or HFts) or monoclonal antibody drug conjugates. In particular, it is currently used as a portal for the delivery of drug-loaded-HFts into malignant cells[38,40,41,42,43,44,47,48,51,52]. As mentioned at the end of section 1.1, the advantage of using HFt instead of Tf or antibodies for hTfR1-targeted nanoparticles delivery lies in the number of molecules that these proteins may load: up to hundreds in HFt cavity, only a few for Tf or antibodies.

However, a key missing piece of information concerns the structural basis of the interaction between CD71 and HFt. We managed to discover the structural determinants that govern HFt-hTfR1 interaction thanks to cryo-electron microscopy, an imaging technique that is increasing its importance in structural biology year over year.

1.3 Single-particle cryo-EM

X-ray crystallography *vs* Cryo-EM

Before 2013, most atomic structures of macromolecules deposited in the protein data bank (PDB) were determined by X-ray crystallography[79]. The physics beyond this technique lies on the diffraction of a coherent X-ray beam incident on a crystal: because of its ordered structure, diffracted X-rays are scattered only in specific directions (Bragg diffraction, Figure 14). By measuring the diffracted pattern in Fourier space, the three-dimensional structure of the crystal can be reconstructed in real space with atomic resolution[80]: this is the reason why X-ray crystallography is one of the most employed experimental techniques to solve protein structures [81].

The determination of a protein structure with this technique completely depends on the quality of the sample, that is a protein transformed into a well-ordered 3D crystal. However, it might be arduous to obtain the favorable experimental conditions in which the good crystal grows, in particular for insoluble proteins; besides, the sample should be extremely homogeneous in both composition and conformation in order for the X-ray beam scattering to occur correctly[82].

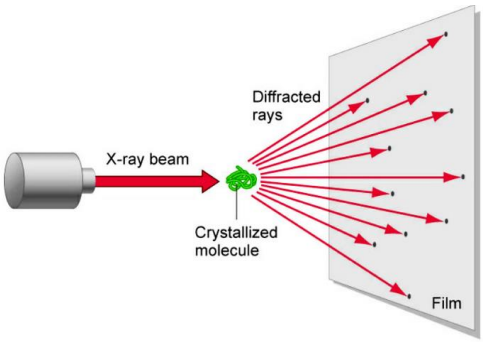


Figure 14 Diffraction of an incident X-ray beam on a crystallized molecule for solving its structure at atomic resolution [80].

Cryo-electron microscopy (cryo-EM) was used for solving the morphology of large protein complexes that resisted crystallization; the electron density maps were reconstructed at much lower resolutions than the crystallographic ones[83]. However, thanks to technological advances in instrumentation (in particular the introduction of direct detectors, as will be explained later) and to more efficient algorithms for data analysis calculations, cryo-EM from 2013 on begun to be more and more used in structural biology[81,84]. In particular, single-particle analysis emerged as the most common approach for inspecting the huge quantity of data produced by an electron microscope.

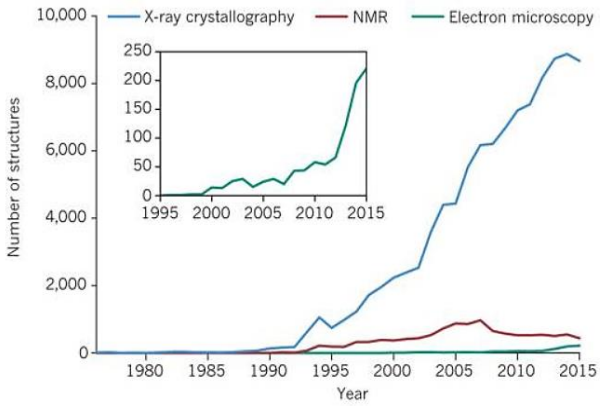


Figure 15 Number of proteins structures deposited in the Protein Data Bank solved with X-ray crystallography (blue), NMR (red) and EM (green) [81]. In 2013, the introduction of Direct Detectors determined a huge increase in the resolutions achievable with cryo-EM [79].

Nowadays (2018), cryo-EM is reaching X-ray crystallography resolution and can be used as a standard method to determine atomic structures of macromolecules, protein complexes or molecules that are difficult to crystallize in specific functional states[79] (as membrane proteins). A compulsory limit of this technique lies in the size of the biological sample to image, that should not be smaller than

≈100 kDa (currently, the smallest protein ever resolved with cryo-EM is haemoglobin, 64 kDa, at a resolution of 3.2 Å[85]).

These new technological progresses have made single-particle cryo-EM the "Method of the Year 2015" for *Nature*[86]; moreover, three electron microscopists were awarded with the 2017 Nobel Prize in Chemistry.

While crystallization locks a protein in its most stable orientation, proteins in cryo-EM are free to move until the moment of freezing. As such, this technique images biological samples in their native state in solution and might visualize multiple states coexisting in a non-homogeneous specimen[87,79]. Moreover, as will be explained below, the sample preparation for cryo-EM is much handier than the crystallization process.

In summary, cryo-EM is now developing into a fast and practical alternative for high-resolution protein structure determination. In this context, solving Ferritins structures with cryo-EM is particularly straightforward: indeed, horse spleen apoferritin has been one of the first proteins structures ever reconstructed at a resolution of 4.7 Å[88].

Sample loading

Few microliters of the protein in solution are deposited on a special grid, consisting of 60 µm squares made by 1-2 µm holes in a carbon film supported by a metal (gold[88] or carbon) frame. When applied, the protein particles distribute within the grid holes in random orientations[89,90].

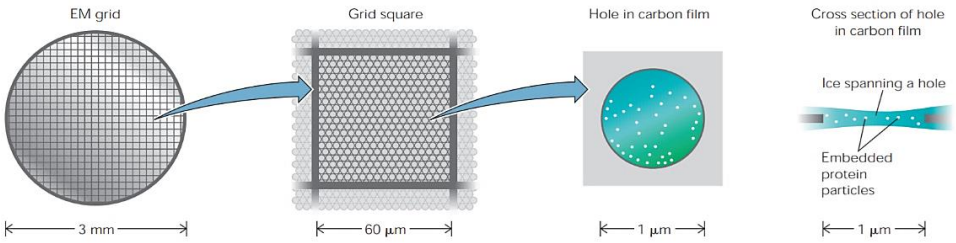


Figure 16 3-mm grid is perforated with holes 1–2 μm in diameter. The particles of interest are thus embedded in a thin layer of ice and are distributed in random orientations [90].

The grid is then rapidly plunged into a cryogen such as liquid ethane (maintained below -160° C to prevent the formation of ice crystals) that freezes and traps the particles in a thin film of vitreous ice[91, 92].

The grid containing the nanoparticles of interest is then loaded on a special specimen holder and is inserted in the microscope.

The cryo-electron microscope

The imagining of a sample in transmission mode is governed by the same physical principles using an optical or an electron microscope[93,94]. Briefly, in both instruments (Figure 17) a coherent and monochromatic radiation (photons of the visible light or electrons) is generated by a source (a laser/lamp or an electron gun) and is collimated through a first pair of lenses (optical lenses made by quartz or electromagnetic lenses made by electromagnets), called *condenser lenses*, on the specimen plane. After the interaction with the specimen, a second and third pairs of lenses, called *objective* and *eyepiece/projector lenses*, focus the transmitted radiation on the image recording system (a photomultiplier tube for light microscopy or a CCD camera for both light and electron microscopy).

The electron microscope column is maintained at high vacuum to avoid scattering between electrons and air molecules.
The diffraction limit of a generic imaging system is in fact equal to $d = \frac{\lambda}{2NA}$ (where *NA* is the numerical aperture of the objective lens): as such, the two microscopes differ deeply in the length scales that they can explore. Even with the most recent super-resolution light microscopy techniques (e.g. STED, PALM, STORM …), it is not possible to have resolutions better than 50 nm[95]. Electrons instead, behaving like waves, have a wavelength of the order of Angstroms: as such, with cryo-EM specimens can be imaged with Å resolution (Figure 18).

Thus, cryo-EM data of proteins are constituted by frames resulting from electrons interaction with the sample. More specifically, these images are bidimensional projections of the ice-embedded vitrified specimen seen from random points of views[79,96] (Figure 19).

In order to avoid the radiation damage on biological matter, it is mandatory to use a low-dose of electrons (*i.e.* a maximum of 20 $e^-/Å^2$ [93]): the acquired cryo-EM images are thus characterized by very low contrast and high noise.

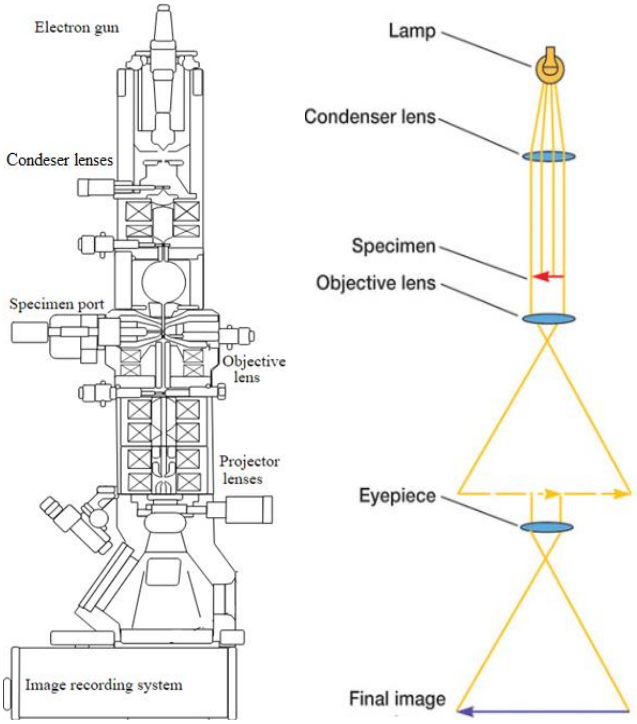


Figure 17 Comparison between an electron (**left** [93]) and a light (**right** [94]) microscope.

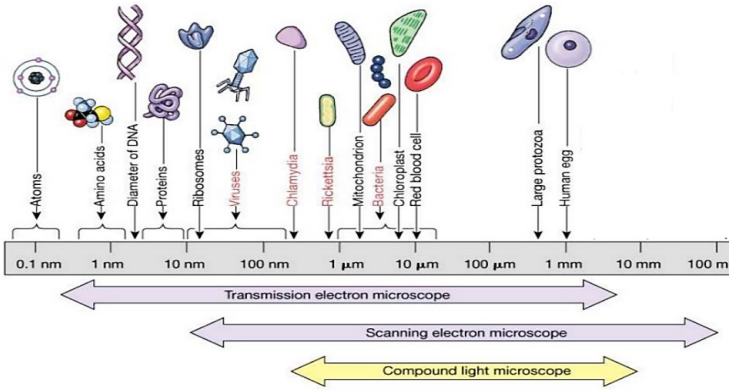


Figure 18 Resolutions that can be reached with different microscopes [94].

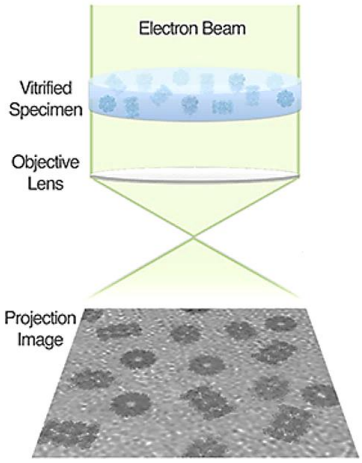


Figure 19 cryo-EM data of proteins are images of repeated, randomly oriented 2D projections of the ice-embedded specimen, whose 3D structure is unknown [79].

Image recording system

Until 2013, it was common to have charge-coupled device (CCD) cameras as image recording devices of an electron microscope, but their resolution was quite poor and their frame rates were slow[83]. While CCD cameras convert electrons in photons to record images, the more recently developed direct detectors devices (DDD) spot the electrons straightforwardly, thus allowing much higher signal to noise ratios in images. Their high frame-rates, moreover, allow to record many frames per second: DDD data are "movies" (called *micrographs*) constituted by subframes, during which the particles move because of electron-beam induced motion[84] (Figure 20). This new mode of data collection is the breakthrough that allowed the so-called "resolution-revolution" in cryo-EM, *i.e.* near-atomic information for protein structure determination.

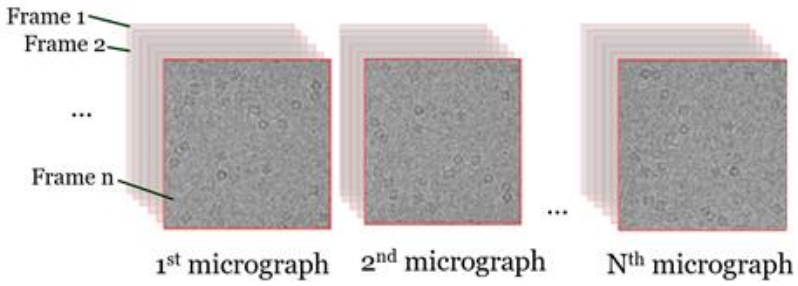


Figure 20 Cryo-EM raw data consist in a collection of N micrographs, each made by n subframes during which the particles move because of electron beam-induced motion. (Image modified [97])

Single-particle data analysis

Single-particle is an analysis technique for cryo-EM data based on the principle that the raw images are constituted by "isolated, unordered particles with identical structure"[93].

As discussed above, the micrographs are constituted by noisy multiple point of views of the same protein: the 3D structure is obtained thanks to single-particle image-processing methods based on deep learning algorithms, that require GPUs and large computer clusters for calculations.

Briefly, the single-particle cryo-EM workflow from raw micrographs to a possible 3D high-resolution structure can be summed-up as follows:

i) **motion correction**: correction of the frames constituting a single micrograph for the electron beam-induced motion through particles tracking algorithms; a single, much more contrasted micrograph is obtained [98,99].

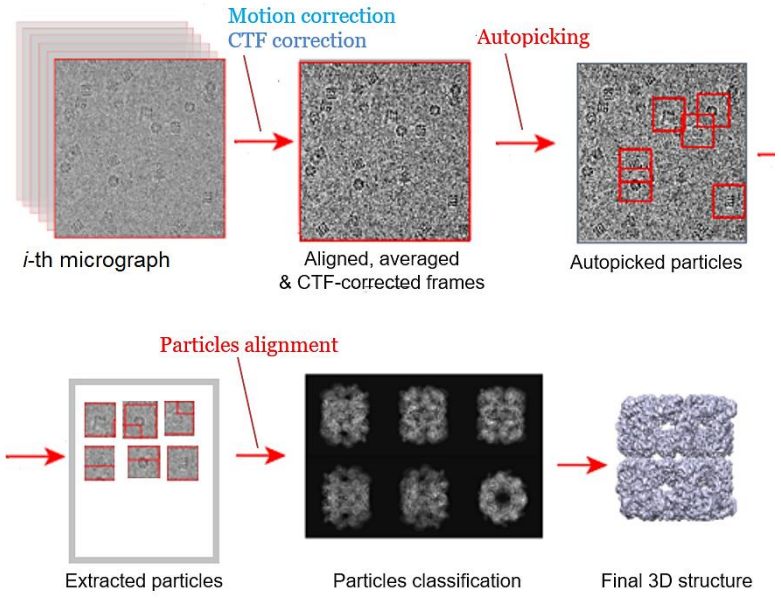


Figure 21 Single-particle workflow for analyzing cryo-EM data. (Image modified [97])

***ii)* CTF correction:** calculation of the Contrast Transfer Function for every micrograph[100]. This function modulates the image contrast in Fourier space and must be calculated to restore the actual contrast of the data. More formally, the CTF is the Fourier Transform of the Point Spread Function (PSF) of an optical system; they both are needed to reconstruct the information (in terms of real distances for PSF, of spatial frequencies for CTF) about an object that is imaged through an instrument[93]. Through CTF fit, the mean defocus of micrographs can be calculated.

***iii)* Autopicking of the particles and extraction**: the algorithm automatically detects the particles of interest in the micrographs. This procedure is generally based on a reference (as, for example, a set of ≈1000

manually-picked particles). In the end, a set of randomly oriented, very noisy and contrast-less particles is extracted from every micrograph[101].

***iv)* Particles alignment**: particles are aligned to a reference by rotations and translations through cross-correlation maximizations[102]. Thus, the set of particles obtained at the previous point is now in the same orientation.

***v)* Particles classification**: once aligned, particles can be compared to each other and then classified. This is one of the most computationally expensive parts of single-particle analysis: sorting algorithms are used to identify clusters of particles having similar characteristics, based on a reference (for example, a structure of a similar protein) or not. These classifications might be obtained in 2D or 3D[103].

***vi)* Three-dimensional model reconstruction:** the algorithm calculates the Euler angles between the classes and arranges all of them in a final 3D structure[102]. This computationally-expensive procedure is based on the central projection theorem (for which the 2D Fourier Transforms of two projected images share a line in the 3D Fourier space[96]). The map is then sharpened through B factors calculations; its resolution is estimated by the Fourier Shell Correlation between volumes reconstructed from two independent random halves of the dataset[93].

At the end, the obtained map may be validated through standard parameters calculated by specific programs[104]. It might also be fitted with crystallographic structures, or vice versa, the crystallographic structures might be modeled on

experimental, more native-like cryo-EM data to check for conformational changes or flexible parts in the biological assembly.

The workflow here shown will be followed throughout all chapter 4 to analyze the cryo-EM data obtained from three different samples.

2. Aim of the work

The principal aim of this thesis is the structural determination of the residues involved in the complex between the human Ferritin H-chain (HFt) and the human Transferrin Receptor 1 (hTfR1 or CD71) at near to atomic resolution.

HFt nanoparticles represent one of the most appropriate vectors for cellular delivery of molecules thanks to their internalization by CD71[20], a transmembrane receptor overexpressed in most cancer cell types[39,55,65,75,76,77,78] (Figure 13). The two proteins have been shown to interact very tightly: SPR data proved their nanomolar affinity[38]. As such, literature is continuously enriched by successful biomedical applications of this interaction, in particular for cancer treatment[38,40,41,42,43,44,47,48,51,52]; nevertheless, the epitopes of their recognition are to date unknown.

Near atomic structural information is fundamental to pinpoint the role of single amino acids directly involved in the HFt-hTfR1 interaction. To this end, we exploited single-particle cryo-electron microscopy (cryo-EM): this technique is currently reaching X-ray crystallography resolutions[79,81] and can be used as a more convenient method to determine atomic structures of protein complexes. We took advantage of modern cryo-electron microscopes (a Titan Halo and a Titan Krios, both equipped with direct detectors) and GPU-equipped computer

clusters for single particle analysis calculations (for which we have followed the workflow described in Figure 21). This investigation is described in chapter 4.1. The structural information here shown will likely represent a fundamental advance for all HFt-related applications in the field of nanotechnology.

Through the CD71-targeted distribution of specific payloads, the cellular internalization property of HFt might be used for a plethora of purposes, ranging from anticancer therapy to image-enhancement for diagnostics. The side aims of this thesis deal with cryo-EM structural analysis of two different Ferritins chimeras, whose constructs have been designed and optimized by our group.
The first (chapter 4.2) is the Humanized *Archaeoglobus fulgidus* Ferritin (HumAfFt), engineered to retain both the reversible, cations-induced association of the archaeal ferritin and the internalization in human cells typical of HFt through CD71 recognition.
The second (chapter 4.3) is the Lanthanide binding Ferritin (HFt-LBT Tb(III)), in which a high Tb^{3+} affinity sequence has been genetically fused to the mouse H-chain Ferritin structure to provide a construct acting both as carrier targeted to CD71 and as a FRET sensitizer by an appropriate antenna effect.

Throughout all chapter 4, the use of cryo-EM has been demonstrated to be crucial to obtain significant structural information that could not be achieved with other experimental techniques.

3. Materials and methods

3.1 Cloning, expression and purification of the proteins

Human HFt and Mutants A, B, C

The genes encoding the human apo-H-chain ferritin (HFt) and mutants A (Q14A/D15A/R22A) and B (F81A/Q83A) were designed, synthesized, optimized for *Escherichia coli* codon usage and cloned in pET11b vector by Geneart (Geneart AG). Mutant C (Q14A/D15A/R22A/F81A/Q83A) was produced by using the Quick Change Lightning Kit (Agilent Technologies, Santa Clara, California) according to the manufacturer's instructions and using mutant A as a template. All apo-HFt variants were expressed in *E. Coli* and purified as previously reported[47].

Briefly, samples dialyzed overnight against phosphate-buffered saline (PBS) pH 7.5 were loaded on a strong anion-exchange column, HiTrap Q Sepharose High Performance (GE Healthcare, Boston, MA, USA) previously equilibrated with the same buffer. In these conditions, HFt samples eluted from the column,

whereas other E. coli proteins and DNA contaminants did not. The recovered HFt samples were ultracentrifugated at 35 000 rpm for 55 min at 6 °C using a BeckmanL8-70M ultracentrifuge (Beckman Coultier). The recovered supernatant was then precipitated using ammonium sulfate at 65% saturation (w/v). The pellet was resuspended and dialyzed overnight against PBS pH 7.5, pooled, and concentrated by means of 30 kDa Amicon Ultra-15 centrifugal filter devices (Millipore, Billerica, MA, USA). Finally, the nanocages were sterile filtered and stored at 4 °C. Typical yields were 100 mg of pure proteins per 1 L of bacteria culture. The purity of all the preparations was assessed by SDS-PAGE, running the nanocages on 15% gels and staining the protein with Coomassie brilliant blue. Protein concentrations were determined spectrophotometrically at 280 nm, using a molar extinction coefficient (on a 24-mer basis) of $4.56 \times 10^5\ M^{-1}\,cm^{-1}$ (ProtParam sofware, http//www.expasy.org).

CD71

The gene encoding the ectodomain of human CD71 (residues 121-760) was extracted by PCR from the plasmid pAcGP67A-TfR[105] (Addgene, Cambridge, MA) and BamHI/EcoRI inserted using the Gibson cloning method and fused at the 3' of the Kozak sequence of the pα−H mammalian expression vector modified by the addition of the hydrophobic leader peptide from the baculovirus protein gp67 (pα−H BiP). An octa-histidine tag was also placed at the C-terminus of the CD71 gene. The expression plasmid pα−H BiP/TfR1 was transiently transfected in HEK293 using polyethylenimine (PEI) as transfection agent. Cells were grown in FreeStyle 293 expression media (ThermoFisher Scientific, Hampton, USA) supplemented with 1% of fetal bovine serum (FBS) at 37°C in a humidified atmosphere of 5% CO_2. After 96 hours, cells were

harvested and CD71 was purified from supernatants using Ni- or Co-affinity chromatography. Supernatant was filtered and incubated with the resin after addition of 50 mM sodium phosphate buffer, pH 8.0, together with 200 mM NaCl and 20 mM imidazole; 250 mM imidazole was used to elute CD71. The protein sample was stored at -80°C in 50 mM sodium phosphate, 200 mM NaCl, pH 8. Quality and quantity of purified protein was evaluated by SDS-PAGE and UV/visible spectra using the theoretical $\varepsilon_{280\ nm}$ 96260 $M^{-1}cm^{-1}$.

"Humanized" *A.Fulgidus* ferritin (HumAfFt)

The gene encoding for a mutated ferritin from *Archaeoglobus fulgidus* was synthesised by GeneArt (ThermoFisher) and subcloned into a pET22b vector (Novagen) between the restriction sites NdeI and HindIII at 5′ and 3′ respectively. The recombinant plasmid was transformed into BL21(DE3) E. coli strain for protein expression.

E. coli cells, containing the HumAfFt plasmid, were grown and induced with 1 mM IPTG (isopropyl-β-D-1-thiogalactopyranoside) at OD_{600} = 0.6. Cells were harvested by centrifugation 3 hours post induction at 37 °C. Cells harvested from 1 L culture were resuspended in 20 mM HEPES buffer, pH 7.5, containing 200 mM NaCl, 1 mM TCEP (tris(2-carboxiethyl)phosphine), and a cOmplete Mini Protease Inhibitor Cocktail Tablet (Roche). Cells were disrupted by sonication and the soluble fraction was purified by heat treatment at 78 °C for 10 minutes. Denatured E. coli proteins were removed by centrifugation at 15 000 rpm at 4 °C for 1 hour. The soluble protein was further purified by ammonium sulfate precipitation. The precipitated fraction at 70% ammonium sulfate was resuspended in 20 mM HEPES, 50 mM $MgCl_2$, pH 7.5 and dialyzed versus the same buffer. As a final purification step, the protein was loaded onto

a HiLoad 26/600 Superdex 200 pg column previously equilibrated in the same buffer using an ÄKTA-Prime system (GE Healthcare). The purified protein was concentrated to obtain the final protein preparation of 1 mg mL^{-1} and protein concentration was calculated by measuring the UV spectrum using an extinction coefficient of 32 400 M^{-1} cm^{-1}. Protein yield was ~40 mg L^{-1} culture.

Lanthanide binding Ferritin (HFt-LBT Tb(III))

A synthetic gene encoding for mouse H chain ferritin (HFt) fused with a lanthanide binding peptide (LBT) was designed, synthesized, and optimized for *Escherichia coli* codon usage by Geneart (Geneart AG). The last five C-terminal aminoacids of H chain ferritin were replaced by the GSG spacer sequence, followed by the LBT sequence YIDTNNDGWIEGDELLA[106]. As such, the LBT sequence is designed to be located within the inner cavity, as a prolongation of the inward directed C-terminal region (or E helix in ferritin secondary structure nomenclature). The resulting HFt-LBT construct, was subcloned into a pET22-b vector (Novagene) between NdeI/XhoI restriction sites. A stop codon was inserted before the His-tag region to avoid transcription of the unwanted tag. HFt-LBT was overexpressed in *Escherichia coli* BL21 cells upon induction with 1mM IPTG (Isopropyl-β-D-1-thiogalactopyranoside) at OD600 = 0.6. Cells were harvested by centrifugation 16 hours post induction at 37°C. Cells harvested from 1 L culture were resuspended in 20 mM HEPES buffer, pH 7.5, containing 300 mM NaCl, 1 mM TCEP, and a cOmplete™ Mini Protease Inhibitor Cocktail Tablet (Roche). Cells were disrupted by sonication and the soluble fraction was purified by heat treatment at 78°C for 10 minutes. Denatured *E. coli* proteins were removed by centrifugation at 12000 rpm at 4°C for 1 hour. The soluble protein was further purified by ammonium sulfate

precipitation. The precipitated fraction at 70% ammonium sulfate was resuspended in 20 mM HEPES, 150 mM NaCl, pH 7.5 and dialyzed versus the same buffer. As final purification step, the protein was loaded onto a HiLoad 26/600 Superdex 200 pg column previously equilibrated in the same buffer using an ÄKTA-Prime system (GE Healthcare). Purified protein was concentrated to obtain the final protein preparation of 1 mg/mL and protein concentration was calculated by measuring the UV spectrum using an extinction coefficient of $26000 M^{-1} cm^{-1}$. A small amount of higher molecular weight aggregates was observed in all ferritins preparations (less than 5%), and attributed to the presence of intermolecular disulphide bridges, possibly involving the surface exposed Cys103. As indicated Protein yield was ~50 mg/L culture. The expected molecular weight of 22662 Da was confirmed by MALDI-TOF Mass Spectrometry (Figure 24).

3.2 Samples preparation for cryo-EM acquisition

CD71/HFt complex

In vitro, incubation of HFt and CD71 at different stoichiometric ratios (HFt:CD71 = 1:1, 1:12, 1:24), temperatures (37 °C, 20 °C, 4 °C) and incubation times (60', 30', overnight) results in protein aggregation due to the presence of multiple binding sites on both HFt and CD71, forming insoluble precipitates. Several attempts in the stabilization of the sample with cryo-compatible Polyethylene Glycol (PEG) at different molecular weights (PEG6000 and

PEG8000) and concentrations (ranging from 1% to 5%) resulted in the same outcome. We managed to isolate a soluble sample of CD71/HFt complex by means of a pull-down experiment, as already established[20]. 250 µg 8xHis-tag fused CD71 was incubated with 100 µL of TALON resin (TALON Superflow Metal Affinity Resin, Ge Healthcare, UK) in 25 mM Tris·HCl, 150 mM NaCl, 1% PEG8000 and 10 mM Imidazole, pH 7.2 (buffer A), for 60 minutes at 4°C, under rotation. After several washes with buffer A, CD71-conjugated beads were incubated with 950 µg of HFt for 90 minutes at 4°C under rotation. The beads were washed increasing imidazole concentration in buffer A up to 30 mM, and the complex was eluted in 350 µL using 290 mM imidazole in buffer A. As control, CD71-unconjugated beads were also incubated with 950 µg of HFt for 90 minutes, following the same procedure adopted to isolate the complex. The pull-down assay final samples were analyzed by SDS/PAGE:

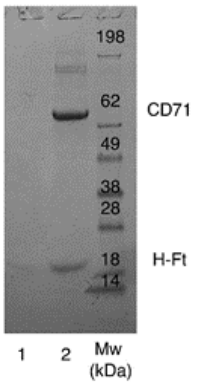


Figure 22 HFt-CD71 pull-down assay analyzed by SDS/PAGE. *Lane 1*: Beads+HFt. *Lane 2:* Beads-8HisCD71 + HFt. *Lane 3*: Molecular weights.

Purified CD71/HFt complex eluted from Talon resin at 0.2 mg/mL was directly used for grid preparation. Two datasets were collected using the same batch of grids to obtain the final map.

HumAfFt

3 µL of 1.2 mg/ml HumAfFt in 20 mM HEPES (pH 7.4) and 50 mM $MgCl_2$ was directly applied to the holey-gold grids. In this buffer the protein is an assembled 24-mer.

HFt-LBT Tb(III)

3 µl of a solution of 20 mM HEPES and 150 mM NaCl, pH 7.5, containing 1 µM (24-mer concentration) HFt-LBT Tb(III) complex was applied to holey-gold grids.

3.3 Grids preparation for cryo-EM

The procedure here described follows the discussion in session 1.3 (Figure 16). Holey-gold R0.9/1 (dataset 1 of CD71/HFt complex, HumAfFt and HFt-LBT Tb(III)) and holey-carbon R1.2/1.3 covered by 2 nm film of carbon (dataset 2 of CD71/HFt complex) grids from Quantifoil Micro Tools, GmbH, were prepared as described[88]. Grid surfaces were treated with plasma cleaning (Fischione Instruments Plasma Cleaner) using a mixture of 75% Ar and 25% O_2 for 60 s before applying 3 µl of sample.

We applied the samples on the grids using a Vitrobot Mark IV (FEI, Hillsboro)[92] kept at 100% humidity and 4°C.

The screening of several blotting conditions revealed that the time between sample application to the grid and plunge into ethane affects enormously the number of particles per field and their distribution: this time is indicated as "waiting time". The waiting times for the different samples were: 135 s for

dataset 1 of CD71/HFt complex, 90 s for dataset 2 of CD71/HFt complex, 30 s for HumAfFt and HFt-LBT Tb(III).

Grids were then blotted on filter paper at force 4 for different times; this time is referred to as "blotting time". The blotting times used were: 1s for CD71/HFt complex (datasets 1 and 2), 4s for HumAfFt, 3s for HFt-LBT Tb(III). After blotting, grids were vitrified by rapidly plunging into liquid ethane at −180 °C.

3.4 Cryo-EM data collections and analysis

Data collections

The parts of cryo-EM microscopes follow the description given in 1.3 (Figure 17, Figure 19, Figure 20).

Micrographs of CD71/HFt complex dataset1 (690 micrographs), HumAfFt (50 micrographs) and HFt-LBT Tb(III) (350 micrographs) were imaged at CUNY Advanced Science Research Center (New York), using a FEI Titan Halo (ThermoFisher Scientific, Eindhoven) operating at 300 kV, while the specimen was maintained at liquid nitrogen temperatures using a Gatan 626 side entry cryo-holder (Gatan, Pleasanton). Datasets were collected with Leginon[107], an automated data collection system.

HumAfFt datasets were imaged with a Volta phase-plate (FEI, Eindhoven)[108] on a FEI Ceta camera (FEI, Eindhoven) with a camera pixel size of 14 µm, corresponding to a calibrated pixel size of 1.49 Å on the specimen scale and with a dose of 18 $e^{-}Å^{-2}$.

CD71/HFt complex (dataset 1) and HFt-LBT Tb(III) data were collected on a

K2 Summit direct detector camera (Gatan, Pleasanton) operating in super-resolution mode, with a calibrated pixel size of 1.15 Å on the object scale, a camera pixel size of 14 µm and a magnification of 94000×. Images were typically recorded with a defocus range between −0.7 and −3.0 µm and using a dose of electrons on the specimen plane ranging between 10 and 20 electrons/Å^2. Movies were acquired with a total exposure time of 12 sec (60 frames/micrograph), with exposure rate of 7.8 electrons/pix/sec.

A second dataset for CD71/HFt complex (573 micrographs) was imaged at European Synchrotron Radiation Facility (ESRF, Grenoble) using a 300 kV Titan Krios (ThermoFisher Scientific, Eindhoven). The dataset was collected automatically using EPU (ThermoFisher Scientific, Eindhoven). Images were recorded on a Gatan K2 Summit direct-detector camera (Gatan, Pleasanton) equipped with a Gatan Bioquantum LS/967 energy filter and operating in super-resolution mode, using a calibrated pixel size of 1.33 Å on the object scale. The defocus was between −1.0 and −3.0 µm. Movies were acquired in electron counting mode, the total exposure time was set to 12 sec (40 frames/micrograph), with exposure rate of 6.2 e^-/pix/sec.

Data analysis

Analysis of the datasets follow the single-particle workflow described in Figure 21. For CD71/HFt complex and HFt-LBT Tb(III) datasets, motion correction for CD71/HFt complex and was performed using MotionCor2[98]; CTF was calculated using Gctf[100]. All subsequent data analysis was carried out using RELION 2.0[109]; occasionally, for CD71/HFt complex, we also used cisTEM[110]. For HumAfFt dataset, nor motion nor CTF correction were applied: the first could not be due since the detector was a CCD camera, the second was not

necessary because the data were collected within 200 nm of focus and the first CTF zero crossing was well beyond the achievable resolution of the dataset. Data analysis was carried out using RELION 1.4.

Cryo-EM maps resolution estimations and sharpening

The overall resolution of the maps was estimated with the Fourier shell correlation (FSC) = 0.143 criterion, based on the 'gold-standard' protocol[111,112,113]. The input map was corrected for the modulation transfer function (MTF) of K2 detector and sharpened using negative temperature B factors as calculated by RELION. When needed, local resolution was estimated using ResMap[114].

The electron density maps for HFt-CD71 complex were sharpened using the autosharpen tool in Phenix[115].

UCSF Chimera[59], Coot[116] and PyMOL (http://pymol.sourceforge.net) were used for graphical visualizations of the electron density maps and crystallographic structures.

3.5 X-ray Crystallography for structure determination

CD71/HFt complex structures refined in cryoEM maps

The program Chimera[59] was initially employed to rigid body fit the crystal structures of HFt (PDB code 3AJO) and CD71 (PDB code 3KAS) into the sharpened electron density map obtained at 5.5 Å resolution, which was further

refined as rigid body with Phenix *real_space_refinement*[115]. In this map, both CD71 and HFt were visible at the level of their secondary structures. The resulting CD71/HFt model was then rigid body fitted into the sharpened 3.9 Å density map using Chimera. Model building of HFt and of the CD71 contact region was performed manually using Coot[116]. The crystallographic model of the CD71/HFt complex was refined once again using the same *real_space_refinement* routine against the overall map at 3.9 Å, imposing secondary structure and Ramachandran restraints. Final visual inspection was performed in COOT to manually correct Ramachandran outliers. The final model was validated using MolProbity[117] and EMRinger [104]: their parameters can be found in Appendix 6.1 (Figure S3).

HumAfFt

The purified protein was concentrated to 20 mg mL^{-1} and initial crystallization screening was performed using a Phenix Robot. Crystals were obtained by mixing in a 2 µL hanging drop 1 mM of the purified protein with a solution containing 22% (vol/vol) polyacrylic acid PAA, 0.1 M Tris, 0.02 M $MgCl_2$, pH 7.4, at 25 °C within a week, cryo-protected by increasing the precipitant concentration and flash-frozen in liquid nitrogen. Diffraction data have been collected at ID23-2 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data were processed with XDS[118] and scaled with Aimless[119] at a final resolution of 2.87 Å. The structure was solved by Molecular Replacement with MolRep (ccp4 suite) using the open pore structure AfFt (pdb code 1S3Q) as the search model. Model Building and refinement were done using Coot[116] and Refmac5 [120], respectively. The final model was analyzed with PROCHECK[121] and Molprobity[117]. The Ramachandran plot showed that

97.8% of residues are in preferred regions, 2.2% in allowed regions and no outlier is observed. The final atomic coordinates and structure factors were deposited with the PDB Data Bank with accession code: 5LS9. Complete data collection and refinement statistics are reported in Appendix 6.2 (Figure S6).

The overall folding corresponded to the wild type structure (pdb code 1S3Q) with a *rmsd* value of 0.4 Å and displayed four wide triangular pores on the surface. B factors analysis shows a mean B factor of 66 $Å^2$ except for the loop region between the D and E helices, which displays higher B factors and a poor electron density on the side chains from 146 to 151 residues. Two magnesium ions have been positioned and successfully refined in the ASU, and they are located in the ferroxidasic site of two different subunits. In both, one magnesium ion is coordinated with both OE1 and OE2 of Glu19 (at 2.5 Å and 2.8 Å distance, respectively), with Glu52 (OE1) at 2.5 Å, with Gln129 (OE1) at 2.6 Å and with a water molecule at 2.6 Å distance. In the other chains, a water molecule has been modelled in the Fo-Fc map and successfully refined in the same position. The loop region, including the conserved terminal turns, spans from aminoacid 68 to aminoacid 86. At the dimeric interface between the two antiparallel BC loops, the main interactions are a hydrogen bond between Arg69(NH2) and Ser80 (O) at 2.8Å distance and two salt bridges, namely Lys71 (NZ) - Glu77 (OE1) at 3.0Å distance and Glu81 (OE2)-Arg69 (NH1) at 2.7Å distance. A weak salt bridge is established by Glu81 (OE2) and Arg69 (NH2), at a distance of 3.0Å. The salt bridge between Lys71 and Glu77 observed in HuHF, is absent in HumAfFt since these positions were mutated into a phenylalanine (Phe71) and a lysine (Lys77). Other interactions are conserved between HFt and HumAfFt and displayed the same distances.

apoHFt-LBT and HFt-LBT Tb(III)

Crystals of wild type apoHFt-LBT and HFt-LBT in complex with Tb(III) were obtained by mixing in a 2 µl hanging drop the purified protein at 15 mg/ml with a solution containing 1.8/2.0 M ammonium sulphate and 0.1 M Tris, pH 8.5, at 25° C within a week, cryo-protected by extensively washing the crystals in sodium malonate and flash-frozen in liquid nitrogen. Diffraction data were collected at $\lambda = 1.0$ Å and 1.4 Å respectively at XRD1 beamline at the Elettra Synchrotron, Trieste, Italy. Data were processed with XDS[118] and scaled with Aimless[119] at a final resolution of 2.85 and 2.65 Å. They both crystallized in I222 space group with 24 identical subunits in the asymmetric unit (ASU) with a solvent content of 64.7%.

The structures were solved by Molecular Replacement with Phaser[122] using the structure of mouse H-chain ferritin (PDB code 3WNW) without waters and ligands as a starting model. Model building and refinement were done using Coot[116] and Refmac5 [120]. Anomalous difference electron density map was calculated from the HFt-LBT Tb(III) crystal diffraction data, collected at the Tb emission peak ($\lambda = 1.4$ Å). The map has been generated using the Fast Fourier Transform Program belonging to the CCP4 suite. The final model was analyzed with PROCHECK[121] and Molprobity[117]. Ramachandran Plot showed more than 98% residues are in preferred regions and no outlier is observed in both structures. The validation of metal binding sites was performed using CheckMyMetal web server. Final atomic coordinates, structure factors of apoHFt-LBT and HFt-LBT with Tb(III) were deposited in the PDB Data Bank with accession code 5OBA and 5OBB, respectively.

The model building process allowed to position residues just from 3 to 176 since the LBT loop was not visible, probably because of the intrinsic high flexibility of the C-terminus region. When looking at the H-chain mouse ferritin structure,

each ferroxidase center and 3-fold axes displayed one and two Mg(II) ions, respectively, due to the crystallization buffer solution. In a similar way, in each ferroxidase center and 3-fold axes of wild type apoHFt-LBT structure, a significant electron density peak was indeed observed and ascribed to iron ions (refined with occupancy <100%), since, although no salt was in the crystallization solution, the X-ray fluorescence scan displayed the presence of residual iron atoms. As a matter of fact, residual iron (5 atoms per ferritin 24-mer) is consistently present in recombinant ferritin H-homopolymers, most likely due to iron uptake within bacterial environment. In addition, in 3WNW structure, one iron ion was modelled in each 4-fold axis while in HFt-LBT structure, one water molecule was positioned and successfully refined. When comparing the wild type apoHFt-LBT and HFt-LBT Tb(III), the LBT loop was still not visible while a very large electron density peak appeared in each ferroxidase center and in the eight 3-fold axes. In the X-ray emission scan, ranging from 4.0 to 21.0 KeV, the typical Xray emission lines of Terbium were clearly identified in the HFt-LBT Tb(III) crystals (Figure 23). Accordingly, the crystals of wild type apoHFt-LBT lacked these emission energies.

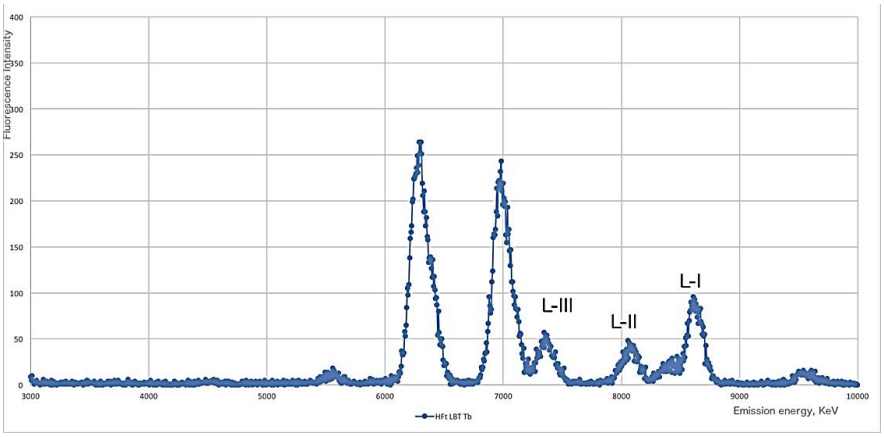


Figure 23 X-ray emission spectrum of HFt-LBT-Tb: L emission lines of Tb at 7.5, 8.2 and 8.5 keV are displayed. Additional lines at 6.3 and 7.0 keV are ascribed to iron ions.

The position of Tb(III) ions was clearly assigned by comparison of the 2Fo-Fc electron density maps of the HFt-LBT Tb(III) and the wild type apoHFt-LBT was used as a control. An additional demonstration of the Terbium assignments was obtained by anomalous difference electron density map, which clearly displayed the positions attributed to Terbium (Figure 62). In the ferroxidase center, each Terbium ion was successfully refined with 70-75% occupancy and was located in a trigonal planar coordination to OE1-Gln141 and to OE1 and OE2-Glu62, and to OE2-Glu27, in a range of 2.6-3.2 Å of distance. In the 3-fold center, the Terbium ion was successfully refined with 100% occupancy and it was tetrahedrally coordinated to OE1-Glu134 of the three subunits (at 2.2-2.4 Å distance). In each 4-fold channel, one water molecule was positioned and successfully refined as well as in the native protein. Complete data collection and refinement statistics are reported in Appendix 6.3 (Figure S9).

3.6 Samples composition assays

CD71/HFt SPR assay

The interactions between the immobilized N-terminal His-tagged CD71 (ligand) and HFt-based constructs (analytes) were measured by surface plasmon resonance (SPR) technique on a Biacore X100 instrument (Biacore, Uppsala, Sweden) according to the procedure previously reported[47]. Briefly, CD71 was immobilized on a Sensor Chip nitrilotriacetic acid (NTA) (GE Healthcare Europe GmbH) according to the manufacturer's instructions. The optimal experimental setup was determined and CD71 was injected at 22 μg/ml for 60s.

Analyte concentration was in the range of 1000–62.5 µg/ml. The sensor chip surface was regenerated using fresh histidine tagged protein after every cycle of the assay. The SPR assay was performed at 25 °C, at flow rate 30 µl/min; the association and dissociation phases were monitored for 180 and 600 s, respectively. Analytes were dissolved in degassed 10 mM Phosphate-buffered saline (PBS) at pH 7.4. To regenerate the chip, complete dissociation of the complexes was achieved by the addition of 10 mM HEPES, 150 mM NaCl, 350 mM EDTA and 0.005% (vol/vol) surfactant P20 (pH 8.3) for 30 sec before starting a new cycle. The k_{on} and k_{off} rates as well as the dissociation constant (K_D) were estimated using the Biacore X100 evaluation software according to a 1:1 binding model or alternatively a heterogeneous analyte binding model (Table 2). All experimental data shown in Figure 45 are reported at analyte concentration of 1 mg/ml. Fits are reported as black lines corresponding to a heterogeneous analyte binding model for wild type, Mutant A and Mutant B respectively, and to a simple 1:1 kinetic model for Mutant C.

HumAfFt self-assembly assessment in solution

Size exclusion chromatography $MgCl_2$-mediated self-assembly (Figure 50) was studied by size exclusion chromatography (SEC) using a Superdex 200 26/600 GL column (GE Healthcare). The molecular size of HumAfFt was determined under different conditions by comparing the elution volume with that of standard proteins. Composition of the mobile phase was 25 mM HEPES (pH 7.5) with different $MgCl_2$ concentrations accordingly to the composition of the protein buffer.

TfR1 silencing in HeLa cells

HeLa cells were transfected with Lipofectamine® RNAiMAX (Life Technologies) following standard procedure. The siRNA against TfR1 was purchased from Sigma (product #NM003234, siRNA ID: SASI_Hs01_00059217). Cells were then incubated with 30µg/ml of HumAfFt-FITC, treated and analyzed as described in session 3.7. The FITC intensity was normalized for the FITC intensity of the scr samples (Figure 56 right).

HFt-LBT Tb(III) MALDI-TOF mass spectrometry

40 µl of protein sample were desalted on a C8 Empore Disk (3M, Minneapolis, MN) homemade stage tip and resuspended in 3 µl formic acid 1%. 1 µl was spotted on a MALDI sample plate and allowed to air dry. Recrystallized sinapinic acid (SA matrix from Thermo Fisher Scientific) was prepared at a concentration of 5 mg/ml in 50:50 acetonitrile/water (0.1% Formic Acid) and spotted directly prior to insertion into the mass spectrometer. Matrix-assisted laser desorption ionization (MALDI) mass spectra were acquired on 4800 MALDI-TOF/ TOF mass spectrometer (Applied Biosystems, Foster City, CA) equipped with a nitrogen laser operated at 336 nm laser. Acquisitions were performed in linear mode averaging 2500 laser shots in a random, uniform pattern. Ions were accelerated with a 20 kV pulse, with a delayed extraction period of 860 ns. Spectra were generated by averaging between 500 and 2000 laser pulses in a mass range from 4 kDa to 50 kDa. Laser intensity was set to optimize the signal-to-noise ratio and the resolution of mass peaks of the analyte. All spectra were externally calibrated and processed via Data Explorer software (version 4.7).

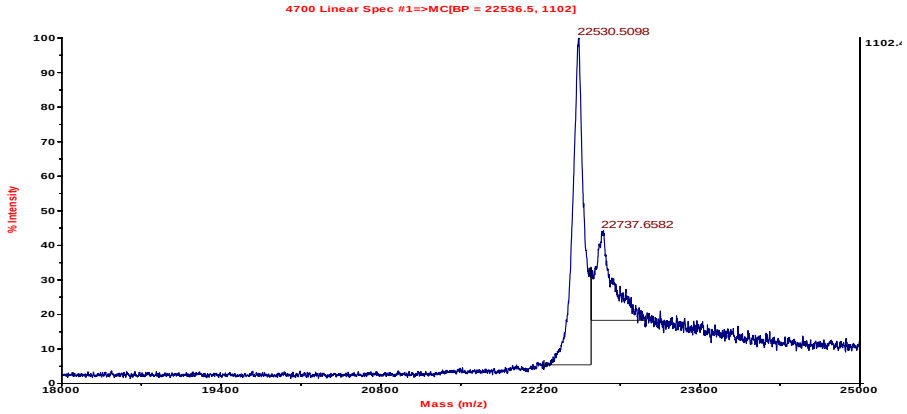


Figure 24 MALDI-TOF mass spectrum of HFt LBT. The molecular mass of 22531 Da corresponds to theoretical protein mass (22662 Da) subtracted of the contribution of a methionine (131 Da)

HFt-LBT Tb(III) Fluorescence spectroscopy

Fluorescence spectra and titrations were performed using FluoroMax 4 (Horiba) spectrofluorimeter with a Haake D8 refrigerated bath at 20 °C. Emission spectra were recorded between 450 and 560 nm, in order to include the luminescent maxima of Tb(III) (490 and 545 nm). The excitation wavelength was chosen at 295 nm to minimize the overlap of second order diffraction (570 nm) with the Tb(III) emission at 545 nm. Emission spectra were taken with excitation and emission band passes of 4 and 8 nm and corrected for the blank contribution and the instrument response at 295 nm in a quartz cell of 1 cm pathlength. Emission spectra were normalized to 1 at 545 nm. Fluorescence static spectra were performed using 1 µM (monomer) apoHFt-LBT and wild type apoHFt as a control in 100 mM MES buffer pH 6.4. A 50 mM $TbCl_3$ anhydrous powder (Sigma Aldrich) stock solution was also prepared in MES buffer at pH 6.4. Under these conditions, precipitation of Terbium hydroxides, easily formed around neutrality, is avoided. Fluorescence spectra of the protein Tb(III) complexes were recorded after 30 min incubation and after addition of 150 µM $TbCl_3$ in buffer solution in order to saturate all possible Tb(III) binding sites

both in HFt-LBT and HFt. Before recording spectra, protein solutions were exchanged with buffer (Terbium free) by centrifugal ultrafiltration on vivaspin MWCO 100 kDa concentrators (Sartorius) in order to remove unbound and weakly bound metal ions (5 exchange steps, 1×10 concentration each). Protein concentration was measured again and adjusted to the final concentration with buffer (1 μM monomer). Samples used for crystallographic and cryo-EM measurements were also prepared according to this procedure. Fluorescence titrations of HFt-LBT were carried out by adding incremental amounts (1.8 μL or multiples) of a 0.5 mM $TbCl_3$ stock solution in 0.1MMES buffer pH 6.4 to 3 mL of 1μM (monomer) protein in the same buffer under stirring. Emission spectra were recorded in 1 cm pathlength cuvette upon excitation at 295 nm at 25 °C, 30 minutes after addition of TbCl3 solution aliquots. Fluorescence intensity of HFt-LBT Tb(III) complex as a function of the Tb(III)/HFt-LBT ratio has been reported in Figure 59 and Figure **60**. Fluorescent intensity was recorded at 545 nm corrected for the dilution factor and normalized to the emission maximum.

3.7 Internalization assays of FITC-labelled proteins in cells

Proteins FITC labeling

When needed, proteins were labeled with fluorescein-isothiocyanide (FITC, ThermoFisher) according to the manufacturer's standard protocol. Briefly, 2 mg/mL of the purified protein was equilibrated in sodium carbonate buffer, pH

9.0. 50 µl of 1 mg/ml (2.6 mM) FITC in DMSO were added to 1 ml of protein solution and the reaction mix was incubated under stirring for 2 hours at room temperature. The non-reacted dye was removed by gel filtration chromatography and the fluorescent dye to protein ratio was determined by UV-spectroscopy. LC-MS spectrometry measurements all FITC-conjugated proteins confirmed that >40% of monomers (HFt mutants), >60% (HumAfFt) and >90% (HFt-LBT) are FITC labeled.

LC-MS measurements of FITC-labelled proteins

LC–MS was performed on protein samples reacted with 10 molar excess of FITC per ferritin monomer in comparison with unreacted protein. Protein samples were diluted with distilled water in the presence of 1 mM EDTA and loaded on a Waters Acquity UPLC connected to Waters Acquity Single Quad Detector. A BEH300 C_4 column was used: 1.7 µm, 2.1 × 150 mm at 220 nm observation wavelength; mobile phase: 0.1% TFA in water: MeCN (0.1% TFA); gradient 20:65% over 60 min; flow rate: 0.25 ml min^{-1}; MS mode was set at a scan range: m/z=200 – 2,000 (ES+); scan time: 0.5 s, electrospray source with a capillary voltage of 3.0 kV and a cone voltage of 45 V. N_2 gas was used as nebulizer and desolvation gas at a total flow of 200 l/hours. Ion series were generated by integration of the ultraviolet-absorbance (at 220 nm) chromatogram over 2.4–2.8 min range. Mass spectra were reconstructed from the ion series by using MassLynx software program.

Cell cultures maintenance

HeLa cells were grown at 37 °C in Dulbecco's Modified Eagle's Medium with Glutamax (DMEM, Gibco) supplemented with 10% FBS (Gibco) and 1%

Penstrep (100 U/ml penicillin and 100 µg/ml streptomycin solution; Gibco). DMEM without phenol red (Sigma) supplemented with Glutamax (Invitrogen), 10% FBS and 1% Penstrep (incubation medium) was used for apo-ferritins internalization assays by FACS and confocal microscopy.

The HeLa cell line stably expressing an inducible TagRFP-FUS protein (HeLaTagRFP) was generated by transfection with epB-Puro-TT-RFP-FUS wt plasmid and the piggyBac transposase vector [123] When needed, the cells were induced with doxycycline 0.2 µg mL^{-1}.

Human prostate cancer cell line DU-145 (ATCC HTB81), human colorectal cancer cell line HCT-116 (ATCC CCL-247), human breast cancer cell line MDA-MB-231 (ATCC HTB-26) and human ovarian cancer cell line SKOV-3 (ATCC HTB-77) were cultured in DMEM medium containing 10% FBS, 100 µg/ml streptomycin and 100 U/ml penicillin G in a humidified 37 °C incubator.

Internalization assays of Ferritins

Cells were seeded one day prior performing internalizations assays. Upon growing medium removal and rinse with PBS, confluent cells were incubated in non-supplemented DMEM medium containing FITC-ferritin nanoparticles (as specified in each experiment) in a humidified 37 °C incubator for the time indicated below. Control cells were treated in the same conditions without ferritins. Cells were then subjected to confocal microscopy and flow cytometry analysis.

These are the final proteins concentrations and incubation times used for the various experiments:

i) *Mutant A, Mutant B, Mutant C, AfFt, HFt:* 0.03 mg/ml for 2.5h (FACS acquisition) or for 20h (confocal microscopy) (Figure 46);

ii) *HumAfFt, Tf, AfFt:* 0.03 mg/ml for 1h, 3h or 20h (FACS acquisition) or for 20h (confocal microscopy) (Figure 57);

iii) *HFt-LBT:* 0.5 mg/ml incubated for 1h (FACS acquisition) or for 20h (confocal microscopy) (Figure 66).

Flow Activated Cytometry Analysis (FACS)

After internalization of FITC-ferritin nanoparticles, cells were washed twice with PBS, detached from the substrate using trypsin-EDTA (Euroclone), then rinsed twice with PBS and resuspended in BD-FACS flow buffer.

For HumAfFt experiment internalization assays, half of each sample was treated with Trypan Blue (TB; Sigma) to quench the FITC signal from membrane-bound nanoparticles that were not internalized; the quenching was performed with 0.04% TB for 10 min on ice.

Internalization of ferritins was evaluated with sample acquisition at the BD FACS Aria III (HFt-LBT data acquisitions) or BD LSRFortessa (BD Biosciences, San Jose, CA, USA) (all the other data acquisitions) equipped with a 488 nm laser and FACSDiva software (BD Biosciences version 6.1.3). Live cells were first gated by forward and side scatter area (FSC-A and SSC-A) plots, then detected in the green channel for FITC expression (530/30 nm filter) and side scatter parameter. As a baseline for FITC fluorescence, control cells not incubated with FITC-ferritins were used. The gate for the specific signal was set based on the control sample. Data were analyzed using FlowJo 9.3.4 software (Tree Star, Ashland, OR, USA).

Live imaging confocal microscopy of live cells

Before imaging, cells were washed twice with PBS to eliminate the unbound

FITC-ferritins and then replaced in an imaging medium (DMEM without phenol red, 10% FBS, 10 mM Hepes, Glutamax and penicillin–streptomycin solution) suitable for confocal imaging purposes.

For HFt, HumAfFt and Tf internalization, the confocal laser-scanning microscope used was an Olympus FV10i platform equipped with a built-in incubator. Images were acquired with a 60×/1.2NA water-immersion objective, LD lasers 473 nm and 559 nm, and filter sets for FITC and TRITC. Phase-contrast channels were acquired simultaneously.

HFt-LBT internalization images were acquired using an inverted confocal microscope IX70 FV 500 (Olympus), with 488 nm laser, 20x objective lens and emission filter 505-560 nm.

Image visualization and processing was performed using ImageJ software[124].

4 Results and discussion

4.1 Human HFt/CD71 complex

Literature is continuously supplemented with successful applications of the human HFt/CD71 complex as a bio-vehicle for targeted delivery of cargos or drugs; nevertheless, the epitopes involved for this recognition are unknown. As shown in section 4.2, in 2017 we identified that the BC loop in the HFt subunit appears to be fundamental for this interaction: when transplanted in AfFt, unable to recognize the human CD71, it is sufficient to induce binding of this chimeric protein to the receptor[46]. Moreover, the possibility of the existence of different epitopes for HFt and Transferrin (Tf), that binds CD71 on the helical and protease-like domains (Figure 10), has been largely proved by a scarce competition between the two ligands for CD71[20,125].

In vitro, the combination of HFt with CD71 at different stoichiometric ratios, temperatures and incubation times results in aggregation. This is probably due

to the presence of multiple binding sites on both HFt (24 identical subunits) and CD71 (2 subunits) that, when combined, form insoluble precipitates. An effort of loading this solution on holey-gold grids resulted in very crowded micrographs that were impossible to analyze with single-particle techniques.

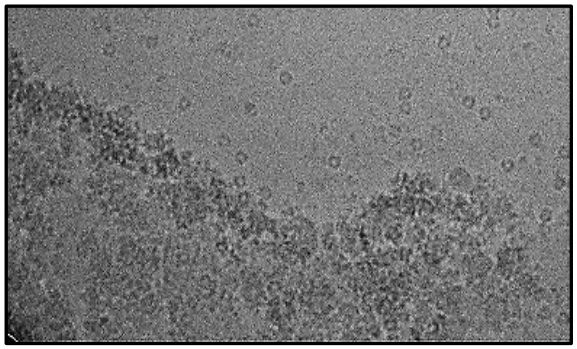


Figure 25 A first effort of loading HFt/CD71 sample resulted in crowded micrographs.

In order to have monodisperse particles on a micrograph, we isolated a soluble sample of CD71/HFt complex by means of a pull-down experiment[20]: issues of aggregation are thus overcome, resulting in successful micrographs in which the complex could be clearly seen even by naked eye (Figure 26).

Two datasets were collected from the same batch of sample, using two 300 kV microscopes, both equipped with a Direct Detector Camera (Gatan K2):
i) "**dataset 1**" was imaged with FEI Titan Halo at the Imaging Facility of CUNY Advanced Science Research Center (New York) (Figure 26, left);
ii) "**dataset 2**" was acquired at a FEI Titan Krios in the CM01 beamline of European Sinchrotron Radiation Facility (ESRF, Grenoble) (Figure 26, right).

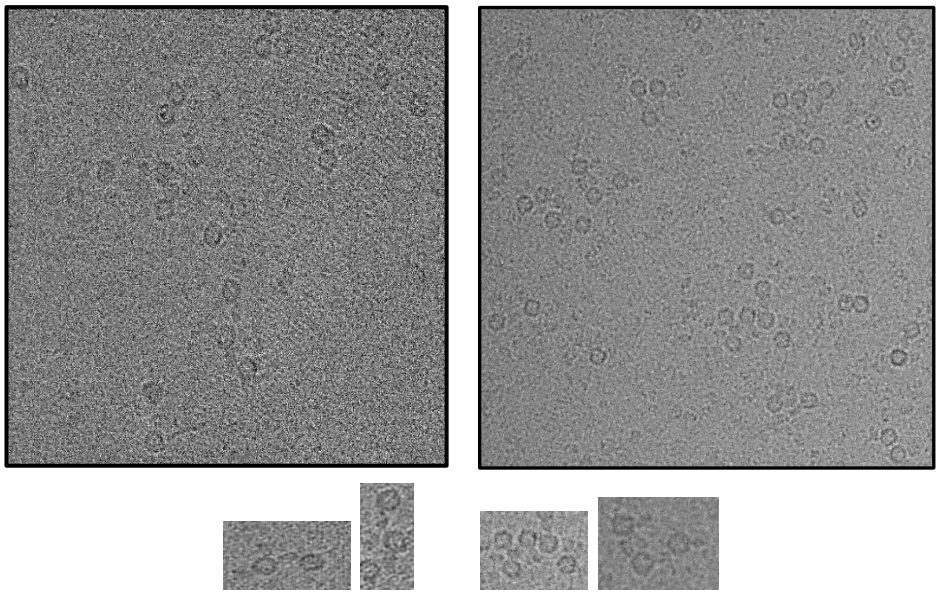


Figure 26 Representative micrographs of CD71/HFt complex with a DDD. In the lower part, zoomed views of particles in which the complex can be clearly seen by eye. **Left**: dataset 1 (FEI Titan Halo). **Right**: dataset 2 (FEI Titan Krios).

Cryo-EM data analysis

The complete data analysis workflow is shown in appendix 6.1 (Figure S1).

Particles collection and extraction.

Micrograph frames collected in both datasets (described in Materials and Methods, sections 3.3 and 3.4), were aligned for beam-induced motion correction and drift with MotionCor2[98]; contrast transfer function was calculated using Gctf[100]. We kept only micrographs with CTF resolution limits better than 6 Å (dataset 1: 690 micrographs) or 5 Å (dataset 2: 573 micrographs).

All subsequent data analysis was carried out using RELION 2.0[109], using the procedures described in Section 1.3 (see Figure 21). More than 1100 particles for both datasets were manually picked to produce a reference for the automated

particle picking procedure implemented in RELION[101]. A total of 140,567 and 194,501 particles were automatically picked from dataset 1 and dataset 2, respectively, and extracted from the original micrographs.

2D Classifications.

After the extraction, particles of each datasets were subjected to 2D classification in two rounds; the first using K=100 classes, the second with 25 classes. The second round of dataset 1 resulted in a set of 27,690 particles, whose classes are shown here:

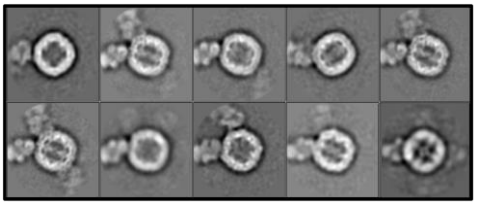


Figure 27 2D classification of dataset 1.

The second round of dataset 2 gave 73,700 good particles:

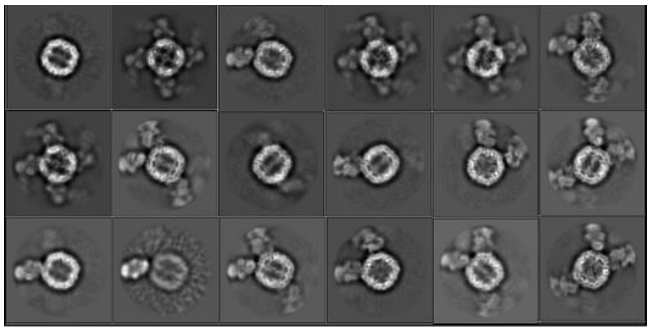


Figure 28 2D classification of dataset 2.

These class averages clearly show CD71/HFt complex in different orientations, revealing some extent of sample heterogeneity due to the presence of alternative populations endowed with different stoichiometry. Provided sufficiently

populated classes, Cryo-EM is anyway able to process heterogeneous data, differently from X-ray crystallography[126].

3D Classifications, Refinement and PostProcessing.

Particles belonging to the classes previously shown were subjected to 3D classification, using K=8 classes and no symmetry.

For dataset 1, the cryo-EM apo-ferritin map (EMDB code: EMD-2788) filtered at 60 Å was firstly used as reference model. This procedure gave only one class (corresponding to 32% of the dataset, *i.e.* 8,860 particles) where a 1:1=HFt:CD71 complex was clearly displayed:

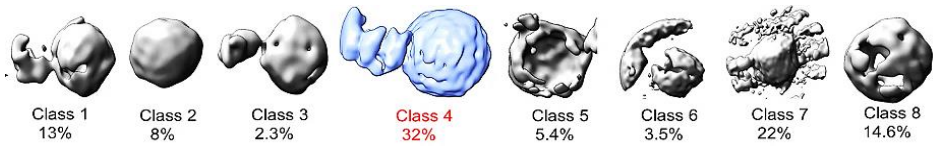


Figure 29 First-round 3D classification of dataset 1.

The resulting map of class 4 was used as reference (filtered at 60 Å) to run a second round of 3D classification on the same dataset, which resulted in two better resolved classes (24.5% + 38.2% = 62.7%, *i.e.* a total of 17,370 particles). Such a deep difference in classification efficiency reflects the dependence upon a starting model of the reference-based 3D classification procedure:

Figure 30 Second-round 3D classification of dataset 1.

These classes were selected for further refinement using the 3DAutorefine, applying a spherical mask of 290 Å diameter: this resulted in a map at 10.1 Å resolution. The post-processing yielded a 8.2 Å map.

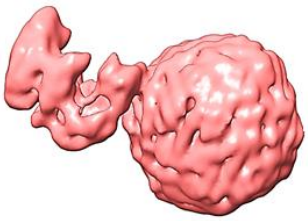


Figure 31 3D Refinement resulting map of dataset 1 (10.1 Å resolution).

This reconstruction of the complex was used as reference model, filtered at 40 Å, for the 3D classification of particles selected from 2D classification of dataset 2:

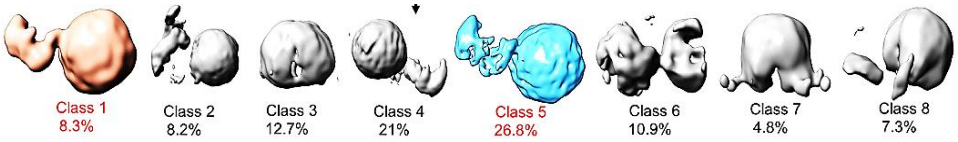


Figure 32 3D classification of dataset 2.

Good particles o(35.1% of the total) were selected for further refinement using the 3D Autorefine procedure (25,870 particles, spherical mask of 290 Å diameter) and yielded a 7.4 Å map (after PostProcessing, 5.5 Å).

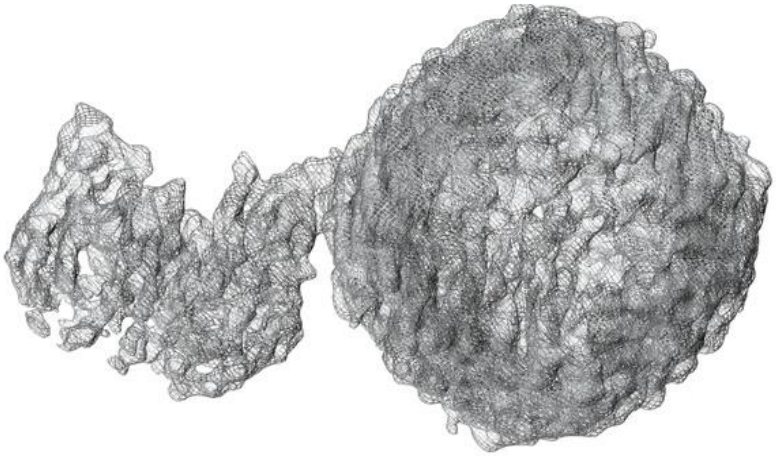


Figure 33 Post-processing map of dataset 2 (5.5 Å resolution).

The complete data analysis workflow of datasets 1 and 2 is outlined in appendix 6.1 (Figure S1).

Merging of the datasets.

With the aim to further increase the resolution of the contacting region between the two molecules, particles used for 3D Refinement of both datasets were joined (total 43,240 particles); however, resolution was not improved with respect to dataset 2:

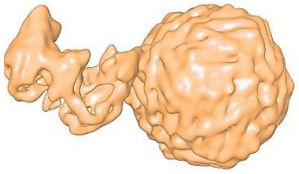


Figure 34 3D Refinement map of joined datasets (43.240 particles, 7.5 Å)

Those particles were subjected to one round of 3D Refinement, imposing O symmetry (*i.e.* Ft point-group symmetry) and applying a smaller mask diameter of 180 Å, to only include the ferritin molecule. This yielded a 4.8 Å map of ferritin re-centered to its center of symmetry:

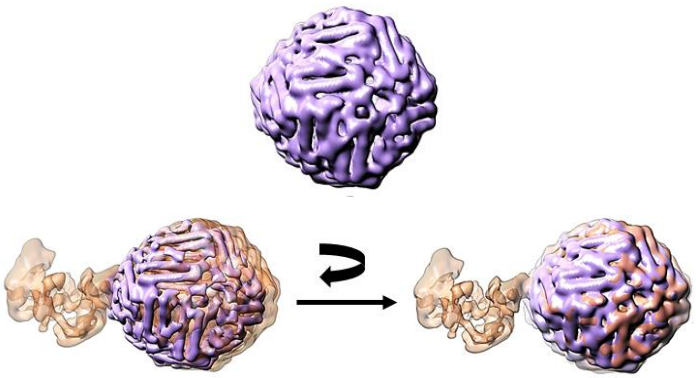


Figure 35 The complex map resulting from the joined particles (Figure 34, orange) was re-centered to its center of symmetry through comparison with the 4.8 Å map of Ft alone (top, purple).

We used the corresponding particle alignment parameters to perform symmetry expansion of this dataset using RELION (command: *relion_particle_symmetry_expand*). This procedure allowed to deal with the pseudo-symmetry of the complex by dividing particles in individual subunits bound to CD71 that could be further classified[126,127]. The dataset was thus artificially expanded according to the pseudo-symmetric O point group and enlarged 24-fold, resulting in 1,037,760 particles.

The particles obtained with this method were 3D Classified in 4 classes using C1 symmetry, no image alignment and a mask generated from the complex with a single receptor bound (displayed above in Figure 35): all the classes were subjected to a second round of 3D Classification, this time with 8 classes and local image alignment. This resulted that only two out of four classes (class 1, 39%, and class 3, 11%) allowed the identification of 3 subclasses of ferritin bearing a receptor (class 1, 8.4% of the former; class 5, 10% and class 6, 13%, of the latter):

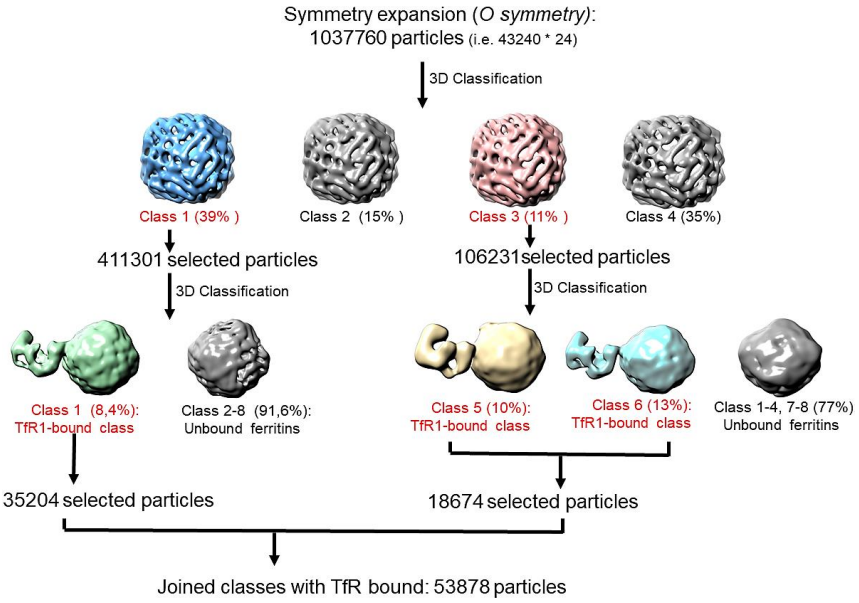


Figure 36 Overview of the data analysis workflow after symmetry expansion.

Therefore, three HFt bound subclasses were combined (total 53,878 particles) and refined, resulting in a map with a global resolution of 4.4 Å; the post-processing procedure, calculated with a mask only including ferritin and the receptor contact sites, was finally estimated to be 4.2 Å resolution.

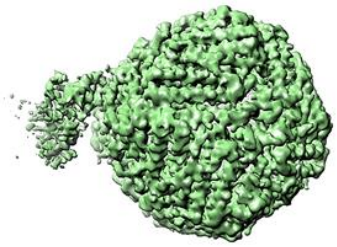


Figure 37 Post-Processing map of joined datasets (53.878 particles, 4.2 Å). Only Ferritin and CD71 binding region can be seen.

The refined particles were exported in cisTEM[110] to calculate the defocus of each particle. This method is more accurate than the average defocus values determined for whole micrographs by Gctf: it has been proved to be effective when particles are about 400 kDa or larger, as our case[110]. Indeed, the resulting final map resolution of HFt and CD71 contact region was improved from 4.2 up Å to 3.9 Å (Figure 39), clearly showing secondary structures and bulky side chains of interacting residues (Figure 38), which allowed more precise model building.

Statistical information of 5.5 Å and 3.9 Å density maps, FSC curves and resolution maps are in Appendix 6.1 (Figure S3, Figure S4).

Model building and refinement.

A rigid-body fit of the crystal structures was done using Chimera at first inside the 5.5 Å map (where the whole density of the receptor can be easily reconstructed, Figure 33), then into the 3.9 Å map (where the contact region was better refined and was modeled manually but CD71 electron density lacked

almost entirely, Figure 39). Given the high quality of the latter map, the side chains of residues exposed at the contacting interface could be modeled into the electron density without ambiguity.

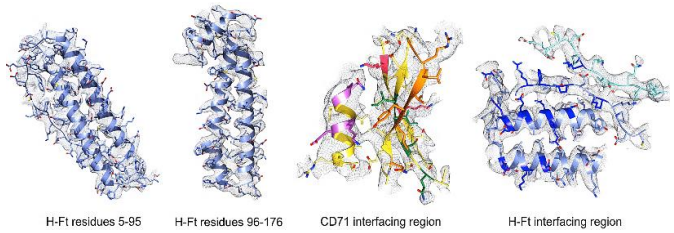


Figure 38 Side chains of crystallographic molecules can be modeled in 3.9 Å map (Figure 39).

The crystallographic model of the CD71/HFt complex was refined once again against the overall map at 3.9 Å, imposing secondary structure and Ramachandran restraints.

The final refined and validated cryo-EM map, with CD71 and HFt crystallographic structures fitted in it, is in Figure 39:

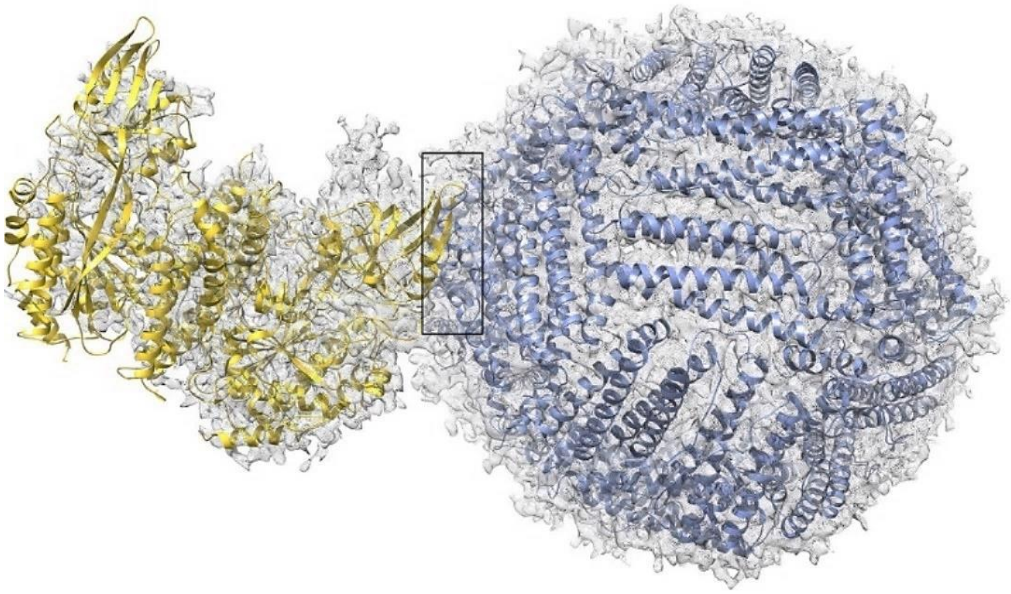


Figure 39 CD71/HFt complex final cryo-EM map (resolution 3.9 Å) with crystallographic structures fitted in (yellow: CD71, pdb 3KAS; purple: HFt, pdb 3AJO). The close-up view is in Figure 40.

Structural analysis of the binding region

CD71 binds HFt in a virus-like fashion, covering an area of ≈1900 Å^2:

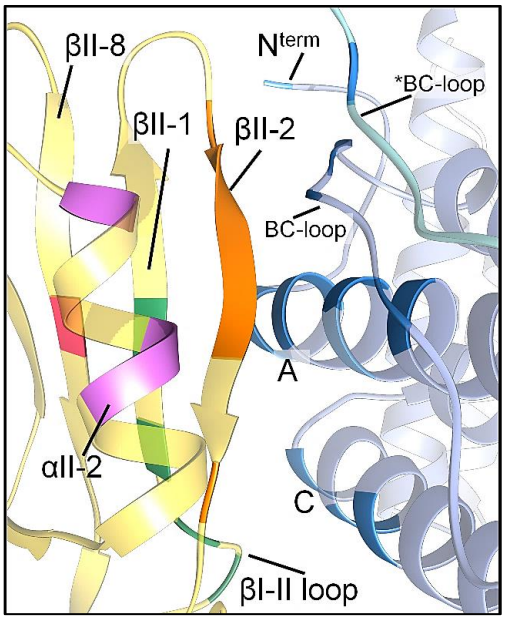


Figure 40 CD71/HFt binding region (black box of Figure 39) with crystallographic structures fitted in the cryo-EM map (yellow: CD71, pdb 3KAS; purple: HFt, pdb 3AJO), colored according to the text.

CD71 interacts through four specific regions on the apical domain, highlighted in different colors: *i)* the βI-II loop and the βII-1 strand (S195, E197, S199, I202: green); *ii)* a residue on the βII-8 strand (K374: magenta). These are specific on CD71 for HFt: we refer to them as "*exclusive contacts*". Additional residues are: *iii)* six amino acids on the βII-2 strand (R208-L212 and N215: orange); *iv)* residues E343, K344 and N348 on the αII-2 helix (purple). We refer to these as "*common contacts*" on CD71, since they represent the key structural determinants for binding also arenaviruses and plasmodial proteins (see Figure 12 for comparison).The HFt binding counterpart regions are three: *i)* the external BC loop (R79, F81, Q83, K86, K87); *ii)* the N-terminus of A-helix (T5, Q14, D15, E17-A19, N21, R22, N25); *iii)* the C-terminus of C helix (E116,

K119, D123).

An exhaustive list of HFt/CD71 pairwise contacts is given in the Appendix 6.1 (Figure S2): a summary is here briefly depicted. The interactions between exclusive contacts and HFt are: *i)* CD71 βII-1 strand and HFt A helix (Figure 41 left); *ii)* CD71 βI-II loop and HFt C helix (Figure 41 left); *iii)* K374 CD71 with Q14 and D15 (Figure 41, right).

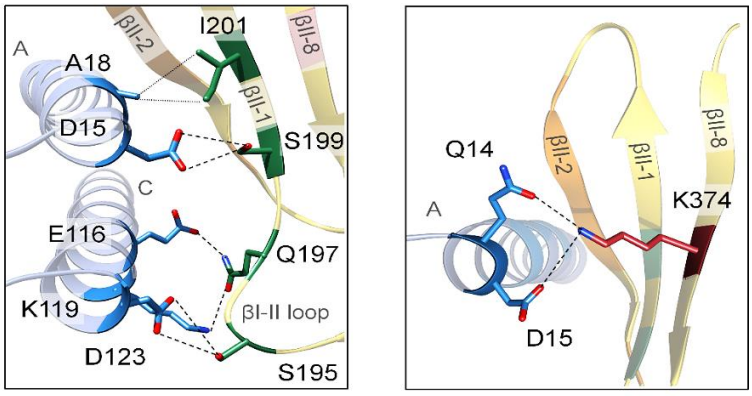


Figure 41 Contacts involving CD71 epitopes that are exclusive for HFt binding. Contacting residues within 5 Å distance are shown in sticks. Dashed black lines: electrostatic interactions, dotted grey lines: hydrophobic contacts. See Figure S2 in the Appendix 6.1 for details.

The interactions between common contacts and HFt are: *iv)* CD71 βII-2 strand and HFt A helix, BC loop, C helix and T5 (Figure 42, left); *v)* CD71 residues 344, 348 and HFt A helix, BC loop and K87 on the loop of a flanking HFt monomer (*K87) (Figure 42, right).

Notably, CD71 amino acids belonging to "common contacts" (figure 43) coincide with those leading a gain or loss of interaction with pathogen binding proteins upon mutations[62,70]. As mentioned in section 1.2 (Figure 12), these cluster in the βII-2 and αII-2 regions, which appear to be essential on CD71 for binding with various partners.

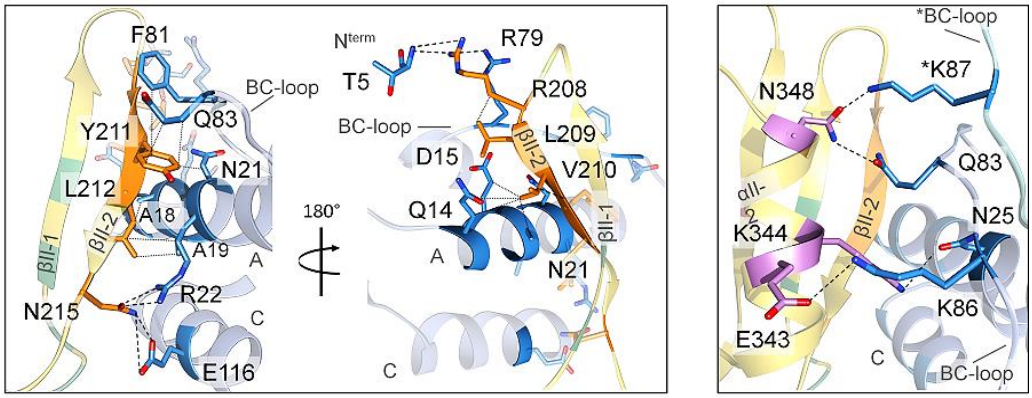


Figure 42 Contacts involving CD71 epitopes used for binding also other pathogens. Contacting residues within 5 Å distance are shown in sticks. Dashed black lines: electrostatic interactions, dotted grey lines: hydrophobic contacts. See Figure S2 in the Appendix 6.1 for details.

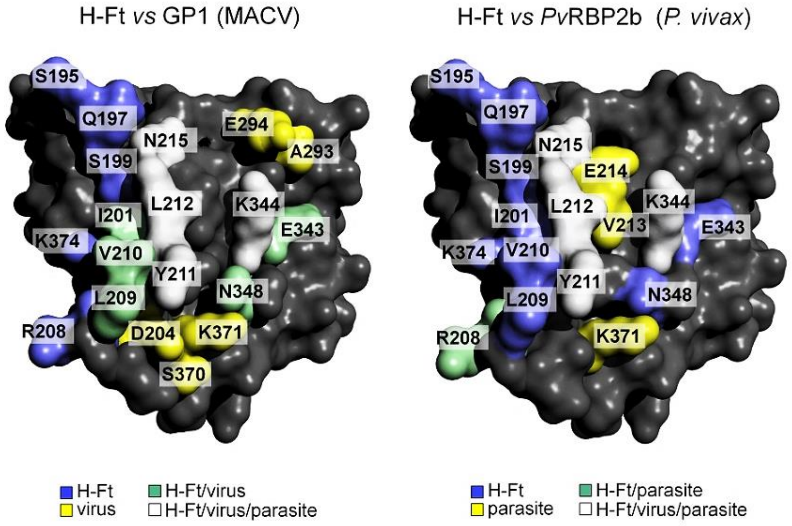


Figure 43 Surface representation of CD71 apical domain (dark gray). Ligand specific and overlapping binding residues are colored following the legend at the bottom. See Figure 12 for comparison.

Human CD71	Human H-Ft	GP1 (MACV)	*Pv*RBP2b (*P. vivax*)
S195	K119, D123		
Q197	E116		
S199	D15		
I201	A18	F226	
D204		D114	
R208	T5, R79		N527, E530, D531
L209	Q14, R79, F81	S97, F98	
V210	Q14, E17, A18, N21	R111, F226, Y228	
Y211	F81, Q83	R111, S113, I115, V117	S541
L212	A18, A19, R22	P223, F226	Y542
V213			Y604
E214			K600
N215	R22, E116	E171	D603
A293		Y122	
E294		M119, Y122, K169	K600
E343	K86	Y122	
K344	N25	M119, D123, K169	Y604
N348	Q83, *K87	D114, S116	
S370		D114	
K371		S116	Y538
K374	Q14, D15		

Table 1 summary of the interfacing residues at CD71 apical domain contacted by human HFt, GP1 of Machupo virus [61] and *Pv*RBP2b of *P. vivax* [62]. CD71 residues contacted by HFt are colored according to the color code used in Figure 40.

Characterization of mutants by SPR, FACS and confocal microscopy

After the identification of the interacting region, we wanted to detect between all HFt residues contacting CD71 (black bold letters in Figure 44) the crucial ones for complex formation. So, amongst HFt residues involved in "common contacts" (T5, Q14, D15, E17-A19, N21, R22, N25, R79, F81, Q83, K86, K87, E116) we selected only those that are not conserved between human HFt and ferritins unable to bind CD71 (*i.e.* human LFt[20] and AfFt[46]): they are Q14, D15, R22, F81, Q83 (black boxes in Figure 44).

To prove their importance, we produced three multiple mutants of HFt:

i) mutant A, lacking the polar residues at the N-terminal of the A helix (Q14A, D15A, R22A);

ii) mutant B, lacking F81 and Q83 on the external BC-loop (F81A, Q83A);

iii) mutant C, combining A and B (Q14A, D15A, R22A, F81A, Q83A).

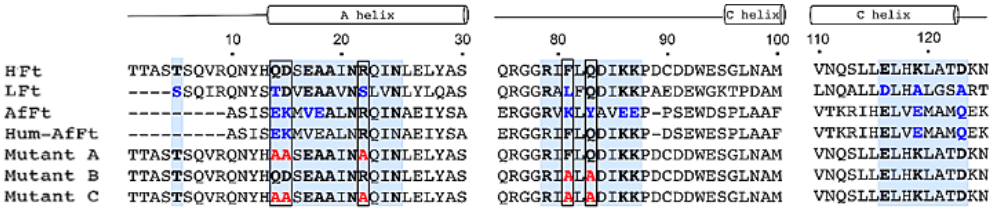


Figure 44 Ferritins (Human H and L chains, AfFt, HumAfFt, Mutants A, B, C) sequence alignment of regions contacting CD71. Elements of secondary structure are shown on the top. Colored background highlights the HFt contact regions with CD71. Conserved residues (respect to HFt) are in **black**, non-conserved in **blue**, mutated in **red**. Black boxes are residues crucial for CD71 binding. (Alignment made with Clustel Omega, figure with ENDscript [128])

Surface Plasmon Resonance (SPR) assays were done to measure the kinetics parameters (association, k_{on}, and dissociation, k_{off}, rates) and the affinity of the ligand-analyte interaction (K_D), using wild-type or mutant HFts as analytes and CD71 as ligand. The reaction between an immobilized ligand (L) and an analyte (A), in a 1:1 binding model, can be assumed to follow a pseudo-first order kinetics:

$$\mathrm{L} + \mathrm{A} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{LA}$$

The differential equation governing the formation of the complex [LA] is[38]:

$$\frac{\mathrm{d[LA]}}{\mathrm{dt}} = k_{on}[\mathrm{L}][\mathrm{A}] - k_{off}[\mathrm{LA}]$$

From its integration, considering the association and dissociation phases, it is possible to fit the experimental curves of SPR sensorgrams and to obtain a measure of the parameters k_{on}, k_{off} and K_D of the reaction.

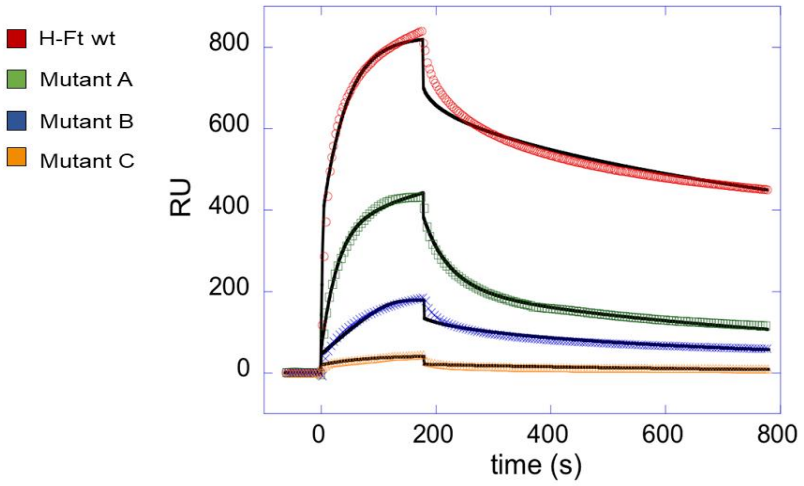


Figure 45 SPR sensorgrams of the interaction between the immobilized His-tagged CD71 receptor and HFts (wild-type and mutants), used as analytes. Fits are reported as black lines; fitted parameters are in Table 2.

Protein Analyte	k_{on1} ($M^{-1}s^{-1}$)	k_{off1} (s^{-1})	K_{D1} (nM)	k_{on2} ($M^{-1}s^{-1}$)	k_{off2} (s^{-1})	K_{D2} (mM)
H-Ft	$(4.78\pm0.07)*10^{5}$	$(3.40\pm0.06)*10^{-3}$	7.1±0.2	$(1.69\pm0.04)*10^{5}$	$(2.51\pm0.06)*10^{-2}$	0.156±0.007
Mutant A	$(1.1\pm0.2)*10^{6}$	$(6.5\pm0.9)*10^{-1}$	570±170	$(1.4\pm0.1)*10^{4}$	$(2.7\pm 0.1)*10^{-2}$	1.9±0.3
Mutant B	$(1.5\pm0.3)*10^{6}$	$(3.3\pm0.2)*10^{-1}$	220±80	$(1.7\pm0.2)*10^{2}$	$(2.76\pm0.04)*10^{-2}$	158±18
Mutant C	Not feasible					

Protein Analyte	k_{on1} ($M^{-1}s^{-1}$)	k_{off1} (s^{-1})	K_{D1} (nM)
H-Ft	$(3.74\pm0.08)*10^{4}$	$(6.72\pm0.08)*10^{-4}$	17.9±0.6
Mutant A	$(1.7\pm0.2)*10^{5}$	$(1.8\pm0.2)*10^{-3}$	140±50
Mutant B	$(6.6\pm0.7)*10^{4}$	$(8.9\pm0.8)*10^{-3}$	130±30
Mutant C	1200 ±100	$(1.55\pm0.03)*10^{-3}$	1300 ±100

Table 2 Kinetic and thermodynamic parameters for SPR experiments: **Top**: heterogeneous analyte fit. **Bottom**: simple 1:1 binding mode

The wild-type HFt-CD71 binding affinity (K_D= 17.9 ± 0.6 nM in the 1:1 binding mode) is compatible with the one already existing in literature[38]. Instead, this parameter is reduced of about two orders of magnitude in mutants A and B and abolished in mutant C. In particular, a critical drop of the k_{on} value is increased across mutants B and C, suggesting that mutations at the BC loop have a dominant role in impairing the CD71/HFt interaction: this is likely due to the loss of contact of HFt residues F81 and Q83 with CD71 residues Y211, and between HFt Q83 with CD71 N348.

Consistently, FACS and confocal microscopy measurements on HeLa cells uptake of the three HFt mutants show reduced (mutant A) or negligible (mutants B and C) internalization with respect to the wild type (Figure 46 and Figure S5 in Appendix 6.1). As such, the substitution of just 2 (mutant B) or 5 (mutant C) residues in HFt monomers is sufficient to make its internalization comparable to the *A. Fulgidus* Ferritin (AfFt), an archaeal ferritin with completely different primary and quaternary structures.

Further considerations on HFt-TfR1 binding

As mentioned in section 1.2 (Figure 12), the CD71 apical domain amino acids R208-L212, N215, E343-K344, N348 are all involved in specific interactions with viral capsid glycoproteins; in particular, Tyr211 plays a pivotal role for viruses entry[61,68]. This is also the recognition site for *Plasmodium vivax* reticulocyte-binding protein 2b (*PvRBP2b*), the key element for red cell invasion by the

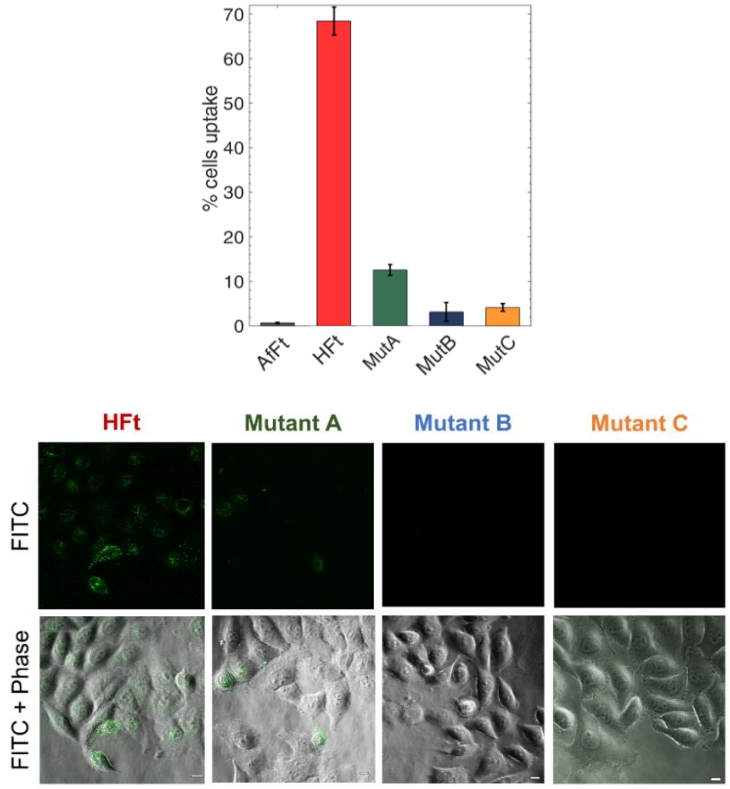


Figure 46 Top: Ferritins uptake in HeLa cells quantified by flow cytometry. The percentage of cells internalizing AfFt-FITC (here used as negative control [46]), HFt-FITC, MutA-FITC, MutB-FITC and MutC-FITC is shown as mean ± s.e.m. for $n=3$ independent experiments.
Bottom: Ferritins internalizations in HeLa cells observed at a 60x confocal microscope, here shown as single FITC and overlay images with phase-contrast. Scale bars: 10 µm.

parasite[62,73]. We found that the very same region is also exploited by HFt through its BC loop and A, C helices (Figure 40).

These results assess the critical relevance for CD71 interaction of HFt BC loop that recognizes the "common contact" residues Y211 and N348, crucial also for

viruses and plasmodial proteins. Upon mutation of BC loop residues F81 and Q83, the K_D is 100-fold reduced and internalization in cells is suppressed. Furthermore, we have also highlighted the relevance of Q14, D15 and R22 at the N-terminus of the HFt A helix, that recognizes both common and specific contacts on CD71. Therefore, the significant binding capability of the Humanized *Archeoglobus fulgidus* ferritin (Hum-AfFt[46]), where the human HFt BC loop had been transplanted, is likely due not only to the presence of F81 and Q83 on the loop, but also to R22 and E14 on the A helix, serendipitously present in this archaeal protein; this nanoparticle will be discussed in session 4.2. Indeed, human LFt, which lacks Q14, R22 and F81, is unable to bind the CD71 receptor[20] (see the alignment in Figure 44).

Notably, we can suggest that selected pathogens may have adapted to exploit the HFt epitope on the apical domain to unlock cellular barriers by mimicking this physiological interaction[61,66,67]. At a glance, in fact, MACV GP1-CD71 and HFt-CD71 complexes are incredibly similar (Figure 47).

As such, with the present research we identified the specific sites on CD71 to be hooked by ferritin for physiological access to cells. We provide structural evidence that Tf and HFt exploit alternative epitopes on the same receptor, allowing differential regulation of iron like HFE[60], as already proved[20, 125]. Importantly, up to now no other complexes associated with HFt have ever been singled out, nor the TfR1 apical domain (present in all mammalians) has ever had a known physiological role.

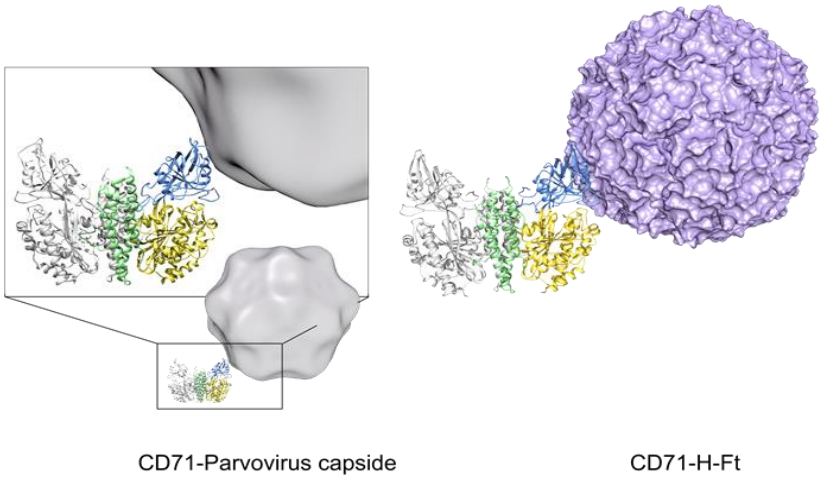


Figure 47 Visual comparison between the P. vivax (*PvRBP2b)* binding mode[62] and HFt with CD71.

In this framework, changes due to Single Nucleotide Polimorphisms (SNPs) within TfR1 apical domain may protect from viruses or parasites entry, but cannot be considered neutral with respect to HFt uptake. Along this line, TfR1 species-specific variants must conveniently match HFt co-evolved homologs in order to conserve its physiological functions. For example, mouse TfR1 (mTfR1) apical domain is rather different from the hTfR1 one in key residues for HFt binding (Figure 48). Indeed, mouse H-chain Ft (92% of identity with the human one[20]) in mice is known to be internalized by the receptor TIM-2, that is not expressed in humans[24,25]. In literature it is still not clear whether TIM-2 is the only receptor suitable for mouse HFt or if mTfR1 plays a similar role to its human counterpart. If mTfR1 didn't bind mouse HFt, then iron uptake phenomenon by humans and mice organisms would be quite different, as already hinted[20]: proving this validation will be of great interest for iron-delivery

mechanisms and to validate mice as animal models for iron-concerned researches. Indeed, other species-specifics partners are known to bind this very region of the apical domain: for example, Mouse Mammary Tumor Virus (MMTV) recognizes specific epitopes on mTfR1 and cannot bind hTfR1[66,129]; vice versa, MACV, JUNV, GTOV and SABV arenaviruses invade only hTfR1 and not mTfR1, nor other rodents TfR1s[70].

Figure 48 Alignment of different TfR1s from necessary regions for HFt and arenaviruses binding. *M. mulatta* (rhesus macaque) TfR1 shows high analogy with hTfR1, and indeed Rhesus macaques can be used as a model for MACV and JUNV infection [130].
Instead, rodents *M. musculus* (house mouse), *R. norvegicus* (brown rat) and *C. griseus* (Chinese hamster) TfR1s are quite different from hTfR1: rodents-derived CHO [125] and BHK cell lines are refractory to entry mediated by MACV, JUNV and GTOV GP [70].

In conclusion, our work provides a sound structural basis to elaborate on the possibility of developing alternative ferritin-like anti-viral or anti-parasite therapeutic ligand, be it an antibody or a peptidomimetic capable of blocking the "common contacts" epitope on receptor apical domain, and to further engineering ferritins as theranostic agents in hTfR1 overexpressing cancer cells.

4.2 Humanized A. Fulgidus Ferritin

Ferritins are widely used as versatile nanovectors for drug-delivery purposes as a result of hTfR1 overexpression in cancer cells. The encapsulation of the compound inside the cage is possible through a pH jump (Figure 8): following HFt disassembly in monomers at extremely acidic or basic pH values, the payload compound binds to the protein internal cavity; when the pH is changed back to neutrality, the tetraicosameric structure is spontaneously re-assembled[13].

The need for such coarse conditions for cargos loading is the most relevant constraint on an efficient use of Ferritins as nanoparticles for biomedical purposes: drug stability might be compromised and the reassembly of the cage might not be complete at all. The inevitable suboptimal load of cargo material, whose chemical structure is required to be pH resistant, often results in poor payload incorporation yields[45]. Consequently, research efforts are devoted in modifying the assembly properties of Ft nanocages either by inter-subunit interface mutagenesis[131] or by genetic engineering of N- or C-terminal regions[132].

Recently, novel ferritins from lower organisms as bacteria and archaea, endowed with different polymer association-dissociation thermodynamic and kinetic features, have emerged as possible alternatives to human HFt for applications requiring cargo material encapsulation[133]. In particular, as already mentioned in paragraph 1.1, the ferritin of *Archaeoglobus fulgidus* (AfFt) is characterized by unique self-assembly properties: in neutral buffers it is present as a dimeric species, while in presence of metal cations it assembles into a non-canonical 24-mer cage with a peculiar tetrahedral geometry having four 45 Å triangular pores

on the shell[17,18] (Figure 6, left).
Nevertheless, the low (31%) sequence similarity between AfFt with HFt runs against the possibility of targeting the archaeal ferritin towards hTfR1; their structural alignment reveals that the BC loops of adjacent subunits display the same geometry in both the proteins, though each couple of loops adopts a different symmetry along the spherical surface of the protein cage, given the different dimer-dimer positioning within each complete 24-mer structure (Figure 6, right).

In this framework, our aim was to genetically modify the AfFt structure in order to have a chimeric nanoparticle characterized both by AfFt peculiar assembly/disassembly properties, that could lead to a completely reversible cargo encapsulation, and by the recognition with hTfR1 typical of HFt, that would allow mammalian cells internalization.
This objective could be achieved simply by replacing the HFt recognition epitopes for hTfR1 in AfFt sequence. However, at the time of this investigation (January 2017) the binding region between CD71 and HFt was not solved yet: we had to guess such epitopes just from structural evidences. From the inspection of HFt and CD71 crystallographic structures alone, it was evident that HFt external loops represent the most significant accessible area to CD71. Actually, more compact patches of residues exposed to solvent are formed by the N-terminal regions or by the iron channels within the 3-fold axis; however, it has been shown that N or C-terminal deleted HFts are efficiently internalized by target cells, thus suggesting that these regions are not relevant for receptor recognition[15,134]. As a consequence, the BC loop, besides the known structural role in stabilization of interdimer interface[2,10], represented the best candidate for hTfR1 recognition. We decided to exploit this loop by mutating 9 AfFt

aminoacids VKLYAVEEP (from residue 70 to 78 in AfFt numbering) into the corresponding HFt residues IFLQDIKKP, located at the center of the human BC loop (Figure 49). In addition, a cysteine in position 54 (AfFt numbering) has been introduced by point mutation to provide a conjugation site of potential thiol reactive derivatives into the Ft cavity. The resulting chimera was called "HumanizedAfFt" (*HumAfFt*).

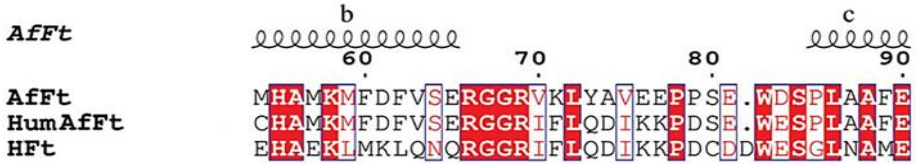


Figure 49 Structure-based sequence alignment of AfFt, HumAfFt, and HFt. Elements of secondary structure for the AfFt are shown on the top. White characters in a red background indicate strict conservation, residues with poor conservation are in black on a white background. (Alignments made using CLUSTAL Omega, the figure with ENDscript [128])

As shown below, HumAfFt nanocarrier displays both the unique assembly/disassembly properties of the archaeal ferritin and is actively internalized by HeLa cells to an extent comparable to transferrin, the preferred ligand for CD71.

Self-assembly assessment

The effect of mutations on the $MgCl_2$-dependent self-assembly of HumAfFt was studied by size exclusion chromatography (SEC) to separate different possible oligomers according to their molecular size.

As shown in the chromatograms, identical peak-positions at 20 mM $MgCl_2$ for both AfFt (blue) and HumAfFt (light blue) confirmed that HumAfFt retains

the cation induced self-assembly properties of native AfFt; in absence of $MgCl_2$, the chimera is possibly disassembled in dimers (cyan).

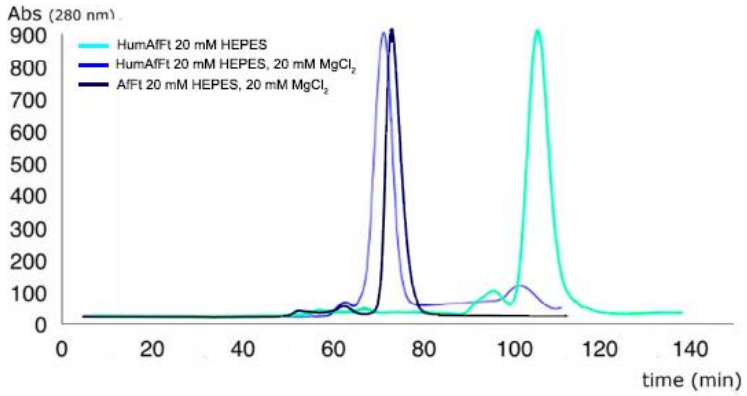


Figure 50: Chromatograms relative to the gel filtration profiles of: HumAfFt in 20 mM HEPES pH 7.5 (cyan); HumAfFt in 20 mM HEPES pH 7.5 and 20 mM $MgCl_2$ (light blue); AfFt in 20 mM HEPES pH 7.5 and 20 mM $MgCl_2$ (blue)

As such, HumAfFt is constituted by stable dimeric species at neutral pH and low ionic strength, capable of associating into non-classical 24-mer in the presence of either monovalent or divalent cations at physiological concentrations (*i.e.* higher than 0.5 M Na^+ or 20 mM Mg^{++}).

X-ray diffraction data

To confirm the tetrahedral symmetry reported for archaeal ferritins, the structure of HumAfFt has been determined by X-ray crystallography at 2.87 Å resolution (Figure 51, left). The protein was crystallized in the presence of Mg^{2+} to maintain the stable quaternary structure of 24-mer. It crystallized in the C2221 space group, as observed also for wild type AfFt; the asymmetric unit (ASU) contains 12 identical subunits with a solvent content of 64.3%. The

overall folding corresponded to the AfFt wild type structure (PDB code 1S3Q) and displayed 4 wide triangular pores on the surface. HumAfFt X-ray data collection and refinement can be found in Appendix 6.2 (Figure S6).

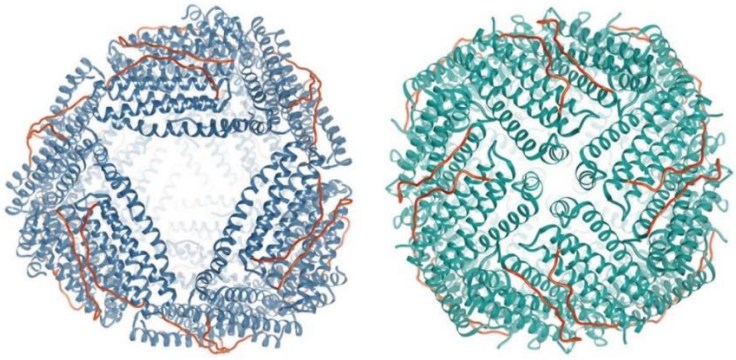


Figure 51 X-ray crystallography experimental data of HumAfFt (**left**) compared to HFt (pdb 3AJO, **right**). BC loops are in red.

The BC loop is well organized in the 3D structure: the analysis of the difference electron density (FoFc) map clearly showed the presence of the mutated residues. It also shows the M54C mutation pointing towards the inner cavity, as can be seen from the structural superpositions between HumAfFt-HFt and HumAfFt-AfFt (Figure 52).

Cryo-EM map reconstruction

HumAfFt samples were loaded on holey-gold grids and analyzed by cryo-EM. The aim of this acquisition was to assess the correct three-dimensional assembly of this chimera in a more native environment than crystallographic conditions.

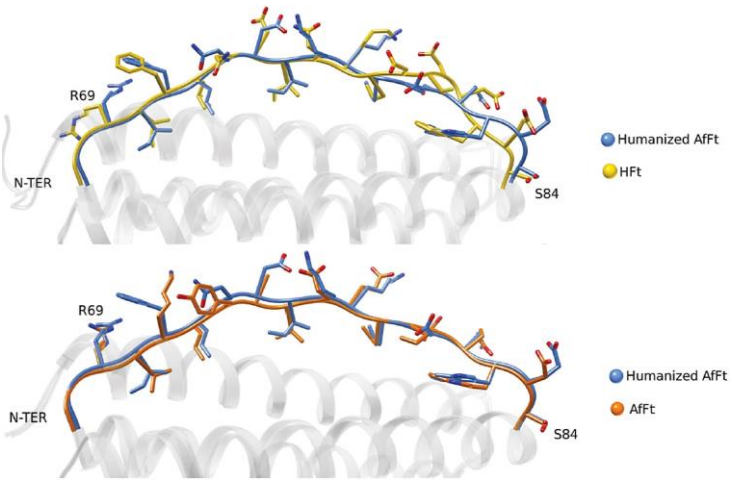


Figure 52 The structural superposition of the region from R69 and S84 (AfFt numbering) is here shown for HumAfFt (blue) *vs.* HFt (yellow) and for HumAfFt (blue) *vs.* AfFt (orange). Residues are depicted as sticks indicating N atoms in blue, O atoms in red and S atoms in yellow.

A sub-nanometer map was not needed, since high-resolution X-ray data already established the expected tetrahedral symmetry of the protein and the occurred loop modification.

For this reason, only 50 micrographs containing HumAfFt at a concentration of 1.2 mg/ml were acquired using a FEI Titan Halo (ThermoFisher Scientific, Eindhoven), a Volta phase-plate (FEI, Eindhoven)[108] and a FEI Ceta camera (FEI, Eindhoven). A representative micrograph is in Figure 53.

It is important to note that the image recording system here used is a CCD camera and is thus limited to much lower signal to noise ratios than a DDD, as can be judged even by naked eye by comparison between HumAfFt micrograph (Figure 53) and that of HFt-CD71 complex (Figure 26) or HFt-LBT Tb(III) (Figure 63). Moreover, because of their low frame rates[84], CCD cameras acquisitions are not constituted by sub-frames: thus, differently from the

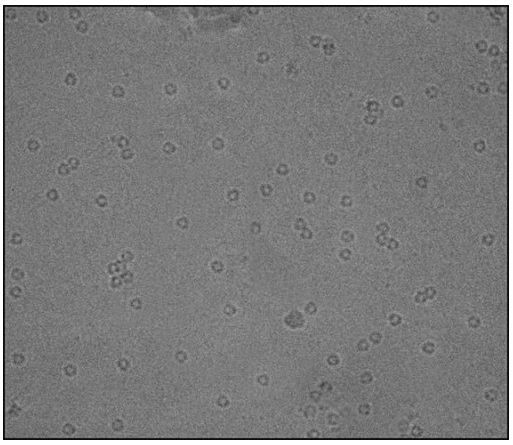


Figure 53: HumAfFt representative micrograph imaged by a CCD camera (FEI Ceta camera). Despite the lower SNR than HFt/CD71 complex or HFt-LBT Tb(III) datasets, it is still possible to see the protein and even the pores on its surface. The sample was very homogeneous and monodisperse.

other samples' datasets, the motion correction procedure could not be applied. In any case, this limit did not affect too deeply the overall final resolution of the map: during the acquisition, we inserted in the microscope column the Volta-phase plate (VPP), a device that enables in-focus cryo-EM imaging, thus improving the image contrast[108]. Because of the in-focus acquisition, VPP introduction generally does not require CTF correction of the frames: this was our case, since the data were collected within 200 nm of focus and the first CTF zero crossing was well beyond the achievable resolution of the dataset.
Hence, no CTF nor motion correction were performed for HumAfFt dataset.

The data were analyzed according to the single-particle workflow in Figure 21. 6500 particles were picked with the reference-based automated particle picking procedure implemented in RELION 1.3[109]. Those particles were subjected to 2D classification with k=100 classes, where the triangular apertures on the protein surface were clearly observable:

Figure 54 Representative 2D classes of HumAfFt sample.

Particles belonging to those classes (5500) were then subjected to 3D classification using K=8 classes and then to the Autorefine procedure. The final map was obtained at 33 Å (FSC=0.143 criterion[112]; FSC curve can be found in Appendix 6.2, Figure S7). Its shape corresponds to that of wild type AfFt: thus, in near-native conditions and in presence of $MgCl_2$ HumAfFt is well assembled as a 24-mer.

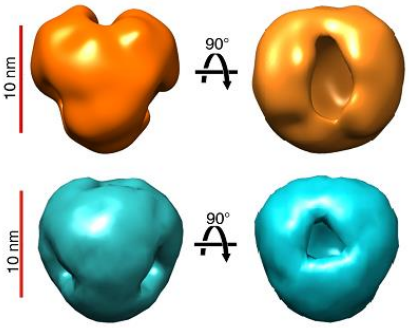


Figure 55 Top: cryo-EM map of HumAfFt (33 Å resolution). **Bottom:** AfFt cryo-EM reconstruction (obtained from the crystal structure [18], filtered at 30 Å), shown for comparison. Scale bars: 10 nm

As displayed in the Conclusions (chapter 5), the data acquisition of HumAfFt in complex with CD71 allowed to obtain a high-resolution cryo-EM map of HumAfFt (Figure 69): this confirms the presence of the pores and the overall molecular symmetry already found at 33 Å resolution. This achievement has been possible mostly through the acquisition with a DDD instead of a "simple" CCD, thus allowing motion correction and a far better reconstruction of the whole protein[84].

FACS, siRNA and confocal microscopy for HumAfFt cellular uptake

After we have demonstrated that HumAfFt maintains AfFt structure and the cation-induced self-assembly, we validated the effects of external loops mutations on the uptake efficiency by HeLa cells. CD71 is in fact highly expressed in this cell line, as in every cancer tissue[74]. We performed time course internalization assays on cells treated with 30µg/mL of AfFt-FITC, HumAfFt-FITC and transferrin-FITC (Tf-FITC) and analyzed them by FACS (Figure 56 left).

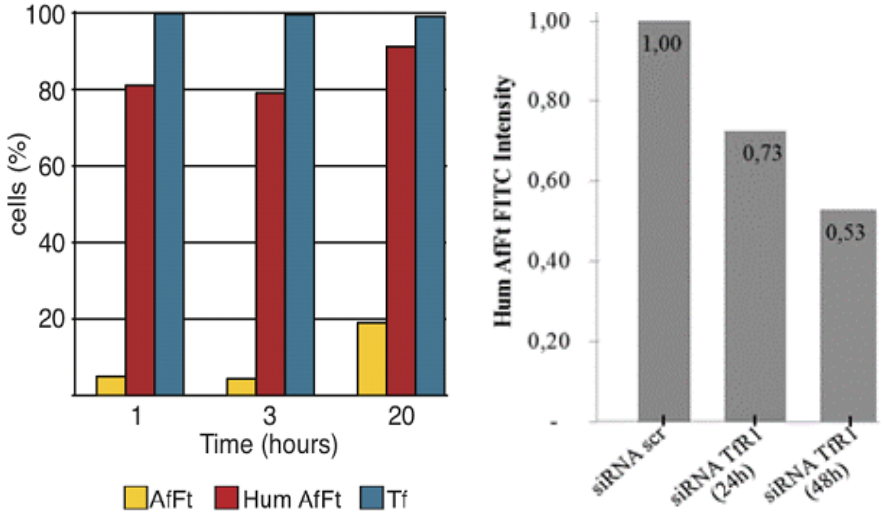


Figure 56 Left: FACS analysis is shown as the percentage of HeLa cells internalizing 30µg/mL FITC-labelled nanoparticles at different incubation times (1h, 3h, 20h). For each sample, 30.000 events gated on live cells have been acquired. **Right**: HeLa cells were transfected with control (scr) and specific anti-hTfR1 siRNA for 24 and 48h. Cells were incubated with HumAfFt-FITC for 3h, collected, washed and analyzed by FACS.

HumAfFt nanoparticles are efficiently internalized by HeLa cells already after one hour incubation, with a much higher percentage compared to AfFt. After 20 hours, the HumAfFtFITC-positive cells are increased at more than 90%, whereas the AfFt-FITC ones are still less than 20% (this increase respect to one

hour is possibly due to unspecific uptake by pinocytosis). Representative FACS acquisition plots are in Appendix 6.2 (Figure S8).

To confirm the involvement of hTfR1 in the internalization process of this nanoparticle, we also performed RNAi experiments against CD71 and analyzed once again the HumAfFt-FITC cellular uptake in the same conditions as before (Figure 56 right): as expected, after 48h of siRNA transfection around 50% of the cells reduced the uptake compared to control.

The uptake has been inspected also by confocal microscopy (Figure 57).

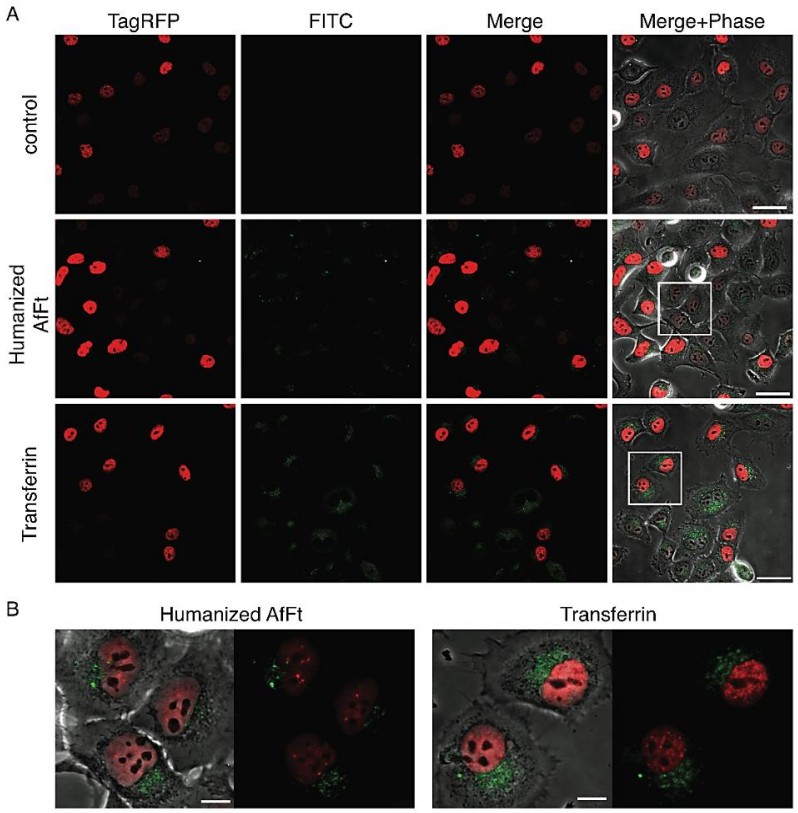


Figure 57 Fts internalization at the confocal microscope. Cells were plated, induced with doxycycline to express TagRFP and then incubated with 30 µg ml^{-1} of HumAfFt-FITC or Tf-FITC for 20 h. After washing, cells were live-imaged in RFP, FITC and phase contrast channels. **(A)** Scale bars: 40 µm. **(B)** Zoomed view of the region highlighted by the white inset in panel *A*. Scale bars: 10 µm.

HeLa TagRFP were incubated for 20 hours with HumAfFt-FITC or transferrin-FITC (Tf-FITC), washed and then imaged by confocal microscope. Images confirmed the high extent of HumAfFt internalization and highlight a cellular distribution in the cytoplasm and in the perinuclear space comparable to that observed in the case of Tf: this suggests a typical hTfR1-mediated endocytosis pathway.

Further considerations

The present data highlight the remarkable properties of a novel chimeric ferritin nanocage, HumAfFt, that combines the flexibility and reversibility in assembly/disassembly of AfFt quaternary structure with mammalian cells uptake by CD71 targeting, thus allowing the delivery of the desired payload to cancer cells as HeLa. This combination has been possible just through the introduction of human BC loops in the AfFt sequence.

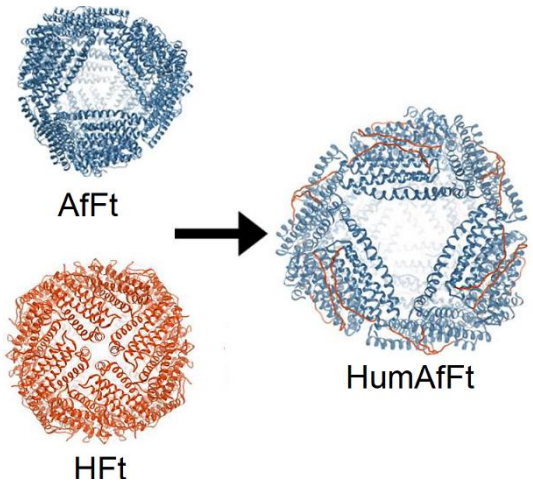


Figure 58 Schematic picture of the different parts constituting HumAfFt: AfFt overall structure for cation-induced assembly, HFt loops (in red) for human cells internalization.

As already mentioned, at the time of this research the HFt-CD71 structure was still not solved. Based on our more recent work (see session 4.1), we can explain the reasons behind the gain of HumAfFt interaction with hTfR1, consequent to the BC loop modification in AfFt. In HFt-CD71 complex, residues Q14, D15, R22, F81 and Q83 on HFt surface have been proved to be fundamental for the recognition with CD71. As such, the natural fortuitous presence of residues E14 and R22 alone (human HFt numbering, see Figure 44) on AfFt A helix is necessary but not sufficient to guarantee hTfR1 recognition; only when coupled to F81 and Q83, introduced in HumAfFt, a HFt-comparable binding capability is achieved.
Vice versa, HumAfFt internalization in mammalian cells proves the importance of BC loop for CD71 recognition, as we proved in section 4.1.

Thus, AfFt represents a uniquely suitable scaffold for incorporating a wealth of diverse substructures inside the Ft cavity in a complete reversible manner and without the harsh side effects of a pH jump.
At present, further experiments with possible payloads for diagnostic or therapeutic applications are under development[135,136]: in particular, a HumAfFt-CD71 complex (discussed in the Conclusions, chapter 5) will be resolved by cryo-EM.

However, *in vivo* applications will have to wait for the evaluation of possible immunological responses against non-human epitopes present on the protein surface.

4.3 Lanthanide binding Ferritin

Fluorescent probes based on trivalent lanthanide ions are widely used for bioimaging purposes thanks to their unique photophysical properties, related to *4f* orbitals[137]: their spectra are indeed characterized by sharp peaks, their high quantum yields are capable of efficiently radiating most of the absorbed energy and their microsecond lifetime fluorescence assures the high spatiotemporal resolution needed for real-time cellular imaging[138].

Nevertheless, their small absorption cross sections pose several limits to their use in advanced optical imaging techniques, that need a bright fluorescent signal for proper biological visualization. To overcome this restrain, lanthanides are currently used in complex with small organic fluorophores which absorb a photon in the UV region and transfer it to the lanthanide, populating its excited state and allowing its emission[139].

Lanthanide-fluorophores complexes, thus, acquire peculiar photophysical properties and an improved luminescent signal. The efficiency of energy transfer between the fluorophore and the lanthanide is highly dependent on their distance[139]: if they are within 5-6 Å, the low extinction coefficient of the ion can be overcome by Förster resonance energy transfer (FRET) that allows bright and photobleaching-less fluorescence signals[140]. This phenomenon is often referred to as "antenna effect". Among lanthanides, Tb(III) and Eu(III) are the most used in biotechnology due to their more intense µs fluorescence in the visible region[141].

Lanthanide ions can also be inserted in biomolecules for *in vivo* imaging. To obtain this result, a lanthanide binding sequence can be directly encoded within a recombinant protein expression construct *via* molecular biology strategies. In

particular, short polypeptides comprising 20 or fewer encoded aminoacids, identified as lanthanide-binding tags[142], have been recently developed. These peptide sequences show low-nanomolar affinities for the target ions, are selective for lanthanides over other common metal ions[143] and can be inserted in the protein construct at the N or C-terminal domain or even at specific loops[144]. The probe nature of these protein tags has been demonstrated by their use for various biotechnological purposes[145].

Apoferritins naturally bind more than one Tb(III) ion[2,146] in the ferroxidase site, threefold channel and nucleation centre (Figure 4), with considerably variable affinities (estimated K_D range from 2 to 666 μM at pH 6.4[147]). The Terbium can be directly incorporated inside the Ferritin cavity after the pH-induced dissociation in dimers[148], but this method may lead to a random and inhomogeneous distribution within the inner hollow, resulting in possible aggregation-induced quenching effects or Tb leakage to the outside through the iron channels. Upon excitation between 280-295 nm, Tb(III)-ferritin complex show characteristic emission bands at 490 nm ($^5D_4$-$^7F_6$ Tb electronic transition) and at 544 nm ($^5D_4$-$^7F_5$ Tb electronic transition) due to a FRET sensitization effect provided by aromatic aminoacids[149]. However, the distance between Terbium ions and aromatic moieties in native Fts make FRET efficiency very poor and suboptimal for any type of fluorescence/luminescence based measurement[145].

To overcome the FRET inefficacy and to achieve an effective Tb(III) internalization in Ferritin, we produced an engineered mouse HFt construct bearing a 17 aminoacids long lanthanide binding tag (LBT), genetically fused at the C-terminal end of each subunit. Mouse HFt was used since its sequence is identical to the human one within the CD71 binding region and because of

more favorable immunogenic profile for forthcoming *in vivo* study in mice. The polypeptide has been designed to be located inside the inner cavity as a prolongation of the Ferritin E helix (see figure 2), in order to trap the lanthanide ions diffusing from the outside through the surface pores. LBT sequence is YIDTNNDGWIEGDELLA and was conceived according to the findings of Martin et al.[106], who demonstrated efficient lanthanide chelating properties of a peptide sequence derived from Ca^{2+} binding sites from Troponin C EF hand motif.

As shown in the following, this biomolecular nanosystem (referred to as HFt-LBT Tb(III) when loaded with Tb(III), or HFt-LBT otherwise) has high affinity for Tb(III) and is endowed with both strong FRET sensitization and CD71 receptor binding properties.

Fluorescence spectroscopy

To find out the affinity of Tb(III) binding to this chimeric nanoparticle, static emission spectra were recorded upon excitation at 295 nm of HFt-LBT Tb(III) and wild-type (wt) mouse HFt after the removal of unbound Tb(III). As reported in Figure 59, the intensity of Terbium emission peak at 544 nm (typical of its excited state decay) for HFt-LBT was narrow and at least two orders of magnitude higher with respect to HFt alone. This is due to the excitation of the tryptophan residue, located in the center of the LBT, that lies at an efficient energy transfer distance from Tb(III) and thus provides an excellent FRET exchange with it.

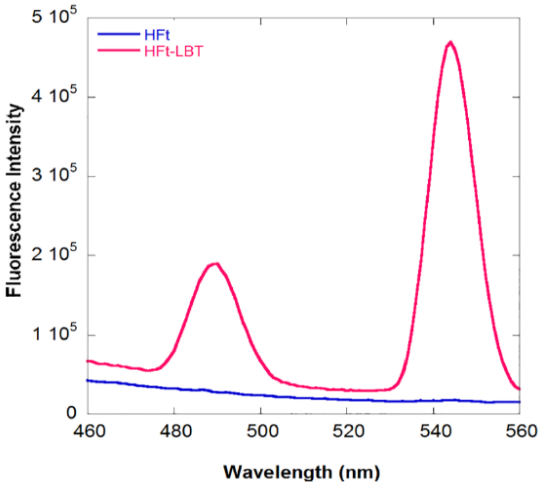


Figure 59 Fluorescence spectra of HFt-LBT Tb(III) complex (red line) and wt mouse HFt Tb(III) (blue line) at the same protein concentration (1 μM monomer). The main emission band due to the $^5D_4$-$^7F_5$ (Tb d-f orbitals) electronic transition at 544 nm is accompanied by the weaker $^5D_4$-$^7F_6$ transition. Spectra were recorded after protein saturation with $TbCl_3$.

Fluorescence titration analysis was thus carried out on HFt-LBT by adding free Tb(III) ions to the apo protein (Figure 60). The titration endpoint was reached at 1.7 equivalent amounts of Tb(III) per HFt subunit instead of the predicted 1 equivalent, based on the presence of one LBT moiety per monomer. The TRP residue in LBTs is thus able to act as an antenna system not only for the Tb(III) ion bound to the tag, but also for a number of extra Tb(III) atoms bound to the natural ferritin binding sites. It also contributes for most of the observed signal, since analogous titrations on wt HFt demonstrated a negligible fluorescence contribution of Terbium bound to the metal binding sites.

Such finding, in agreement with previous reports[146,149] and with the X-ray crystallography data shown below, confirms that in the wild type protein Tb(III) is bound to the canonical iron binding sites, where it receives extremely weak energy transfer from the unique TRP residue (W94).

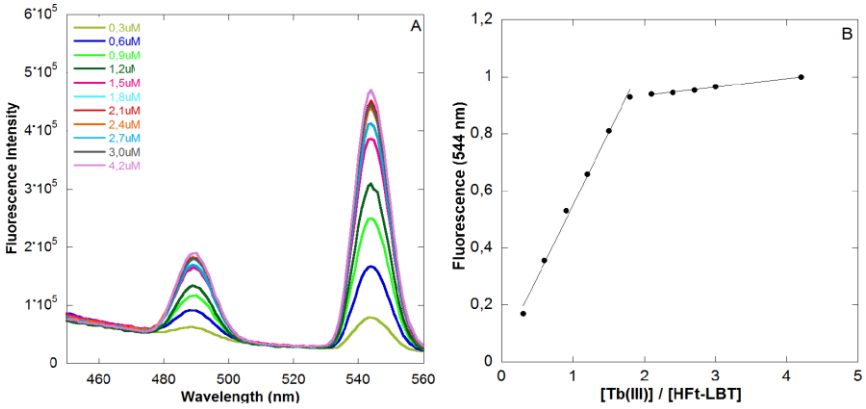


Figure 60 Left: Fluorescence titration of HFt-LBT with incremental concentration of Tb(III). Emission spectra were recorded upon excitation at 295 nm. **Right**: Fluorescence intensity of HFt-LBT Tb(III) complex as a function of the [Tb(III)]/[HFt-LBT] ratio. Fluorescent intensity was recorded at 545 nm, corrected for dilution factor and normalized to the emission maximum.

Structural analysis by X-ray christallography and Cryo-EM

Once we proved the FRET properties of our innovative nanoparticle, we verified by X-ray crystallography that the LBTs do not affect the overall assembly of the protein and that they are located inside HFt interior cavity.

The X-ray-structures of apoHFt-LBT and HFt-LBT Tb(III) (Figure 61A) were determined at 2.85 Å and 2.65 Å respectively; they both crystallized in I222 space group with 24 identical subunits in the asymmetric unit (ASU) with a solvent content of 64.7%. LBT presence, as expected, does not alter the structure, that matches with the native H-chain mouse Ft (PDB code 3WNW) with a *rmsd* value of 0.1 Å.

Taken together, these data highlight that HFt-LBT construct is capable of:

i) high affinity binding of 24 Tb(III) atoms, one per each LBT;

ii) intermediate affinity binding of 20 Tb(III) atoms at the ferroxidase sites (depicted as magenta spheres in Figure 61 A, B and in Figure 62 A) out of 24 available, as demonstrated by 75% occupancy in X-ray structure, by carboxyl residues E27 and E62, with minor participation of H65;

iii) lower affinity binding of 8 Tb(III) atoms at the entrance of the 3-fold axis channels by side chains of E131 and D134 (depicted as orange spheres in Figure 61 A, C and in Figure 62 B).

Hence, within the whole 24-mer a total of 56 Terbium atoms can be hosted, if we consider full occupancy of the ferroxidase sites.

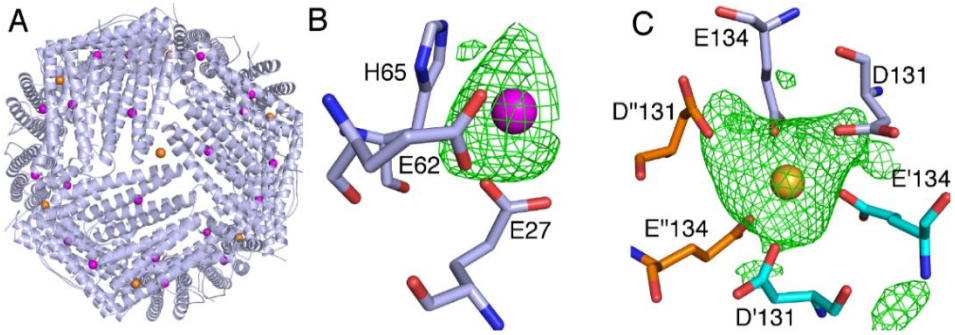


Figure 61 Tb(III) binding sites from X-ray crystallography. **A)** An overall view of the 24-meric shell of HFt-LBT Tb(III) showing the positions of Tb(III). The omit map contoured at 3σ as a green mesh is shown for **B)** ferroxidase center (where Tb(III) is in magenta) and for **C)** 3-fold axes (where Tb(III) is in yellow).

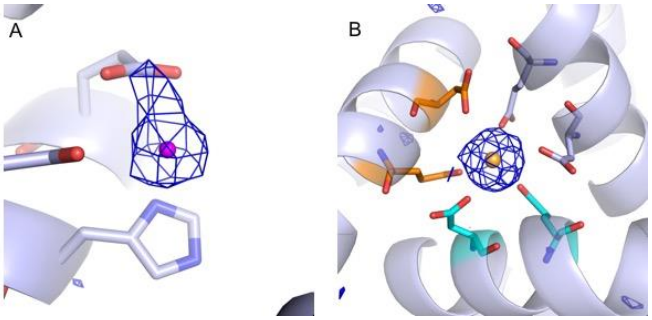


Figure 62 Anomalous difference map peaks of Tb(III). Terbium is represented as spheres and the map, contoured at 3 σ, is represented as a blue mesh. **A)** Tb(III) at the ferroxidase center; **B)** Tb(III) at the 3-fold axes.

As expected, the flexibility around the GSG spacer of the peptide arm connecting the LBT to the E-helix does not allow by crystallography even a partial detection of the occurred polypeptide insertion; in addition, also LBT bound Tb(III) atoms were not visible in the crystal structure. X-ray data refinement tables are shown in the Appendix 6.3 (Figure S9).

With the aim to detect the LBTs, we loaded HFt-LBT Tb(III) nanoparticles on holey-gold grids and analyzed their structure with cryo-EM. A qualitative analysis of the collected micrographs showed the high level of homogeneity and monodispersion of the sample in solution in near native conditions (Figure 63 left). We obtained a cryo-EM map at 7.1 Å resolution (Figure 63 right).

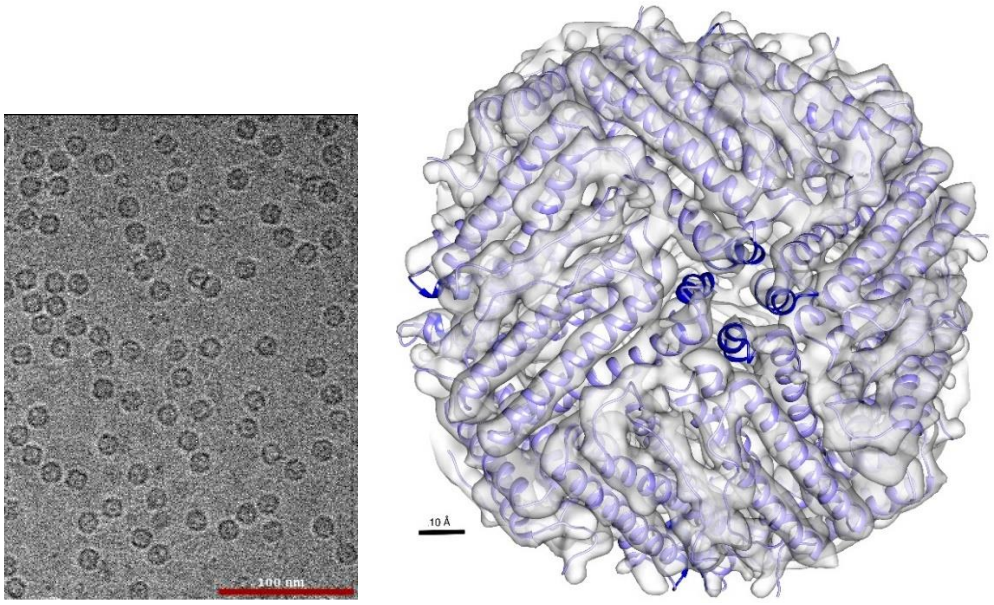


Figure 63 Left: representative micrograph of HFt-LBT Tb(III) imaged with a DDD; the sample was very homogeneous and monodisperse. Scale bar=100 nm. **Right:** cryo-EM map of HFt-LBT Tb(III), obtained at a resolution of 7.1 Å, with crystallographic data of the same sample (displayed in Figure 61 A, in lilac) fitted in; E-helices are in highlighted in blue. From this point of view, the LBTs cannot be seen. Scale bar=10 Å.

Unlike crystallography, that must deal with a very pure sample population, cryo-EM imaging is capable of reconstructing flexible parts of non-homogeneous samples of complexes or proteins, provided a sufficient number of particles having that particular conformation[126,150]. In our case, indeed the cryo-EM reconstruction did allow the identification of low-resolution but definite density patterns (14 Å long) relative to the first 6-7 residues of the tags (main chain length). However, good classes were not populated enough to reconstruct all the possible multiple orientations, thus precluding the electronic density building of the whole polypeptide.

The tags are seen as kinked small segments attached to the C-terminal part of each E-helix; they protrude, as expected, to the center of the protein:

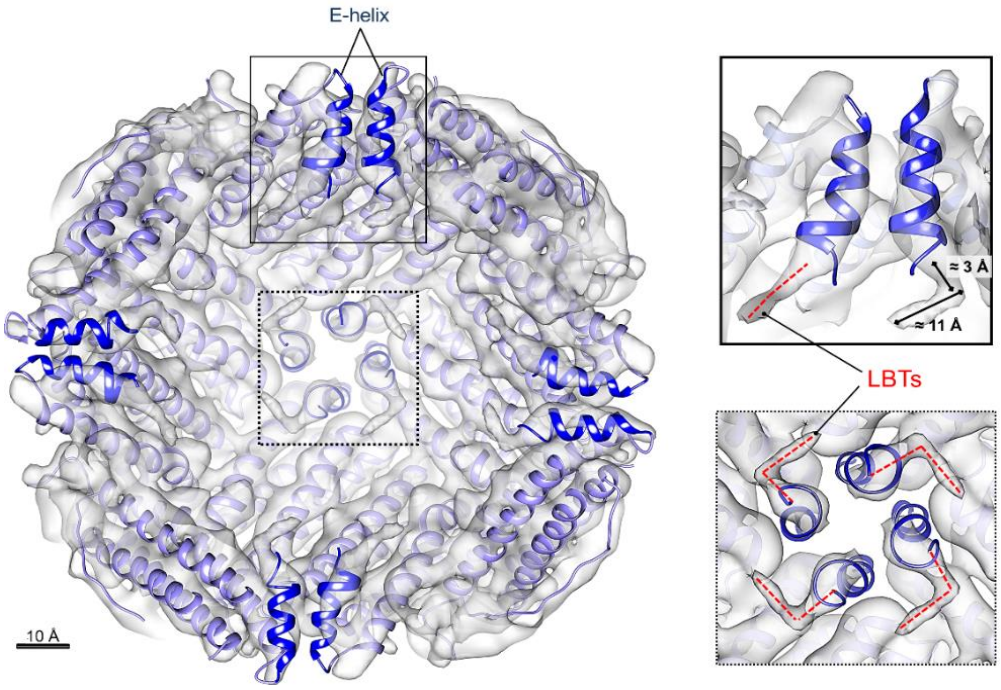


Figure 64 Same HFt-LBT Tb(III) cryo-EM map as Figure 63, here seen from the internal view, sliced at the protein center. E helices, in blue, are here more visible. LBT tails are enlarged on the right panels showing the top view (dashed line box) and the side-view (continuous line box). Electron density (red dotted lines) extends over the end of the X-ray model as due to the presence of the terminal LBTs, that approximately are 14 Å long (*i.e.* 6-7 amino acids). Scale bar = 10 Å.

We have analyzed the data according to cryo-EM workflow (Figure 21).

The map of Figure 63 and Figure 64 has been obtained with RELION 2.0[109]. Following CTF[100] and motion correction[98], 91947 particles were automatically picked from the original 350 micrographs; after the extraction, particles were 2D classified using 100 classes. Particles (48047) belonging to good 2D classes were selected and subjected to 3D classification (Figure 65), using as reference model the mouse H-ferritin map generated from its PDB structure, ultimately yielding 8 classes.

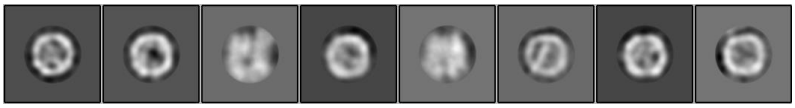


Figure 65 3D classes averages on a 2D view. Classes 3 and 5 contain noisy contributions; other classes have been selected for PostProcessing. Even at such low resolution, it is possible to notice non-trivial structures inside the cavity.

A selection of 35625 particles were subjected to another round of 3D classification. This resulted in only one good class (7345 particles) in which LBTs could be seen as "blobs"; it was refined with the 3D Autorefine procedure, imposing the 'O' symmetry (Ferritin space group symmetry), that yielded a 9 Å resolution. The final Post Processing 3D map resolution (equal to 7.1 Å) was estimated with the Fourier shell correlation=0.143 criterion, based on the 'gold-standard' protocol[112]; FSC curve is shown in Appendix 6.3 (Figure S10).

A rigid-body fit, calculated in Chimera, of this map with the high-resolution structure of the same HFt-LBT Tb(III) sample obtained by X-ray crystallography (shown in Figure 63 and Figure 64 in purple) indicates a perfect matching of the subunit assembly and the helix axes; it clearly shows that the LBTs are in "flip" conformation (*i.e.* with the E-helix oriented inside the cavity)

in solution, as we wanted. The tags can also be seen in 3D classes (Figure 65) as densities that point to the Ft internal cavity; possible "flop" conformations (*i.e.* with the E-helix oriented outside the cavity) are not observed in any of the classes. The arrangement of LBTs inside the inner cavity is a fundamental property of this nanoparticle, since only in this position they can bind potential lanthanide ions diffusing from the external through HFt surface pores.

Uptake by tumor cell lines

After the demonstration that HFt-LBT Tb(III) maintained the overall structure of the wild-type mouse HFt and that it is capable to bind multiple Terbium ions, internalization of 0.5 mg/ml HFt-LBT Tb(III) and wt HFt was analyzed in selected human cancer cell lines by flow cytometry and confocal microscopy. Although HFt-LBT Tb(III) is an engineered mouse HFt, it can be recognized by human cells thanks to its high homology with human HFt.

In any case, despite several efforts, attempts to image live cells after nanoparticles uptake by direct excitation at 290-375 nm of the tryptophan residue in the LBT yielded very poor results within the typical Terbium emission interval, due to the fluorescence background of cytoplasmic proteins. Clearly, a time resolved fluorescence microscopy approach or a red shifted antenna system will be needed to expand the scope of the present work.

We thus report FACS and confocal microscopy measurements of cells in which HFt-LBT Tb(III) or wt mouse HFt have been labeled with FITC (Figure 66 and Figure S11 in Appendix 6.3).

Results demonstrate that the HFt-LBT Tb(III) complex is very efficiently

internalized by all four human tumor cell lines selected. In particular, DU145 (from a central nervous system metastasis, of primary prostate adenocarcinoma origin), MDA-MB-231 (from invasive ductal carcinoma, Hepatoma cell line) and SK-OV-3 (highly resistant ovarian cancer cell line) were demonstrated for the first time to display a high uptake of both wild type mouse HFt and HFt-LBT Tb(III): this fact is noteworthy, since these cancer cell lines overexpress CD71 receptor and are indeed subject of cancer therapy studies focused on this receptor[55,75].

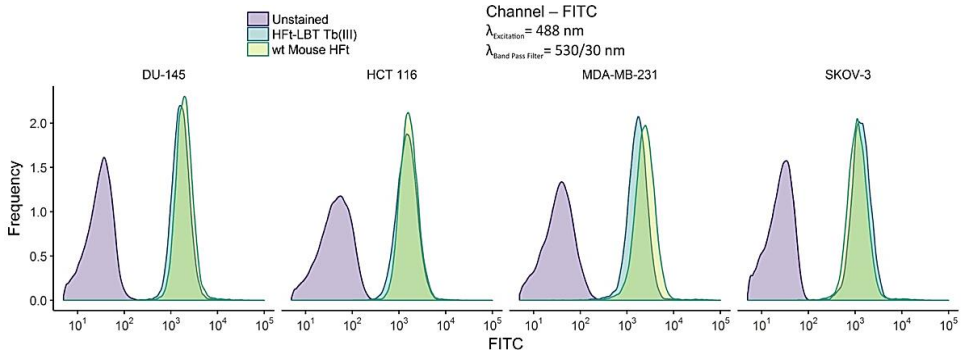


Figure 66 FACS analysis of HFt-LBT Tb(III) uptake by selected human tumor cell lines DU-145, HCT-116, MDA-MB-231 and SKOV-3, incubated with either HFt-LBT Tb(III) or wild type mouse HFt (0.5 mg/ml) for 60 min.

.

Further considerations

We have shown that a mouse HFt nanocage can be engineered by introducing appropriate metal binding tags (LBTs) inside its cavity to provide additional metal sites in topologically selected positions. The LBTs are 17 amino acids long polypeptides with a TRP residue in the center that acts as the organic fluorophore that correctly transfers energy to the Tb(III) *via* an antenna effect. As such, HFt-LBT Tb(III) construct shows great FRET properties (Figure 59

and Figure 60) and binds a total of 56 Tb(III) atoms (Figure 61 and Figure 62). The tags are efficiently located inside the HFt cavity (namely, at the C-terminal part of E helices) as proved by cryo-EM (Figure 63 and Figure 64).
Additionally, the key physical property of the tag as chelating antenna system for Terbium can be coupled to CD71 receptor recognition properties of HFt. The LBTs insertion does not hinder the contact between Ft and CD71 since these tags are a prolongation of the E helix, a region that is not essential for CD71 recognition. Thus, the same internalization in cancer cells as the wild-type can occur, with additional bioimaging applications. Indeed, HFt-LBT Tb(III) binds to different types of human cancer cells (Figure 66, Figure S11).

In conclusion, we propose the approach here described as an alternative to the *quasi* random insertion of metal clusters or small organic fluorophore into the Ft cavity by free diffusion through the macromolecule pores or by disassembly-reassembly of the 24-mer[145]. The importance of a guided allocation of metal sites inside the cavity is considered essential to proceed with rational positioning of antenna systems close to lanthanide sites, with the aim of designing the best geometry for efficient FRET. Positioning of additional tryptophan residues next to these binding sites could also provide an enhancement in fluorescence/luminescence Tb(III) yields and it will be evaluated in the future. Rational design of such metal binding sites would foster more advanced applications such as the construction of up-converting nanoparticles or ultra-bright fluorescent organic polymers for single molecule detection.

5.Conclusions & perspectives

This thesis has accomplished its principal aim (discussed in section 4.1): the structural determination using cryo-EM of HFt-CD71 complex, an interaction widely exploited for nanotechnological applications of drug-delivery systems[38,40,41,42,43,44,47,48,51,52], but so far unexplained in atomic details.

The final cryo-EM map, obtained at 3.9 Å resolution (Figure 39), allowed to fit the atomic structures of the contact region without ambiguity: HFt and CD71 bind in a "virus-like" manner, *i.e.* HFt contacts the apical domain of CD71 (see Figure 10 for comparison).

HFt binding region (Figure 40, Figure S2 in Appendix 6.1) is constituted by the external BC loop (R79, F81, Q83, K86, K87) and the N-terminus of both A and C helices (T5; Q14, D15, E17-A19, N21, R22, N25; E116, K119, D123). From sequence alignments (Figure 44) and binding assays of mutated residues both *in vitro* (SPR, Figure 45) and *in vivo* (FACS and microscopy, Figure 46), we identified as HFt most crucial residues for hTfR1 binding in Q14, D15, R22, F81 and Q83. Notably, this is the first time that HFt in complex with another

protein has ever been imaged and resolved.
CD71 counterpart (Figure 40, Figure S2 in Appendix 6.1) is located on its apical domain and is established by epitopes that are exclusive for HFt binding (S195, E197, S199, I202, K374) or common to the recognition of several arenaviruses and of malaria plasmodium parasite *PvRBP2b* (R208-L212, N215, E343, K344, N348): compare Figure 12 with Figure 43. It is important to stress that to date the apical domain (present in all mammalian TfR1s[70]) has never had a known physiological role.

The CD71 interaction with HFt is so similar to the other known arenaviruses and parasites complexes (Figure 47) that we suggested that these pathogens may have adapted to enter cells by reproducing Ferritin physiological binding mode. Moreover, the involvement of the apical part for HFt recognition assures the independent and non-competitive binding of Tf and HFE on the same receptor, already been noticed[20,125], thus allowing differential regulation of iron. Finally, this very region might be deeply dependent on species-specific variants: if TfR1 is mutated in a species, the corresponding residues on Ferritin have to match to preserve its vital functions. For example, mouse TfR1 (mTfR1) apical domain is quite different to hTfR1 in HFt binding key residues (Figure 48), while mouse HFt is similar to human HFt: it is thus unclear whether iron uptake phenomena by humans and mice organisms may have diverged or not[20]. The only known receptor for mouse HFt is in fact TIM-2, not expressed in humans[24,25]. Proving this validation could be appealing for studying iron-regulatory mechanisms and to validate mice as animal models for iron-concerned

researches.

Two other side projects that deal with potential applications of this finding have been reached in the present thesis.

The first one (described in section 4.2) aimed to verify the three-dimensional cryo-EM structure of a novel chimera that our group engineered. This protein, the Humanized *A.Fulgidus* Ferritin (HumAfFt), is endowed with the reversible cation-dependent assembly/disassembly properties proper of AfFt cage[17,18] (Figure 50) and the cellular internalization by hTfR1 recognition[20] typical of HFt (Figure 56 and Figure 57). It has been possible to implement the CD71 interaction just by transplanting the human HFt BC loop in the AfFt sequence (Figure 58). This result confirms the importance of the loop for CD71 recognition; the gained binding is obtained also thanks to the natural presence of E14 and R22 in AfFt sequence.

Here, we obtained the structure of the chimera by X-ray crystallography (Figure 51) and by cryo-EM at 33 Å resolution (Figure 53), sufficient to assure its correct 24-mer assembly in near-native conditions.

Currently, a major effort is dedicated to the development of HumAFt as a more versatile carrier in the delivery of small protein-based payloads[135,136]. Extensive use of Cryo-EM measurements is fundamental in order to single out the structural determinants that govern the encapsulation by this very promising construct.

The aim of the second project (section 4.3) was to assess the correct insertion of lanthanide binding tags (LBTs) at the C-terminal domain of mouse HFt.

LBTs are short polypeptides[106] that are able to bind Tb^{3+} with high affinity; the tryptophan in their sequence confers strong FRET properties when bound to Terbium through an antenna effect (Figure 59 and 60). Cryo-EM was crucial for determining that LBTs are in "flip" conformation in the interior cavity of HFt (Figure 64): their low-resolution but defined patterns were enough to assure the presence of the tags in the occurred position. Being attached to E-helix, they do not interfere with CD71 binding region, allowing HFt-LBT Tb(III) uptake by several cancer cells, as verified by FACS and confocal microscopy (Figure 66 and S11).

The structural results here shown were obtained with single-particle cryo-EM: we have highlighted its importance for the determination of the three structures at different resolution, following the workflow in Figure 21.

Even before the resolution of the HFt/CD71 complex, experimental efforts were made trying to investigate the complex of HumAfFt with CD71 by cryo-EM. After the incubation of the proteins a pull-down step was not performed: yet, aggregation was not observed on the micrographs. Differently than the HFt/CD71 first attempt (Figure 25), the grids were not over-crowded and no precipitation could be seen; in most micrographs the proteins split apart and remained quite separated, though sizeable amounts of HumAfFt appeared to be in complex with CD71 (Figure 67 right).

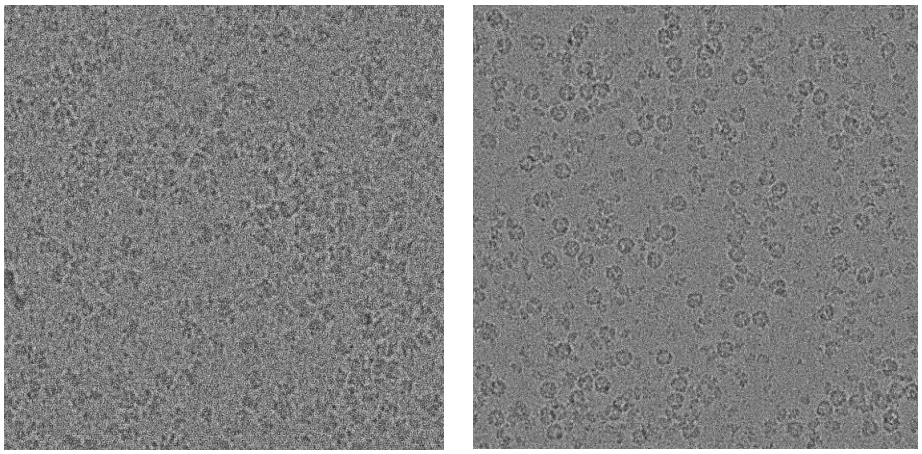


Figure 67 Effort of HumAfFt-CD71 cryo-EM acquisition on holey-gold grids, shown here after motion correction. The proteins did not form a complex and split apart. **Left**: micrograph with mostly receptors. **Right**: micrograph with mostly HumAfFts; many of them seem to be in complex with CD71.

Data analysis, performed with RELION 2.0, confirmed the unsuccessful occurrence of discrete and isolated complexes formation: in all the classes, CD71 and HumAfFt were separated:

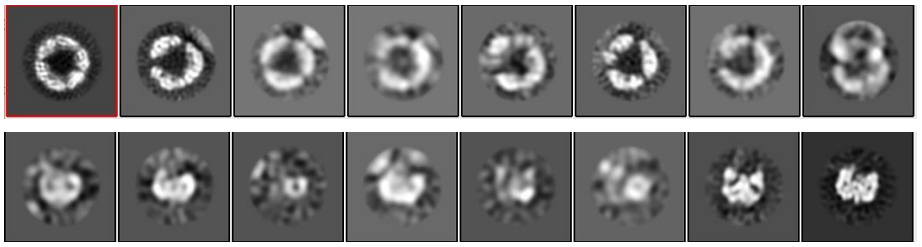


Figure 68 2D classification of HumAfFt-CD71 sample. Only HumAfFt or CD71 alone can be seen; as such, we may conclude that the proteins did not bind.

The final maps were obtained at 6.1 Å (HumAfFt) and 8.2 Å (hTfR1) resolutions:

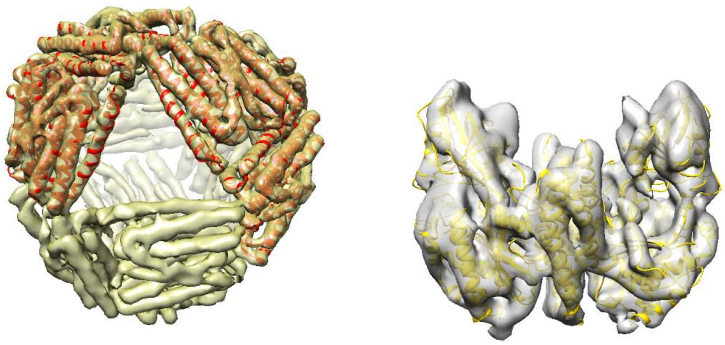


Figure 69 HumAfFt and CD71 cryo-EM reconstructions at 6.1 Å and 8.2 Å resolution. Notice the difference between this HumAfFt structure and the one in Figure 55, imaged with a CCD camera.

It is worth considering that the great difference between the HumAfFt sub-nanometer resolution structure here obtained and the one, at 33 Å, shown in chapter 4.2 (Figure 55) is mostly due to the direct detector acquisition instead of a CCD camera; the increased number of frames in this acquisition (200 frames *vs* 50 frames of HumAfFt alone) was probably another cause of the better reconstruction. In summary, though the HumAfFt-CD71 complex formation could be partially visible at the micrographs by naked eye, there were no significantly populated classes to reconstruct a map of the complex.

Consequently, as a last comment, the final part of this project will join our results on HFt/CD71 binding on HumAfFt structure. We will create a "twice chimeric" nanoparticle having point modifications in key residues for hTfR1 binding and not a generical BC loop substitution: these mutations will be targeted to increase the receptor affinity of the chimera towards CD71. In such manner, this protein will be equipped with an affinity to hTfR1 possibly higher than the nanomolar one of HFt and with a fully reversible cage

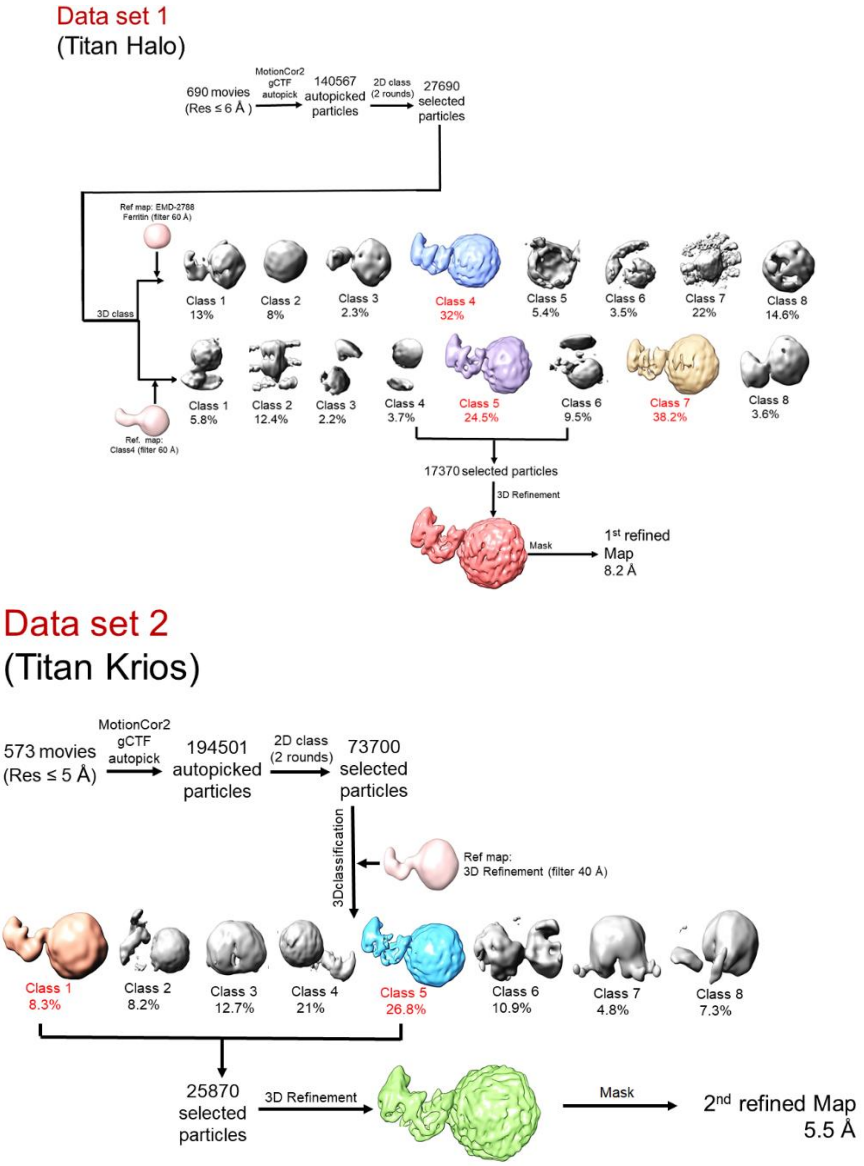
assembly/disassembly, thus allowing cargos encapsulation without the side effects typical of a pH jump.

The nanoparticle will likely increase the potential applications of Ferritins in biomedicine and represent a fundamental breakthrough for cancer treatment.

6. Appendix

6.1 CD71/HFt complex

Data analysis workflow



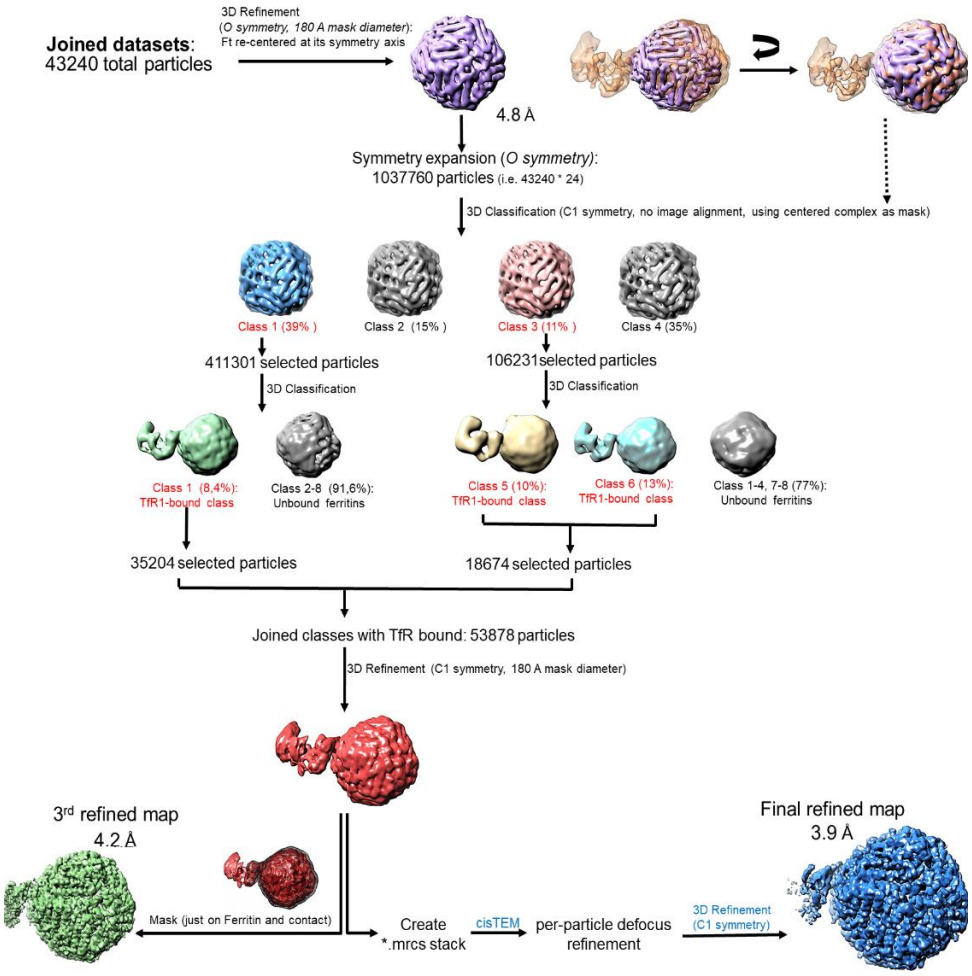


Figure S1 Complete workflow for CD717HFt complex.

CD71/HFt contacts

*CD71	Group	Location	H-Ft	Group	Location	Distance (Å)	+exclusive/ common
Electrostatic interactions (hydrogen bonds, salt bridges)							
Ser195	OG	loop βI-1-βII-1	Asp123	OD1	helix C	4.4	exclusive
Ser195	OG	loop βI-1-βII-1	Asp123	OD2	helix C	3.7	exclusive
Gln197	OE1	loop βI-1-βII-1	Lys119	NZ	helix C	3.3	exclusive
Gln197	NE2	loop βI-1-βII-1	Glu116	OE2	helix C	4.8	exclusive
Ser199	OG	strand βII-1	Asp15	OD1	helix A	4.8	exclusive
Ser199	OG	strand βII-1	Asp15	OD2	helix A	5.0	exclusive
Arg208	NH1	loop βII-1-βII-2	Thr5	N	N-term	3.7	common
Arg208	NH2	loop βII-1-βII-2	Thr5	O	N-term	4.0	common
Arg208	NH2	loop βII-1-βII-2	Arg79	NH2	loop BC	3.6	common
Val210	O	strand βII-2	Asn21	ND2	helix A	3.0	common
Asn215	OD1	loop βII-2-βII-3	Arg22	NH1	helix A	3.3	common
Asn215	OD1	loop βII-2-βII-3	Arg22	NH2	helix A	3.3	common
Asn215	ND2	loop βII-2-βII-3	Glu116	OE1	helix C	2.7	common
Asn215	ND2	loop βII-2-βII-3	Glu116	OE2	helix C	4.0	common
Lys344	NZ	helix αII-2	Asn25	OD1	helix A	3.6	common
Gly347	O	helix αII-2	*Lys87	NZ	*loop BC	3.6	common
Asn348	OD1	helix αII-2	*Lys87	NZ	*loop BC	3.0	common
Asn348	ND2	helix αII-2	Gln83	OE1	loop BC	3.5	common
Lys374	NZ	strand βII-8	Gln14	OE1	helix A	3.4	exclusive
Lys374	NZ	strand βII-8	Asp15	OD1	helix A	5.0	exclusive

*CD71	Group	Location	H-Ft	Group	Location	Distance (Å)	+exclusive/ common
			Hydrophobic interactions				
Ile201	CD1	strand βII-1	Ala18	CB	helix A	4.2	common
Ile201	CG2	strand βII-1	Ala18	CB	helix A	4.5	common
Leu209	CD1	strand βII-2	Gln14	CG	helix A	4.5	common
Leu209	CD2	strand βII-2	Arg79	CG	loop BC	3.8	common
Val210	CG1	strand βII-2	Gln14	CB	helix A	4.8	common
Val210	CG2	strand βII-2	Glu17	CB	helix A	4.4	common
Val210	CG2	strand βII-2	Glu17	CG	helix A	4.4	common
Val210	CG2	strand βII-2	Ala18	CB	helix A	4.4	common
Tyr211	CD1	strand βII-2	Asn21	CB	helix A	3.7	common
Tyr211	CE1	strand βII-2	Gln83	CB	loop BC	4.7	common
Tyr211	CE2	strand βII-2	Gln83	CG	loop BC	3.6	common
Tyr211	CE2	strand βII-2	Gln83	CD	loop BC	4.4	common
Tyr211	CD2	strand βII-2	Gln83	CD	loop BC	4.5	common
Tyr211	CG	strand βII-2	Phe81	CD2	loop BC	4.9	common
Leu212	CB	strand βII-2	Ala18	CB	helix A	3.6	common
Leu212	CB	strand βII-2	Arg22	CB	helix A	4.2	common
Leu212	CD2	strand βII-2	Ala19	CB	helix A	4.3	common
Leu212	CG	strand βII-2	Arg22	CG	helix A	4.7	common

Figure S2 Complete list of CD71/HFt contacts.
*: The colors follow the code of Figure 40.
+: the explanation common/exclusive is given in the text.

Cryo-EM maps refinement

	CD71/H-Ft complex at 5.5 Å **EMD-0046** **PDB 6GSR**	**CD71/H-Ft complex at 3.9 Å** **EMD-0140** **PDB 6H5I**
Refinement		
Initial model used (PDB code)	3KAS, 3AJO	3KAS, 3AJO
Model resolution (Å)	5.5	3.9
FSC threshold	0.143	0.143
Map sharpening *B* factor (Å^2)	-350	-272
Symmetry imposed	C1	C1, O
Model composition		
Non-hydrogen atoms	45308	43900
Protein residues	5392	5392
B factors (Å^2)		
Proteins (min, max)	21.4 (5.5, 181.1)	169.2 (71.5, 505.5)
R.m.s. deviations		
Bond lengths (Å)	0.0059	0.0065
Bond angles (°)	1.00	1.11
Validation		
MolProbity score	1.45	1.36
Clashscore	5.94	2.7
Poor rotamers (%)	1.20	0.17
EMRinger score	1.06	1.92
Ramachandran plot		
Favored (%)	97.73	95.69
Allowed (%)	2.15	4.31
Disallawed (%)	0.11	0

Figure S3 cryo-EM maps refinement for the 5.5 Å and 3.9 Å maps

Cryo-EM maps resolutions

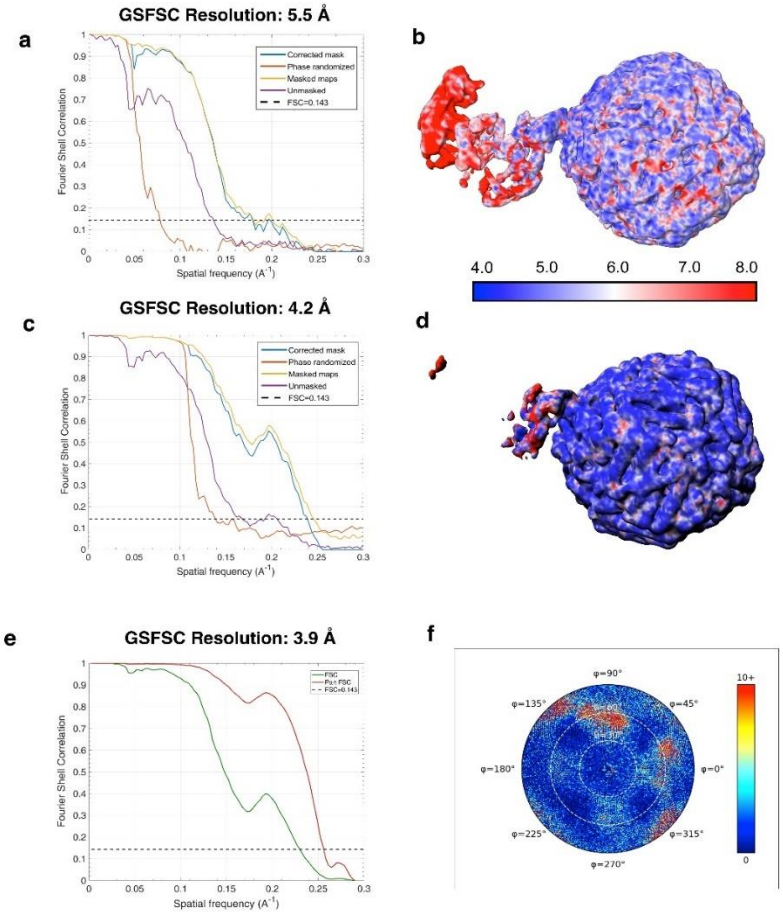


Figure S4 cryo-EM maps resolutions.

FACS acquisition plots

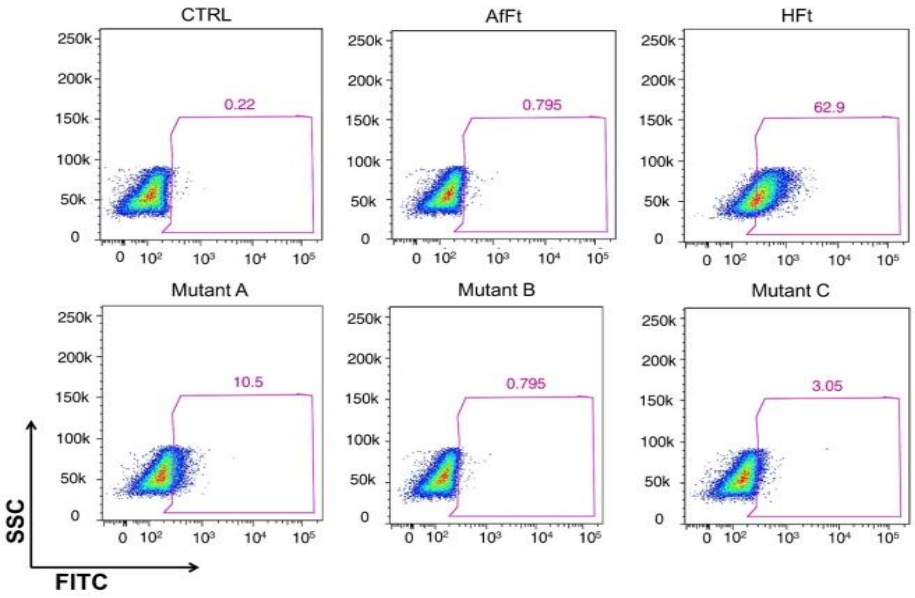


Figure S5 Representative FACS plot of the analysis shown in Figure 46.

6.2 Humanized HumAfFt

X-ray data collection and refinement

Data Collection	
Space Group	$C222_1$
Cell Dimensions a, b, c (Å)	185.80 190.65 176.54
Resolution (Å)	48.86 - 2.94 (3.00 - 2.94)
R_{merge}	0.134 (0.751)
Unique Reflections	59926 (4378)
$I/\sigma I$	7.8 (1.8)
Completeness (%)	95.8 (95.6)
Redundancy	4.2 (4.3)
Refinement	
Resolution (Å)	49.69 - 2.94 (3.00 - 2.94)
No. reflections	56823 (346)
R_{factor} R_{free}	0.27 (0.61) 0.30 (0.59)
Ion (Mg^{2+})	2
water	13
R.m.s deviations	
Bond lengths (Å)	0,007
Bond angles (°)	0,99

Figure S6 HumAfFt X-ray data collection and refinement.

Cryo-EM map FSC curve

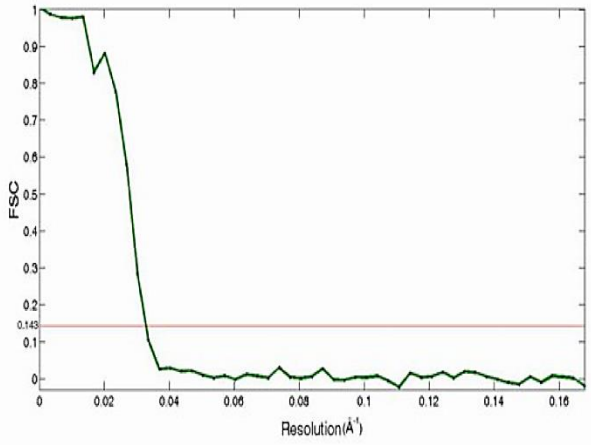


Figure S7 FSC curve of the 33 A map shown in Figure 55.

FACS acquisition plots

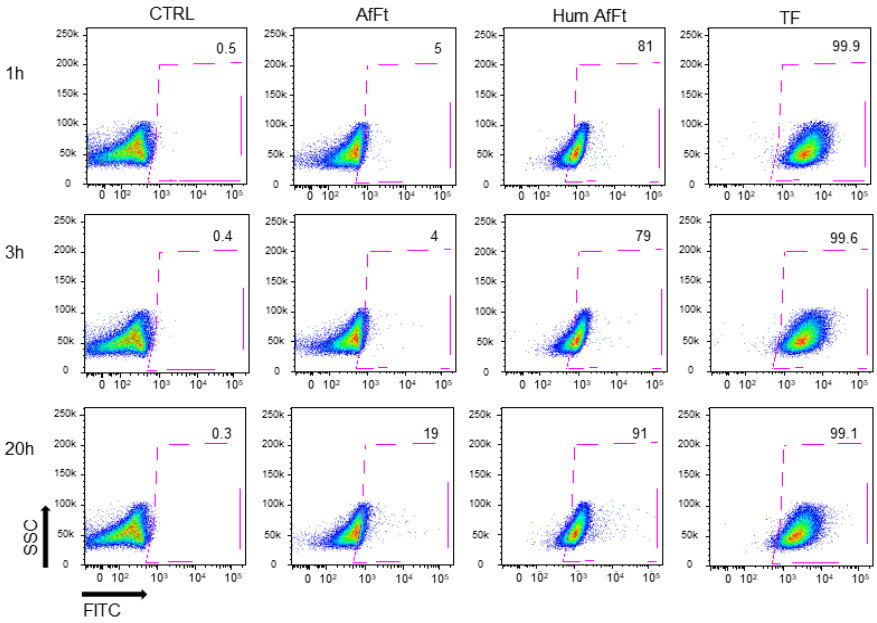


Figure S8 Representative FACS plot of the analysis shown in Figure 56.

6.3 Lanthanide Binding Ferritin

X-ray data collection and refinement

Data Collection	apoHFt_TBP	HFt_TBP_TB
Wavelength (Å)	1.0	1.4
Space group	I222	I222
Cell Dimensions	237.31 237.47 237.53 90.00 90.00 90.00	237.25 237.56 237.57 90.00 90.00 90.00
Resolution (Å)	48.48-2.85 (2.9-2.85)	48.48-2.65 (2.70-2.65)
R_{merge}	0.109 (0.833)	0.133 (1.135)
I/σI	17.6 (2.5)	12.6(1.4)
Completeness (%)	100 (100)	99.2 (94.3)
Redundancy	6.9 (6.8)	6.7 (5.3)
Refinement		
Resolution (Å)	48.48-2.85 (2.9-2.85)	48.48-2.65 (2.70-2.65)
No. unique reflections	61020 (4475)	191449 (8966)
R_{work}	0.137 (0.352)	0.164 (0.266)
R_{free}	0.167 (0.374)	0.176 (0.282)
Ligand/Ion	32 Iron Ions	32 Terbium Ions
Water	12	11
R.m.s. Deviations		
Bond length (Å)	0.004	0.004
Angles (°)	0.711	0.623

Figure S9 HFt-LBT Tb(III) X-ray data refinement.

FSC Cryo-EM map curve

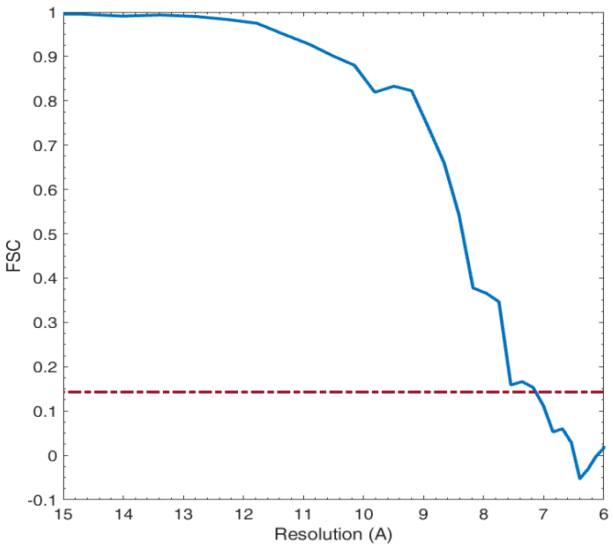


Figure S10 FSC curve of the cryo-EM map shown in Figure 63 and Figure 64

FACS and confocal microscopy data

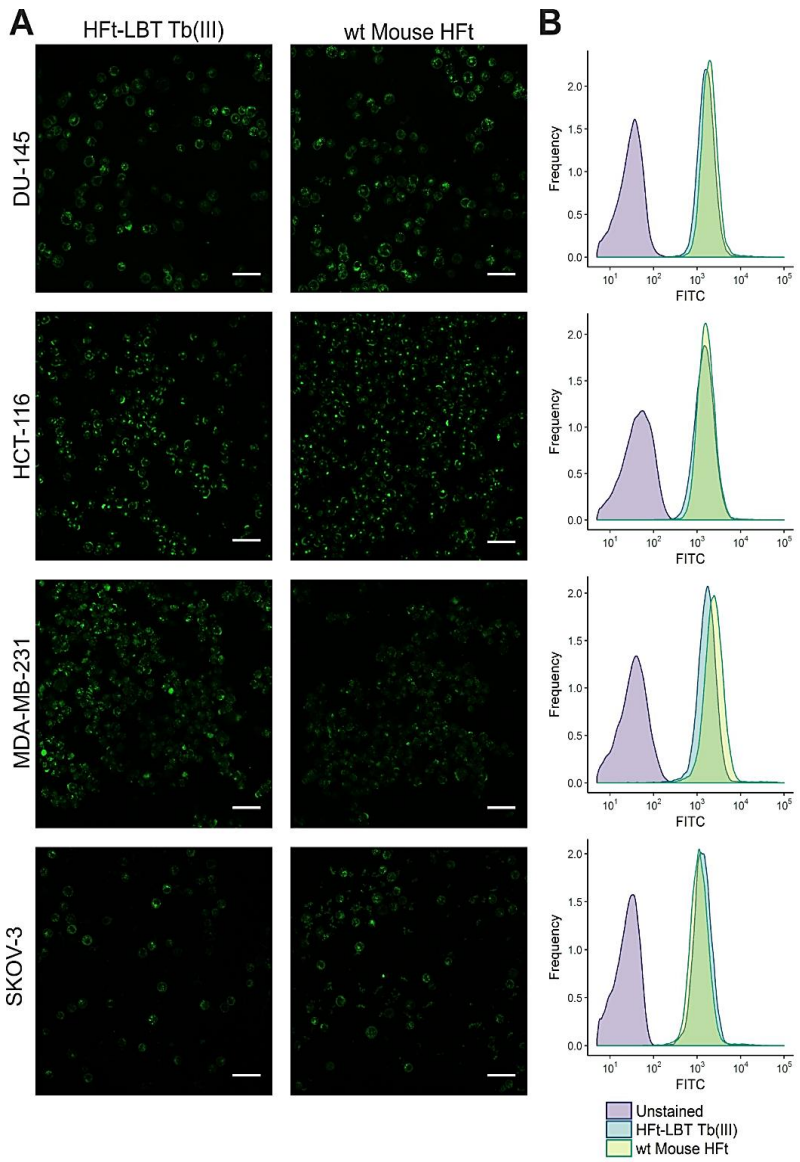


Figure S11 FACS and confocal data of Figure 66.

References

[1] P. M. Harrison, "The structure and function of ferritin" *Biochemical education,* vol. 14(4), pp. 154-162, 1986.

[2] K. H. Ebrahimi, E. Bill, P. L. Hagedoorn and W. R. Hagen, "The catalytic center of ferritin regulates iron storage via Fe(II)-Fe(III) displacement" *Nature chemical biology,* vol. 8(11), p. 941, 2012.

[3] A. R. Bogdan, M. Miyazawa, K. Hashimoto and Y. Tsuji, "Regulators of iron homeostasis: new players in metabolism, cell death, and disease" *Trends in biochemical sciences,* vol. 41(3), pp. 274-286, 2016.

[4] N. C. Andrews, "Disorders of iron metabolism" *New England Journal of Medicine,* vol. 341(26), pp. 1986-1995, 1999.

[5] B. Silva and P. Faustino, "An overview of molecular basis of iron metabolism regulation and the associated pathologies" *Biochimica et Biophysica Acta,* vol. 1852(7), pp. 1347-1359, 2015.

[6] J. K. Grady and N. D. Chasteen, "Some Speculations on the Role of Oxyradicals in the Conversion of Ferritin to Hemosiderin" *Iron Biominerals,* pp. 315-323, 1991.

[7] F. Bou-Abdallah, "The iron redox and hydrolysis chemistry of the ferritins" *Biochimica et Biophysica Acta,* vol. 1800, pp. 719-731, 2010.

[8] L. F. Dickey, S. Sreedharan, E. C. Theil, J. R. Didsbury, Y. H. Wang and R. E. Kaufman, "Differences in the regulation of messenger RNA for housekeeping and specialized-cell ferritin. A comparison of three distinct ferritin complementary DNAs, the corresponding subunits, and identification of the first processed in amphibia" *Journal of Biological Chemistry,* vol. 262(16), pp. 7901-7907, 1987.

[9] M. C. Sammarco, S. Ditch, A. Banerjee and E. Grabczyk, "Ferritin L and H subunits are differentially regulated on a post-transcriptional level" *Journal of Biological Chemistry,* vol. 283(8), pp. 4578-4587, 2008.

[10] K. H. Ebrahimi, P. L. Hagedoorn and W. R. Hagen, "Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin" *Chemical reviews,* vol. 115(1), pp. 295-326, 2014.

[11] D. J. Price and J. G. Joshi, "Ferritin. Binding of beryllium and other divalent metal ions" *Journal of Biological Chemistry,* vol. 258(18), pp. 10873-10880, 1983.

[12] M. Truffi, L. Fiandra, L. Sorrentino, M. Monieri, F. Corsi, S. Mazzucchelli, "Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer" *Pharmacol Res,* vol. 107, pp. 57-65, 2016.

[13] R. R. Crichton and C. F. Bryce, "Subunit interactions in horse spleen apoferritin. Dissociation by extremes of pH" *Biochemical Journal,* vol. 133(2), pp. 289-299, 1973.

[14] S. Stefanini, P. Vecchini and E. Chiancone, "On the mechanism of horse spleen apoferritin assembly: a sedimentation velocity and circular dichroism study" *Biochemistry,* vol. 26(7), pp. 1831-1837, 1987.

[15] S. Levi, A. Luzzago, F. Franceschinelli, P. Santambrogio, G. Cesareni and P. Arosio, "Mutational analysis of the channel and loop sequences of human ferritin H-chain" *Biochemical Journal,* vol. 264(2), pp. 381-388, 1989.

[16] D. Sato, H. Ohtomo, Y. Yamada, T. Hikima, A. Kurobe, K. Fujiwara and M. Ikeguchi, "Ferritin assembly revisited: a time-resolved small-angle X-ray scattering study" *Biochemistry,* vol. 55(2), pp. 287-293, 2016.

[17] J. Swift, C. A. Butts, J. Cheung-Lau, V. Yerubandi and I. J. Dmochowski, "Efficient self-assembly of Archaeoglobus fulgidus ferritin around metallic cores" *Langmuir,* vol. 25(9), pp. 5219-5225, 2009.

[18] E. Johnson, D. Cascio, M. R. Sawaya, M. Gingery and I. Schröder, "Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic archaeon Archaeoglobus fulgidus" *Structure,* vol. 13(4), pp. 637-648, 2005.

[19] B. Todorich, X. Zhang and J. Connor, "H-ferritin is the Major Source of Iron for Oligodendrocytes" *GLIA,* vol. 59, pp. 927-935, 2011.

[20] L. Li, C. J. Fang, J. C. Ryan, E. C. Niemi, J. A. Lebrón, P. J. Björkman ... and W. E. Seaman, "Binding and uptake of H-ferritin are mediated by human transferrin receptor-1" *Proceedings of the National Academy of Sciences,* vol. 107(8), pp. 3505-3510, 2010.

[21] Y. Cheng, O. Zak, P. Aisen, S.C. Harrison, and T. Walz, "Single particle reconstruction of the human apo-transferrin-transferrin receptor complex." *J Struct Biol ,* vol. 152(3), pp. 204-210, 2005.

[22] Y. Cheng, O. Zak, P. Aisen, S.C. Harrison, and T. Walz, "Structure of the human transferrin receptor-transferrin complex" *Cell,* vol. 116(4), pp. 565-576, 2004.

[23] M. J. Bennett, J. A. Lebrón and P. J. Bjorkman, "Crystal structure of the hereditary haemochromatosis protein HFE complexed with transferrin receptor" *Nature,* vol. 403(6765), p. 46, 2000.

[24] T. T. Chen, L. Li, D.-H. Chung, C. D. Allen, S. V. Torti, … and M. Daws, "TIM-2 is expressed on B cells and in liver and kidney and is a receptor for H-ferritin endocytosis" *The Journal of Experimental Medicine,* 2005.

[25] B. Todorich, X. Zhang, B. Slagle-Webb, W. Seaman and J. Connor, "Tim–2 is the receptor for H-ferritin on oligodendrocytes." *Journal of Neurochemistry,* vol. 107(6), p. 1495–505, 2008.

[26] B. Bhushan, S. U. Kumar, I. Matai, A. Sachdev, P. Dubey and P. Gopinath, "Ferritin nanocages: a novel platform for biomedical applications.," *Journal of biomedical nanotechnology,* vol. 10(10), pp. 2950-2976, 2014.

[27] S. Zhang, J. Zang, H. Chen, M. Li, C. Xu and G. Zhao, "The size flexibility of ferritin nanocage opens a new way to prepare nanomaterials" *small,* vol. 13(37), p. 1701045.

[28] S. Abe, Maity, B. and T. Ueno, "Design of a confined environment using protein cages and crystals for the development of biohybrid materials" *Chem Commun (Camb),* vol. 52(39), pp. 6496-6512, 2016.

[29] Z. Zhen, W. Tang, T. Todd and J. Xie, "Ferritins as nanoplatforms for imaging and drug delivery" *Expert opinion on drug delivery,* vol. 11(12), pp. 1913-1922, 2014.

[30] E. Blanco, H. Shen and M. Ferrari, "Principles of nanoparticle design for overcoming biological barriers to drug delivery" *Nature biotechnology,* vol. 33(9), p. 941, 2015.

[31] H. S. Choi, W. Liu, P. Misra, E. Tanaka, J. P. Zimmer, B. Ipe, … and J. V. Frangioni, "Renal clearance of quantum dots" *Nat. Biotechnol.,* vol. 25, pp. 1165-1170, 2007.

[32] F. Braet, E. Wisse, … and S. Ringer, "Contribution of high-resolution correlative imaging techniques in the study of liver sieve in three-dimensions" *Microsc. Res. Tech,* vol. 70, pp. 230-242, 2007.

[33] Z. Wang, H. Gao, Y. Zhang, G. Liu, G. Niu and X. Chen, "Functional ferritin nanoparticles for biomedical applications" *Frontiers of chemical science and engineering,* pp. 1-14, (2017).

[34] M. Ferrari, "Cancer nanotechnology: opportunities and challenges" *Nature reviews cancer,* vol. 5(3), p. 161, 2005.

[35] W. H. De Jong and P. J. Borm, "Drug delivery and nanoparticles: Applications and hazards" *International Journal of Nanomedicine,* vol. 3(2), pp. 133-149, 2008.

[36] S. Marcazzan, E. M. Varoni, E. Blanco, G. Lodi and M. Ferrari, "Nanomedicine, an emerging therapeutic strategy for oral cancer therapy" *Oral oncology,* vol. 76, pp. 1-7, 2018.

[37] S. Nizzero, A. Ziemys and M. Ferrari, "Transport Barriers and Oncophysics in Cancer Treatment" *Trends in cancer,* vol. 4(4), pp. 277-280, 2018.

[38] G. Fracasso, E. Falvo, G. Colotti, … and P. Ceci, "Selective delivery of doxorubicin by novel stimuli-sensitive nano-ferritins overcomes tumor refractoriness" *J Control Release,* vol. 239, pp. 10-18, 2016.

[39] S. Tortorella and T. C. Karagiannis, "Transferrin receptor-mediated endocytosis: a useful target for cancer therapy" *The Journal of membrane biology,* vol. 247(4), pp. 291-307, 2014.

[40] Z. Zhen, W. Tang, C. Guo, H. Chen, X. Lin, G. Liu ...and J. Xie, "Ferritin nanocages to encapsulate and deliver photosensitizers for efficient photodynamic therapy against cancer" *ACS nano,* vol. 7(8), pp. 6988-6996, 2013.

[41] Y. Lei, Y. Hamada, J. Li, L. Cong, N. Wang, Y. Li ...and X. Jiang, "Targeted tumor delivery and controlled release of neuronal drugs with ferritin nanoparticles to regulate pancreatic cancer progression" *Journal of Controlled Release,* vol. 232, p. 13, 2016.

[42] E. Falvo, E. Tremante, R. Fraioli, C. Leonetti, C. Zamparelli, A. Boffi …and P. A. Giacomini, "Antibody–drug conjugates: targeting melanoma with cisplatin encapsulated in protein-cage nanoparticles based on human ferritin" *Nanoscale,* vol. 5(24), pp. 12278-12285, 2013.

[43] C. Sun, H. Yang, Y. Yuan, X. Tian, L. Wang, Y. Guo … and X. J. Liang, "Controlling assembly of paired gold clusters within apoferritin nanoreactor for in vivo kidney targeting and biomedical imaging." *Journal of the American Chemical Society,* vol. 133(22), pp. 8617-8624, 2011.

[44] X. Lin, J. Xie, G. Niu, Z. F. H. Gao, Y. M. … and R. Leapman, "Chimeric ferritin nanocages for multiple function loading and multimodal imaging" *Nano Letters,* vol. 11(2), pp. 814-819., 2011.

[45] M. Kim, Y. Rho, K. S. Jin, B. Ahn, S. Jung, H. Kim and M. Ree, "pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly" *Biomacromolecules,* vol. 12(5), pp. 1629-1640, 2011.

[46] V. de Turris, M. C. Trabuco, G. Peruzzi, A. Boffi, C. Testi, ... and P. Baiocco, "Humanized archaeal ferritin as a tool for cell targeted delivery" *Nanoscale,* vol. 9(2), pp. 647-655, 2017.

[47] E. Falvo, E. Tremante, A. Arcovito, M. Papi … and P. Ceci, "Improved doxorubicin encapsulation and pharmacokinetics of ferritin-fusion protein nanocarriers bearing proline, serine and alanine elements" *Biomacromolecules ,* vol. 17(2), pp. 514-522, 2016.

[48] X. Huang, J. Chisholm, J. X. Zhuang, D. G. Y., X. Chen ... and J. Hanes, "Protein nanocages that penetrate airway mucus and tumor tissue" *Proceedings of the National Academy of Sciences,* vol. 114(32), pp. E6595-E6602., 2017.

[49] M. A. Kostiainen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen and P. Ceci, "Electrostatic assembly of binary nanoparticle superlattices using protein cages" *Nature nanotechnology,* vol. 8(1), p. 52, 2013.

[50] M. Kanekiyo, C. J. Wei, H. M. Yassine, P. M. McTamney, J. C. Boyington, J. R. Whittle ... and G. J. Nabel, "Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies" *Nature,* vol. 499(7456), p. 102, 2013.

[51] K. Fan, Cao, C., Pan, Y., Lu, D., Yang, D., Feng, J., Song, L., Liang, Y. M. and X., "Magnetoferritin nanoparticles for targeting and visualizing tumour tissues." *Nat Nanotechnol ,* vol. 7(7), pp. 459-464 , 2012.

[52] C. Cao, X. Wang, Y. Cai, L. Sun, L. Tian, H. Wu, ... and R. Zhu, "Targeted in vivo imaging of microscopic tumors with ferritin-based nanoprobes across biological barriers" *Advanced Materials,,* vol. 26(16), pp. 2566-2571, 2014.

[53] L. A. Cohen, L. Gutierrez, A. Weiss, Leichtmann-Bardoogo, Z. D. L. Y., D. Crooks ... and F. J. Lazaro, "Serum ferritin is derived primarily from macrophages through a non-classical secretory pathway" *Blood,* 2010.

[54] R. Aleksandrowicz, B. Taciak and M. Krol, "Drug delivery systems improving chemical and physical properties of anticancer drugs currently investigated for treatment of solid tumors" *Journal of Physiology and Pharmacology,* vol. 68(2), pp. 165-174, 2017.

[55] T. R. Daniels, E. Bernabeu, J. A. Rodríguez, S. Patel, M. Kozman, D. A. Chiappetta ... and M. L. Penichet, "The transferrin receptor and the targeted delivery of therapeutic agents against cancer" *Biochimica et Biophysica Acta (BBA),* vol. 1820, pp. 291-317, 2012.

[56] M. Liang, K. Fan, M. Zhou, D. Duan, J. Zheng, D. Yang … and X. Yan, "H-ferritin–nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection" *Proceedings of the National Academy of Sciences,* vol. 111(41), pp. 14900-14905, 2014.

[57] E. Gammella, P. Buratti, G. Cairo and S. Recalcati, "The transferrin receptor: the cellular iron gate." *Metallomics,* vol. 9(10), pp. 1367-1375, 2017.

[58] C. M. Lawrence, Ray, S., Babyonyshev, M., Galluser, R., Borhani, D.W. and S. Harrison, "Crystal structure of the ectodomain of human transferrin receptor" *Science,* vol. 286(5440), pp. 779-782, 1999.

[59] E. F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng and T. Ferrin, "UCSF Chimera--a visualization system for exploratory research and analysis" *J Comput Chem,* vol. 25(13), pp. 1605-1612, 2004.

[60] J. A. Lebrón, M.J. Bennett, D.E. Vaughn, A.J. Chirino, P.M. Snow, … and P. Bjorkman, "Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with Transferrin Receptor" *Cell ,* vol. 93, pp. 111-123 , 1998.

[61] J. Abraham, K. Corbett, M. Farzan, H. Choe and S. Harrison, "Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses." *Nat Struct Mol Biol 17,* p. 438–444, 2010.

[62] J. Gruszczyk, R.K. Huang, Chan … and W.H. Tham, "Cryo-EM structure of an essential Plasmodium vivax invasion complex" *Nature,* vol. 559(7712), pp. 135-139, 2018.

[63] K. M. Mayle, A. Le and D. Kamei, "The intracellular trafficking pathway of transferrin" *Biochim Biophys Acta,* vol. 1820(3), pp. 264-281, 2012.

[64] H. T. McMahon and E. Boucrot, "Molecular mechanism and physiological functions of clathrin-mediated endocytosis" *Nature reviews Molecular cell biology,* vol. 12(8), p. 517, 2011.

[65] T. R. Daniels-Wells and M. Penichet, "Transferrin receptor 1: a target for antibody-mediated cancer therapy" *Immunotherapy,* vol. 8(9), pp. 991-994 , 2016.

[66] A. Demogines, J. Abraham, H. Choe, M. Farzan and S. Sawyer, "Dual host-virus arms races shape an essential housekeeping protein." *PLoS Biol 11(5), e1001571,* 2013.

[67] S. A. Kerr, E. Jackson, O. Lungu, A. Meyer, Demogines … and S. Sawyer, "Computational and Functional Analysis of the Virus-Receptor Interface Reveals Host Range Trade-Offs in New World Arenaviruses" *J Virol,* pp. 89(22), 11643-11653, 2015.

[68] S. R. Radoshitzky, J. Abraham, C. Spiropoulou, J. Kuhn ... and H. Choe, "Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses" *Nature,* vol. 446, pp. 92-96, 2007.

[69] S. R. Radoshitzky, L. Longobardi, J. Kuhn ... and S. Bavari, "Machupo virus glycoprotein determinants for human transferrin receptor 1 binding and cell entry" *Plos One,* vol. 6, p. e21398, 2011.

[70] S. R. Radoshitzky, J.H. Kuhn, C.F. Spiropoulou, …and M. Farzan, "Receptor determinants of zoonotic transmission of New World

hemorrhagic fever arenaviruses" *Proc Natl Acad Sci U S A,* vol. 105(7), pp. 2664-2669, 2008.

[71] S. Hafenstein, L. Palermo, V. Kostyuchenko, C. Xiao, M. Morais … and M. Rossmann, "Asymmetric binding of transferrin receptor to parvovirus capsids" *Proc Natl Acad Sci U S A,* vol. 104(16), pp. 6585-6589, 2007.

[72] D. N. Martin and S. Uprichard, "Identification of transferrin receptor 1 as a hepatitis C virus entry factor." *Proc Natl Acad Sci U S A,* vol. 110(26), pp. 10777-10782, 2013.

[73] J. Gruszczyk, U. Kanjee, L. Chan and W. Tham, "Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax." *Science,* vol. 359(6371), pp. 48-55, 2018.

[74] S. Singh, A. DuPage, A. Y. Weaver, J. Sagert, C. White, M. Krimm and J. Richardson, "Development of a probody drug conjugate (PDC) targeting CD71 for the treatment of solid tumors and lymphomas" in *AACR 107th Annual Meeting 2016*, New Orleans, 2016.

[75] T. R. Daniels, T. Delgado, G. Helguera and M. L. Penichet, "The transferrin receptor part II: targeted delivery of therapeutic agents into cancer cells" *Clinical immunology,* vol. 121(2), pp. 159-176, 2006.

[76] T. R. Daniels, T. Delgado, J. A. Rodriguez, G. Helguera and M. L. Penichet, "The transferrin receptor part I: Biology and targeting with cytotoxic antibodies for the treatment of cancer" *Clinical immunology,* vol. 121(2), pp. 144-158, 2006.

[77] C. J. Greene, K. Attwood, N.J. Sharma, K.W. Gross, G.J. Smith, B. Xu and E. Kauffman, "Transferrin receptor 1 upregulation in primary tumor and downregulation in benign kidney is associated with progression and mortality in renal cell carcinoma patients" *Oncotarget,* vol. 8(63), pp. 107052-107075, 2017.

[78] C. R. Hopkins and I. S. Trowbridge, " Internalization and processing of transferrin and the transferrin receptor in human carcinoma A431 cells" *The Journal of cell biology,* vol. 97(2), pp. 508-521, 1983.

[79] Y. Cheng, "Single-particle cryo-EM at crystallographic resolution" *Cell,* vol. 161(3), pp. 450-457, 2015.

[80] J. Drenth, Principles of protein X-ray crystallography, Springer Science & Business Media, 2007.

[81] R. Fernandez-Leiro and S. Scheres, "Unravelling the structures of biological macromolecules by cryo-EM" *Nature,* vol. 537(7620), pp. 339-346, 2016.

[82] A. McPherson Jr, "The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis" *Methods of biochemical analysis,* pp. 249-345, 1976.

[83] A. Doerr, "Single-particle cryo-electron microscopy" *Nature Merthods,* vol. 13(1), p. 23, 2016.

[84] K. R. Vinothkumar and R. Henderson, "Single particle electron cryomicroscopy: trends, issues and future perspective." *Quarterly reviews of biophysics,* vol. 49, 2016.

[85] M. Khoshouei, M. Radjainia, W. Baumeister and R. Danev, "Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate" *Nature communications,* vol. 8, p. 16099, 2017.

[86] Nature Editorial, "Method of the Year 2015" *Nature Methods,* vol. 13(1), p. 1, 2016.

[87] A. des Georges, Y. Hashem, S. N. Buss, F. Jossinet, Q. Zhang, H. Y. Liao ... and C. Bajaj, "High-resolution cryo-EM structure of the Trypanosoma brucei Ribosome: a case study" *Computational Methods for Three-Dimensional Microscopy Reconstruction,* pp. 97-132, 2014.

[88] C. J. Russo and L. Passmore, "Ultrastable gold substrates for electron cryomicroscopy" *Science,* vol. 346(6215), pp. 1377-1380, 2014.

[89] R. F. Thompson, M. Walker, C. A. Siebert, S. P. Muench and N. A. Ranson, "An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology" *Methods,* vol. 100, pp. 3-15, 2016.

[90] L. Wang and F. J. Sigworth, "CryoEM and single particles" *Physiology,* vol. 21, pp. 13-18, 2006.

[91] J. Dubochet, J. Chang, R. Freeman, J. Lepault and A. McDowall, "Frozen aqueous suspensions" *Ultramicroscopy,* vol. 10, pp. 55-61, 1982.

[92] J. Dubochet, M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, J. McDowall and P. A.W Schultz, "Cryo-electron microscopy of vitrified specimens" *Q Rev Biophys,* vol. 21(2), p. 129–228, 1998.

[93] J. Frank, Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state, Oxford University Press, 2006.

[94] D. B. Murphy, Fundamentals of light microscopy and electronic imaging, John Wiley & Sons., 2002.

[95] L. Schermelleh, R. Heintzmann and H. Leonhardt, "A guide to super-resolution fluorescence microscopy" *The Journal of cell biology,* vol. 190(2), pp. 165-175, 2010.

[96] J. L. Milne, M. J. Borgnia, A. Bartesaghi, E. E. Tran, L. A. Earl, D. M. Schauder ... and S. Subramaniam, "Cryo-electron microscopy–a primer for the non-microscopist." *The FEBS journal,* vol. 280(1), pp. 28-45, 2013.

[97] M. Carroni and H. R. Saibil, "Cryo electron microscopy to determine the structure of macromolecular complexes" *Methods,* vol. 95, pp. 78-85, 2016.

[98] S. Q. Zheng, Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y. and D. Agard, "MotionCor2: anisotropic correction of beam-induced motion

for improved single-particle electron cryo-microscopy" *Nat Methods,* vol. 14(4), p. 331–332, 2017.

[99] J. L. Rubinstein and M. A. Brubakerd, "Alignment of cryo-EM movies of individual particles by optimization of image translations" *Journal of Structural Biology,* vol. 192(2), pp. 188-195, 2015.

[100] K. Zhang, "Gctf: Real-time CTF determination and correction" *J Struct Biol ,* vol. 193(1), pp. 1-12 , 2016.

[101] S. H. W. Scheres, "Semi-automated selection of cryo-EM particles in RELION-1.3" *J Struct Biol ,* vol. 189(2), pp. 114-122, 2015.

[102] J. Vargas, J. Otón, R. Marabini, J. M. Carazo and C. O. S. Sorzano, "Particle alignment reliability in single particle electron cryomicroscopy: a general approach" *Scientific reports,* vol. 6, p. 21626, 2016.

[103] D. Lyumkis, A. F. Brilot, D. L. Theobald and N. Grigorieff, "Likelihood-based classification of cryo-EM images using FREALIGN" *Journal of structural biology,* vol. 183(3), pp. 377-388, 2013.

[104] B. A. Barad, N. Echols, R.Y. Wang, Y. Cheng, F. DiMaio, P.D. Adams and J. Fraser, "EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy" *Nat Methods ,* vol. 12(10), pp. 943-946, 2015.

[105] A. C. Dalton and W. A. Barton., "Over-expression of secreted proteins from mammalian cell lines" *Protein Science,* vol. 23(5), pp. 517-525, 2014.

[106] L. Martin, M. Hahnke, M. Nitz, J. Wohnert, R. Nicholas, A. K. Silvaggi, H. Schwalbe and B. Imperiali, "Double-Lanthanide-Binding Tags: Design, Photophysical Properties and NMR Applications" *J. Am. Chem. Soc.,* vol. 129(22) , pp. 7106-7113, 2007.

[107] C. Suloway, J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C.S. Potter and B. Carragher, "Automated molecular microscopy: the new Leginon system" *Struct Biol ,* vol. 151(1), pp. 41-60, 2005.

[108] R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko and W. Baumeister, "Volta potential phase plate for in-focus phase contrast transmission electron microscopy" *Proceedings of the National Academy of Sciences,* vol. 111(44), pp. 15635-15640, 2014.

[109] S. H. W. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination" *J Struct Biol,* vol. 180(3), pp. 519-530, 2012.

[110] T. Grant, A. Rohou and N. Grigorieff, "cisTEM; user-friendly software for single-particle image processing" *Elife 7, e35383 ,* 2018.

[111] S. Chen, G. McMullan, A.R. Faruqi, G.N. Murshudov, J.M. Short, S.H. Scheres and R. Henderson, "High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy" *Ultramicroscopy,* vol. 135, pp. 24-35, 2013.

[112] P. Rosenthal and R. Henderson, "Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy" *J. Mol. Biol. ,* vol. 333, pp. 721-745, 2003.

[113] S. H. Scheres and S. Chen, "Prevention of overfitting in cryo-EM structure determination" *Nature methods,* vol. 9(9), p. 853, 2012.

[114] A. Kucukelbir, F.J. Sigworth and H. D. Tagare, "Quantifying the local resolution of cryo-EM density maps" *Nat Methods ,* vol. 11(1), pp. 63-65, 2014.

[115] P.D. Adams, P.V. Afonine ...and P. Zwart, "PHENIX: a comprehensive Python-based system for macromolecular structure solution" *Acta Crystallogr D Biol Crystallogr,* vol. 66, pp. 213-22, 2010.

[116] P. Emsley, B. Lohkamp, W.G. Scott and K. Cowtan, "Features and development of Coot" *Acta Crystallogr D Biol Crystallogr,* vol. 66, pp. 486-501, 2010.

[117] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino ... and D. C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography" *Acta Crystallographica Section D: Biological Crystallography,* vol. 66(1), pp. 12-21, 2010.

[118] W. Kabsch, "Xds" *Acta Crystallographica Section D: Biological Crystallography,* vol. 66(2), pp. 125-132, 2010.

[119] M. D. Winn, C. C. Ballard ... and S.J. McNicholas "Overview of the CCP4 suite and current developments" *Acta Crystallogr. D Biol. Crystallogr.,* vol. 67(4), pp. 235-242, 2011.

[120] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls ... and A. A. Vagin, "REFMAC5 for the refinement of macromolecular crystal structures" *Acta Crystallographica Section D: Biological Crystallography,* vol. 67(4), pp. 355-367, 2011.

[121] R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures" *Journal of applied crystallography,* vol. 26(2), pp. 283-291, 1993.

[122] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, "Phaser crystallographic software" *Journal of applied crystallography,* vol. 40(4), pp. 658-674, 2007.

[123] M. Morlando, S. D. Modigliani, G. Torrelli, A. Rosa, V. Di Carlo, E. Caffarelli and I. Bozzoni, "FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment" *The EMBO journal,* vol. 31(24), pp. 4502-4510, 2012.

[124] M. Abramoff, P. Magalhaes and S. Ram, "Image Processing with ImageJ" *Biophotonics International,* p. 11, 2004.

[125] S. Sakamoto, H. Kawabata, T. Masuda, T. Uchiyama and C. Mizumoto "H-Ferritin Is Preferentially Incorporated by Human Erythroid Cells

through Transferrin Receptor 1 in a Threshold-Dependent Manner" *PLOS ONE,* p. 10(10), 2015.

[126] S. H. Scheres, "Processing of structurally heterogeneous cryo-EM data in RELION" *Methods in enzymology,* vol. 579, pp. 125-157, 2016.

[127] M. Zhou, Y. Li, Q. Hu …and Y. Shi, "Atomic structure of the apoptosome: mechanism of cytochrome c-and dATP-mediated activation of Apaf-1" *Genes Dev,* vol. 29(22), pp. 2349-2361, 2015.

[128] X. Robert and P. Gouet, "Deciphering key features in protein structures with the new ENDscript server" *Nucleic Acids Res.,* vol. 42, p. 320–324, 2014.

[129] E. Wang, L. Albritton and S. R. Ross, "Identification of the segments of the mouse transferrin receptor 1 required for mouse mammary tumor virus infection" *Journal of Biological Chemistry,* vol. 281(15), pp. 10243-10249, 2006.

[130] K. J. McKee, B. Mahlandt, J. Maiztegui, G. Eddy and CJ Peters, "Experimental Argentine hemorrhagic fever in rhesus macaques: Viral strain-dependent clinical response" *J Infect Dis ,* vol. 152, p. 218–221, 1985.

[131] H. Chen, S. Zhang, C. Xu and G. Zhao, "Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions" *Chemical Communications,* vol. 52(46), pp. 7402-7405, 2016.

[132] S. Kim, J. O. Jeon, E. Jun, J. Jee, H. K. Jung, B. H. Lee ... and S. Kim, "Designing peptide bunches on nanocage for bispecific or superaffinity targeting" *Biomacromolecules,* vol. 17(3), pp. 1150-1159, 2016.

[133] B. Sana, E. Johnson, P. Le Magueres, A. Criswell, D. Cascio and S. Lim, "The role of non-conserved residues of Archaeoglobus fulgidus ferritin on its unique structure and biophysical properties" *Journal of Biological Chemistry,* vol. 288, p. 32663–32672, 2013.

[134] S. Levi, A. Luzzago, G. Cesareni, A. Cozzi, F. Franceschinelli, A. Albertini and P. Arosio, "Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferroxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants" *J. Biol. Chem.,* vol. 263, p. 18086–18092, 1988.

[135] I. Benni, M. C. Trabuco, E. Di Stasio, A. Arcovito, A. Boffi, F. Malatesta ... and P. Baiocco, "Excimer based fluorescent pyrene–ferritin conjugate for protein oligomerization studies and imaging in living cells" *RSC Advances,* vol. 8(23), pp. 12815-128, 2018.

[136] S. Tetter and D. Hilvert, "Enzyme Encapsulation by a Ferritin Cage" *Angewandte Chemie International Edition,* vol. 56(47), pp. 14933-14936., 2017.

[137] M. H. V. Werts, "Making sense of lanthanide luminescence" *Science Progress,* vol. 88(2), pp. 101-131, 2005.

[138] H. Handl and R. Gillies, "Lanthanide-based luminescent assays for ligand-receptor interactions" *J. Life Sci.,* vol. 77, pp. 361-371, 2005.

[139] E. G. Moore, A. P. Samuel and K. N. Raymond, "From antenna to assay: lessons learned in lanthanide luminescence" *Accounts of chemical research,* vol. 42(4), pp. 542-552, 2009.

[140] E. Gudgin Dickson, P. A and D. EP, "Time-resolved detection of lanthanide luminescence for ultrasensitive bioanalytical assays" *Photochem.Photobiol. B ,* vol. 27(1), pp. 3-19, 1995.

[141] PR Selvin, "Principles and biophysical applications of lanthanide-based probes" *Annu. Rev. Biophys. Biomembr.,* vol. 42, pp. 39-75, 2002.

[142] K. Franz, M. Nitz and B. Imperiali, "Lanthanide-binding tags as versatile protein coexpression probes" *Chembiochem.,* vol. 4(4) , pp. 265-271, 2003.

[143] M. Nitz, M. Sherawat, K. Franz, E. Peisach, K. Allen and B. Imperiali, "Structural origin of the high affinity of a chemically evolved lanthanide-

binding peptide" *Angew. Chem. Int. Ed. Engl.* , vol. 43(28) , pp. 3682-3685, 2004.

[144] A. Barb, T. G. Ho, H. Flanagan-Street and J. Prestegard, "Lanthanide binding and IgG affinity construct: Potential applications in solution NMR, MRI, and luminescence microscopy" *Protein Sci.* , vol. 21(10) , pp. 1456-1466, 2012.

[145] N. Goda, T. Tenno, K. Inomata, N. Iwaya, Y. Sasaki, M. Shirakawa and H. Hiroaki "LBT/PTD dual tagged vector for purification, cellular protein delivery and visualization in living cells" *Biochim.Biophys. Acta* , vol. 1773(2), pp. 141-6, 2007.

[146] S. Stefanini, E. Chiancone, E. Antonini and A. Finazzi-Agro, "Binding of Terbium to apoferritin: a fluorescence study" *Arch. Biochem. Biophys,* vol. 222(2), p. 430±434, 1983.

[147] A. Treffry and P. Harrison, "Spectroscopic studies on the binding of iron, terbium, and zinc by apoferritin" *J. Inorg. Biochem.,* vol. 21(1), pp. 9-20, 1984.

[148] L. Tian, Z. Dai, X. Liu, B. Song, Z. Ye and J. Yuan, "Ratiometric Time-Gated Luminescence Probe for Nitric Oxide Based on an Apoferritin-Assembled Lanthanide Complex-Rhodamine Luminescence Resonance Energy Transfer" *System Anal. Chem.,* vol. 87(21) 1, pp. 10878-10885, 2015.

[149] F. Bou-Abdallah, P. Arosio, S. Levi, C. Janus-Chandler and N. Chasteen, "Defining metal ion inhibitor interactions with recombinant human H- and L-chain ferritins and site-directed variants: an isothermal titration calorimetry study" *J. Biol. Inorg. Chem.,* vol. 8(4), pp. 489-497, 2003.

[150] S. Rawson, M. G. Iadanza, N. A. Ranson and S. P. Muench, "Methods to account for movement and flexibility in cryo-EM data processing" *Methods,* vol. 100, pp. 35-41, 2016.

Attachments

During the three years of my PhD (November 2015- November 2018), the following works were published:

- *January 2017*: de Turris V., Trabuco M., Peruzzi G., Boffi A., **Testi C.**, Vallone B., Montemiglio L.C., Des Georges A., Calisti L., Benni I., Bonamore A., and Baiocco P. *Humanized archaeal ferritin as a tool for cell targeted delivery.* Nanoscale (2017) 9, 647-655. (**Attachment I**)

- March *2017*: Scaglione A., Montemiglio L.C., Parisi G., Asteriti I. A., Bruni R., Cerutti G., **Testi C.**, Savino C., Mania F., Lavia P. and Vallone B. *Subcellular localization of the five members of the human steroid 5alfa-reductase family.* Biochimie Open 4, 99-106. (**Attachment II**)

- *August 2018*: Calisti L. & Trabuco M., Boffi A., **Testi C.**, Montemiglio L.C., des Georges A., Benni I., Ilari A., Taciak B., Białasek M., Rygiel T., Król M., Baiocco P. and Bonamore A. *Engineered Ferritin for lanthanide binding.* PloS ONE (2018), 13(8). (**Attachment III**)

- *September 2018*: Montemiglio L.C. **& Testi C.**, Ceci P., Falvo E., Pitea M., Savino C., Arcovito A., Peruzzi G., Baiocco P., Mancia F., Boffi A., des Georges A. and Vallone B. *Cryo-EM structure of the human Ferritin-Transferrin Receptor 1 complex.* Under review by Nature Communications. (**Attachment IV**)

Attachment I

Nanoscale

PAPER

Cite this: *Nanoscale*, 2017, **9**, 647

Humanized archaeal ferritin as a tool for cell targeted delivery†

Valeria de Turris,[a] Matilde Cardoso Trabuco,[b] Giovanna Peruzzi,[a] Alberto Boffi,[a,c] Claudia Testi,[a] Beatrice Vallone,[c] Linda Celeste Montemiglio,[d] Amédée Des Georges,[e] Lorenzo Calisti,[d] Irene Benni,[d] Alessandra Bonamore[d] and Paola Baiocco*[a]

Received 8th September 2016,
Accepted 23rd November 2016

DOI: 10.1039/c6nr07129e

www.rsc.org/nanoscale

Human ferritins have been extensively studied to be used as nanocarriers for diverse applications and could represent a convenient alternative for targeted delivery of anticancer drugs and imaging agents. However, the most relevant limitation to their applications is the need for highly acidic experimental conditions during the initial steps of particle/cargo assembly, a process that could affect both drug stability and the complete reassembly of the ferritin cage. To overcome this issue the unique assembly of *Archaeoglobus fulgidus* ferritin was genetically engineered by changing a surface exposed loop of 12 amino acids connecting B and C helices to mimic the sequence of the analogous human H-chain ferritin loop. This new chimeric protein was shown to maintain the unique, cation linked, association–dissociation properties of *Archaeoglobus fulgidus* ferritin occurring at neutral pH values, while exhibiting the typical human H-homopolymer recognition by the transferrin receptor TfR1. The chimeric protein was confirmed to be actively and specifically internalized by HeLa cells, thus representing a unique nanotechnological tool for cell-targeted delivery of possible payloads for diagnostic or therapeutic purposes. Moreover, it was demonstrated that the 12 amino acids' loop is necessary and sufficient for binding to the transferrin receptor. The three-dimensional structure of the humanized *Archaeoglobus* ferritin has been obtained both as crystals by X-ray diffraction and in solution by cryo-EM.

Introduction

Ferritin proteins have been extensively used as nanocarriers for diverse applications due to their hollow cage-like structures and their unique, reversible, 24-mer assembly.[1,2] In more detail, they represent the most convenient alternative to viral carriers for targeted delivery of anticancer drugs and imaging agents and have been successfully utilized as reaction nanovessels for the synthesis of non-native metallic nanoparticles in the inner core, with applications in nanoelectronic devices.[3–5] The external and internal surfaces of ferritin are chemically and genetically modifiable allowing for the attachment site for drugs, nucleic acids, fluorophores or magnetic moieties. Along this line, recent studies further established heavy (H) or light (L)-chain homopolymers as versatile multifunctional nanocarriers for targeted cancer diagnosis and therapy.[6–10] In fact, human ferritins constitute biocompatible nanocarriers that stabilize and shelter the enclosed particles, thus preventing immunogenic responses. Moreover, ferritins are naturally targeted toward ubiquitously expressed TfR1 transferrin receptors (H-chain specific) or hepatic SCARA 5 receptors (L-chain specific). Such properties have been widely exploited for the efficient delivery of antitumor drugs to iron-avid, fast replicating, tumor cells overexpressing the TfR1 receptor.[11] Thus, ferritin based protein cages have been developed as versatile platforms for multiple applications in nanomedicine.

Current development of human ferritin based particle is however facing intrinsic limitations due to the experimental conditions linked to their assembly–disassembly equilibrium, whose control is a prerequisite in order to achieve encapsulation of the cargo within the internal cavity. In the conventional *in vitro* encapsulation procedures, ferritin must be disassembled at extreme pH values (pH = 2.0) and re-assembled in the presence of highly concentrated payload compounds.[12]

[a]*Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, V.le Regina Elena 291, Rome 00161, Italy. E-mail: paola.baiocco@iit.it*
[b]*Molirom srl, via Ravenna 8, 00161 Rome, Italy*
[c]*Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro, 7, 00185, Rome, Italy*
[d]*Department of Biochemical Sciences "Alessandro Rossi Fanelli", Sapienza University of Rome, P.le A. Moro, 5, 00185, Rome, Italy*
[e]*The City University of New York Advanced Science Research Center 85 St. Nicholas Terrace, New York, NY 10031, USA*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6nr07129e

Such a procedure leads to a suboptimal load of cargo material, whose chemical structure is required to be pH resistant. Moreover, the pH jump procedure is only partially reversible and re-assembly may not be complete, depending on complex equilibrium and kinetic parameters.[13] Currently, extensive research efforts are devoted to adjust the assembly properties of ferritin nanocages to the desired applications either by intersubunit interface mutagenesis[14] or by genetic engineering of N- or C-terminal regions.[15]

Recently, novel ferritins from lower eukaryotes, bacteria and archaea, endowed with different polymer association–dissociation thermodynamic and kinetic features have emerged as possible alternatives to human ferritin homopolymers for several biotechnological applications requiring cargo material encapsulation.[16,17] Despite low sequence similarity, these ferritins display a highly conserved quaternary structure consisting of a four-helix bundle namely A, B, C and D and a short E helix at the C-terminus. Helices B and C are connected by a 12 amino acid (aa) long loop involved in stabilizing interactions at the 2-fold dimer interface. Unlike most eukaryotic and prokaryotic ferritins, the archaeal ferritin from *Archaeoglobus fulgidus* (AfFt) is characterized by unique self-assembly properties. In fact, in neutral buffers it is present as dimeric species, easily combining into a non-canonical 24-mer cage in the presence of metal cations.[18] AfFt assembles in a distinctive tetrahedral geometry as a result of particular packing between four hexametric units into a 24-mer structure different from those observed so far. Such unusual assembly results in the formation of four wide triangular pores (45 Å) on the protein shell.[19] As demonstrated by Sana *et al.*,[19] amino acid substitutions in the turn motif that connects D and E helices forming the 4-fold iron channels, namely K150A and R151A, are sufficient to restore the canonical octahedral symmetry observed in vertebrate and bacterial ferritins,[18,19] thus providing a rationale for the atypical tetrahedral architecture of AfFt. Nevertheless, the low sequence similarity of AfFt with mammalian ferritins runs against the possibility of targeting AfFt onto the TfR1 receptors in mammalian cells.

Transferrin receptor TfR1, or CD71, has been reported to be a preferred target for human ferritin, due to the specific interaction of the receptor extracellular moiety with epitopes of the H-ferritin subunit.[20,21] Inspection of the three-dimensional structure of the human H-homopolymer external surface indicates that the most significant accessible area is occupied by the external 12 aa long loop connecting B and C helices (up to 19 aa including the turn regions). Patches of a lesser extent exposed to solvent are formed by the N-terminal regions or by the iron channels within the threefold axis of the intersubunit assembly. However, it has been shown that N-terminal or C-terminal deleted human H-homopolymers are efficiently taken up by target cells, thus suggesting these regions are not required for receptor recognition or uptake.[22] Therefore the BC loop, besides its structural role in stabilization of the interdimer interface,[23] appeared to be the best candidate for TfR1 receptor recognition of the ferritin molecule.

We decided then to engineer the archaeal ferritin to devise a chimeric construct, named humanized *Archaeoglobus* ferritin (HumAfFt), in which the external 12 aa loop connecting the B and C helices was mutated to reproduce an analogous one in the corresponding region of the human H chain homopolymer (HuHF). It was observed that this chimeric protein (HumAfFt) was actively internalized by HeLa cells to an extent comparable to transferrin, the preferred ligand for TfR1.

Results and discussion

"Humanized" *Archaeoglobus* ferritin design

Heavy chain human ferritin and *Archaeoglobus* ferritin display 31% sequence identity. The structural alignment of HuHF and AfFt monomers (pdb code 3AJO and 1S3Q, respectively) reveals a shorter N-terminus and shorter loops between three of the four helices, with the notable exception of the long loop connecting B and C helices.

As shown in Fig. 1, the BC loops of two adjacent subunits run in an antiparallel fashion establishing significant, mutual interactions. Thus the BC loops of adjacent subunits display the same overall geometry in both HuHF and AfFt, though each couple of loops adopts a different symmetry along the spherical surface of the protein cage, given the different dimer–dimer positioning within each complete 24-mer structure.

In order to preserve the unique assembly properties of AfFt, while implementing a potential cellular uptake, we decided to exploit this external loop by mutating 9 residues into the corresponding residues found in HuHF, according to the sequence alignment shown in Fig. 2. In particular, the nine amino acids sequence IFLQDIKKP, typical of a human H ferritin homopolymer, located at the center of the 12 aa loop was inserted in place of VKLYAVEEP (from residue 70 to 79 in AfFt numbering, see Fig. 2). In addition, a cysteine residue in position 54 (AfFt numbering) has been introduced by point mutation in order to provide a conjugation site of potential thiol reactive derivatives into the cavity. As previously mentioned, we will refer to the chimeric protein described above as humanized *Archaeoglobus* ferritin (HumAfFt).

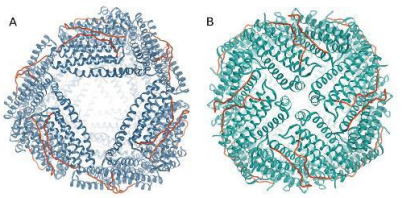


Fig. 1 Three-dimensional structures determined by X-ray crystallography. A cartoon representation of (A) HumAfFt and (B) HuHF (pdb code 3AJO). Models are coloured in blue and green, respectively, and the external loop connecting the helices B and C of each monomer is shown as red ribbons. Molecular graphics were performed using a UCSF Chimera package.[35]

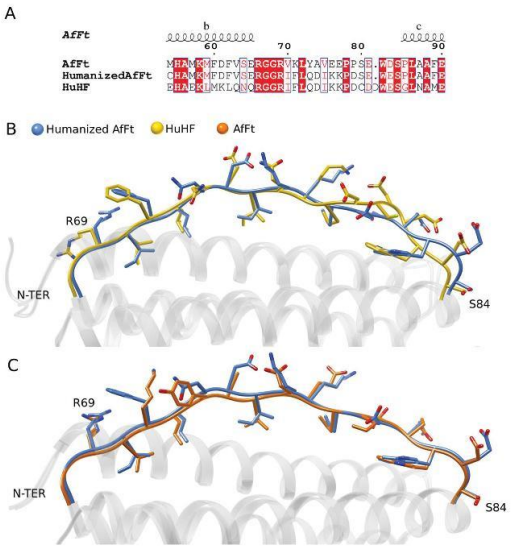


Fig. 2 Structure-based sequence alignment of AfFt, HumAfFt, and HuHF. (A) Close-up view of the sequence alignment. Elements of secondary structure for the AfFt are shown on the top. White characters in a red background indicate strict conservation while residues with poor conservation are drawn in black on a white background. Alignments were made using CLUSTAL Omega, and the figure was generated using ENDSCRIPT.[37] The structural superposition of the region from R69 and S84 (AfFt numbering) is shown in (B) HumAfFt (in blue) *vs.* HuHF (in yellow) and in (C) HumAfFt (in blue) *vs.* AfFt (in orange). Residues are depicted as sticks indicating N atoms in blue, O atoms in red and S atoms in yellow.

Self-assembly assessment

The effect of mutations on the $MgCl_2$-mediated self-assembly of HumAfFt was studied by size exclusion chromatography (SEC), in order to separate different possible oligomers according to their molecular size. Identical peak-positions confirmed that HumAfFt retains the $MgCl_2$-mediated self-assembly properties of native AfFt[24] (see Fig. S2†). As shown in the chromatograms, the increasing of $MgCl_2$ concentration triggered the self-assembly of dimers until they reached a stable polymeric structure around 500 kDa, roughly corresponding to the expected 24-mer cage-like structure, at 20 mM $MgCl_2$. The data highlighted that the chimeric HumAfFt maintained the cation induced association/dissociation properties of archaeal ferritin and is possibly assembled into a 24-mer typical structure.[24]

X-ray diffraction data on HumAfFt show the tetrahedral symmetry of archaeal ferritins

The humanized AfFt crystallized under different conditions with respect to wild type AfFt. Crystals were exposed to the synchrotron light in order to confirm the tetrahedral symmetry reported for archaeal ferritins. The protein was crystallized in the presence of Mg^{2+} in order to maintain the 24-mer assembly.

The structure of HumAfFt has been determined by X-ray crystallography at a 2.87 Å resolution. It crystallized in the $C222_1$ space group, as also observed for wild type AfFt. The asymmetric unit (ASU) contains 12 identical subunits with a solvent content of 64.3%. The overall folding corresponded to the wild type structure (pdb code 1S3Q) with a rmsd value of 0.4 Å and displayed four wide triangular pores on the surface (Fig. 1A). B factor analysis shows a mean B factor of 66 $Å^2$ with the exception of the loop region between the D and E helices, which displays higher B factors and a poor electron density on the side chains from 146 to 151 residues. Conversely, the BC loop is well organized and the analysis of the difference electron density (Fo–Fc) map clearly showed the presence of the mutated residues in the loop between the helices B and C as well as the M54C mutation, pointing towards the inner cavity.

Two magnesium ions have been positioned and successfully refined in the ASU, and they are located in the ferroxidasic site of two different subunits. In both, one magnesium

ion is coordinated with both OE1 and OE2 of Glu19, (at 2.5 Å and 2.8 Å distance, respectively), with Glu52 (OE1) at 2.5 Å, with Gln129 (OE1) at 2.6 Å and with a water molecule at 2.6 Å distance. In the other chains, a water molecule has been modelled in the Fo–Fc map and successfully refined in the same position.

The loop region, including the conserved terminal turns, spans from amino acid 68 to amino acid 86. The sequence alignment and the structural superposition between HumAfFt and the human H-homopolymer, as well as between HumAfFt and AfFt, are shown in Fig. 2B and C. At the dimeric interface between the two antiparallel BC loops, the main interactions are a hydrogen bond between Arg69(NH2) and Ser80 (O) at 2.8 Å distance and two salt bridges, namely Lys71 (NZ)–Glu77 (OE1) at 3.0 Å distance and Glu81 (OE2)–Arg69 (NH1) at 2.7 Å distance. A weak salt bridge is established by Glu81 (OE2) and Arg69 (NH2), at a distance of 3.0 Å. The salt bridge between Lys71 and Glu77 observed in HuHF, is absent in HumAfFt since these positions were mutated into a phenylalanine (Phe71) and a lysine (Lys77). Other interactions are conserved between HuHF and HumAfFt and measured the same distances.

Cryo-electron microscopy confirms the canonical AfFt architecture in solution

Humanized AfFt samples were prepared in thin ice layer and analyzed by cryo-EM in order to assess the three-dimensional structure of the chimeric protein in a near-native environment. Particles were picked with a reference-based automated particle picking procedure. Single ferritin particles were visualized at a nominal resolution of 33 Å demonstrating that their shape and dimensions correspond to that of wild type AfFt in its 24-mer assembly (Fig. 3A). Moreover, the triangular apertures on the protein surface were clearly observable in a bi-dimensional view (Fig. 3B) and better displayed in a three-dimensional reconstruction as shown in Fig. 3C.

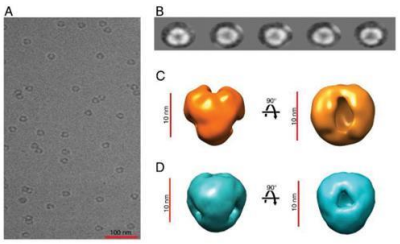


Fig. 3 Electron microscopy characterization of humanized *Archaeoglobus fulgidus* ferritin. (A) Sample micrograph of the HumAfFt data set. Scale bar: 100 nm. (B) Five representative 2D class averages obtained with RELION. (C) 3D reconstruction of HumAfFt obtained with RELION and visualized with UCSF Chimera.[35] Map final resolution: 33.1 Å. Scale bars: 10 nm. Left: side view. Right: top view. (D) *Archaeoglobus fulgidus* ferritin crystal structure (from 16) filtered to 30 Å, shown for comparison. Scale bars: 10 nm. Left: side view. Right: top view.

Flow cytometry analysis shows HumAfFt cellular uptake

After we have demonstrated that our HumAfFt maintained its structure with large open pores and the self-assembly property characteristic of the original AfFt, we aimed to verify that the modified version also gained access to eukaryotic cancer cells such as HeLa cells. It is known that HuHF is recognized and internalized by the TfR1, which is overexpressed in many types of tumor cells but not in normal cells and healthy tissues.[21] To validate the effect of our mutations on the external loop related to the uptake efficiency by HeLa cells, we performed time course experiments on cells treated with the same amount (30 μg ml^{-1}) of AfFt-FITC, HumAfFt-FITC and transferrin-FITC (TF-FITC) and analyzed them by flow cytometry. As a baseline for FITC fluorescence, control cells not incubated with FITC-ferritins were used. Moreover, to exclude any signal generated from outside particles sticking on the cell membrane due to unspecific binding or remains from the washing steps, trypan blue quenching was performed before using FACS. In Fig. 4, the FACS analysis is summarized, shown as the percentage of cells internalizing the nanoparticles at different times. These data highlighted that HumAfFt nanoparticles are efficiently taken up by HeLa cells already after one hour incubation with a much higher percentage compared to AfFt (81% and 5% respectively). After a

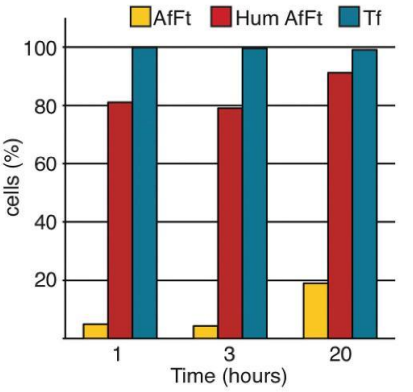


Fig. 4 Humanized AfFt is internalized with higher efficiency than the original ferritin. Ferritins taken up in HeLa cells have been quantified by flow cytometry. Cells have been treated with 30 μg ml^{-1} of AfFt-FITC, HumAfFt-FITC and transferrin-FITC (Tf). The percentage of cells internalizing the nanoparticles at the time indicated is shown. For each sample 30 000 events gated on live cells have been acquired.

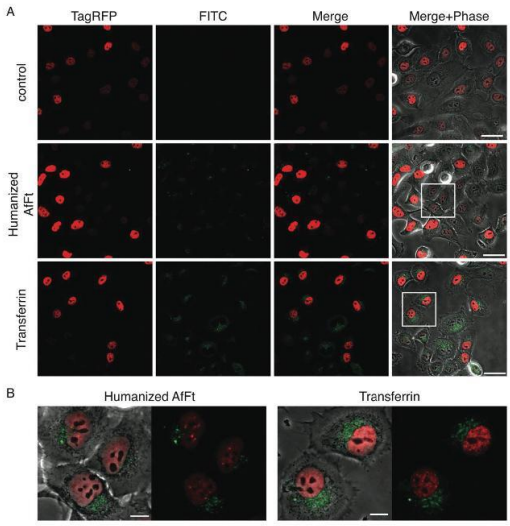


Fig. 5 Ferritin internalization observed at the confocal microscope. Cells were plated on an 8-well µ-slide (ibidi), induced with doxycycline to express TagRFP and then incubated with 30 µg ml^{-1} of humanized AfFt-FITC or transferrin-FITC for 20 h. After washing, cells were live-imaged using a confocal microscope. (A) Confocal images of live HeLa TagRFP cells are shown as single, merged channels and overlay images with the phase contrast. Scale bars: 40 µm. (B) Images are a magnified view of the region highlighted by the white inset in panel A; merged channels and the overlay with the phase contrast images are shown. Scale bars: 10 µm.

longer incubation time (20 hours), the FITC-positive cells for the humanized samples are increased to more than 90% whereas for native AfFt they are still less than 20%. The latter increment is possibly due to the unspecific uptake by pinocytosis. Each acquired plot is shown in Fig. S4.†

Moreover we performed the RNAi experiment against the transferrin receptor and analyzed the cellular uptake by FACS (Fig. S5†) obtaining around 50% HumAfFt uptake reduction after 48 hours of siRNA transfection, confirming the involvement of TfR1 in the internalization process.

Visualization of ferritin nanoparticles inside living cells by confocal microscopy

In order to visualize uptaken HumAfFt–FITC nanocages, we performed the internalization assay directly on an ibidi 8-well µ-slide and incubated the cells for 20 hours before confocal microscopy. We exploited the properties of the HeLa TagRFP cell lines available in the lab in order to have a reference fluorescence signal confirming that we were imaging inside the cell. This cell line contains a TagRFP-FUS protein under the control of a doxycycline-inducible promoter that allows for a controlled expression of the protein and hence permits visualization of the nucleus in the red channel. In this case the cells were not selected to eliminate the untransfected cells, allowing us to visualize different levels of expression and also unstained nuclei. Just before imaging, the cells were washed to eliminate the unbound FITC-nanoparticles and then acquired by confocal laser-scanning microscopy. Confocal representative images of the entire field of view of live HeLa TagRFP cells alone (control) or incubated with HumAfFt-FITC or TF-FITC are shown in Fig. 5A. A detailed view of the boxed region in panel A is shown in Fig. 5B. Images confirmed the high extent of HumAfFt internalization and highlight a cellular distribution in the cytoplasm and in the perinuclear space comparable to that observed in the case of transferrin, thus suggesting a typical clathrin-coated endocytosis pathway, mediated by TfR1.

Conclusion

The present data highlight the remarkable properties of a novel chimeric ferritin nanocage suitable for the design of efficient and versatile scaffolds for the intracellular delivery of bioactive small molecules and/or diagnostic probes.

Archaeoglobus fulgidus ferritin (AfFt) is a prominent example of this versatility due to its unique association/dissociation properties that lead to the presence of stable dimeric species at neutral pH and low ionic strength capable of associating into non-classical 24-mer species in the presence of either monovalent or divalent cations at physiological concentrations (*i.e.* higher than 0.5 M Na^{+} or 20 mM Mg^{++}). Moreover, AfFt also displays a unique subunit assembly, based on tetrahedral symmetry, which leads to the formation of four large openings in the protein shell. As such, AfFt represents a uniquely suitable scaffold for incorporating a wealth of diverse substructures inside the protein cavity, either by an assembly/disassembly process at neutral pH or by diffusion through the large triangular pores on the surface. Notable examples have been reported in recent literature.[17] Nevertheless, one of the key properties of ferritin nanocages in biomedical applications is the possibility of targeting receptors on human cells, thus allowing the delivery of the desired payload within the cytoplasm. The engineered HumAfFt described here combines the versatility in assembly and cargo incorporation of AfFt by binding to TfR1 and by cellular uptake of HuHF. At present, further experiments with possible payloads for diagnostic or therapeutic applications are under development. However, *in vivo* applications will have to wait for the evaluation of possible immunological responses against non-human epitopes present on the protein surface. In fact, though the mutated loops appear to account for most of the solvent exposed surface, N-terminal and pore lining regions might still offer windows for non-self-recognition by the immune system.

Materials and methods

"Humanized" *Archaeoglobus* ferritin design

The gene encoding for a mutated ferritin from *Archaeoglobus fulgidus* was synthesised by GeneArt (ThermoFisher) and subcloned into a pET22b vector (Novagen) between the restriction sites NdeI and HindIII at 5′ and 3′ respectively. The recombinant plasmid was transformed into BL21(DE3) *E. coli* strain for protein expression.

Protein expression and purification

E. coli cells, containing the HumAfFt plasmid, were grown and induced with 1 mM IPTG (isopropyl-β-D-1-thiogalactopyranoside) at OD_{600} = 0.6. Cells were harvested by centrifugation 3 hours post induction at 37 °C.

Cells harvested from 1 L culture were resuspended in 20 mM HEPES buffer, pH 7.5, containing 200 mM NaCl, 1 mM TCEP (tris(2-carboxiethyl)phosphine), and a cOmplete™ Mini Protease Inhibitor Cocktail Tablet (Roche). Cells were disrupted by sonication and the soluble fraction was purified by heat treatment at 78 °C for 10 minutes. Denatured *E. coli* proteins were removed by centrifugation at 15 000 rpm at 4 °C for 1 hour. The soluble protein was further purified by ammonium sulfate precipitation. The precipitated fraction at 70% ammonium sulfate was resuspended in 20 mM HEPES, 50 mM $MgCl_2$, pH 7.5 and dialyzed *versus* the same buffer. As a final purification step, the protein was loaded onto a HiLoad 26/600 Superdex 200 pg column previously equilibrated in the same buffer using an ÄKTA-Prime system (GE Healthcare). The purified protein was concentrated to obtain the final protein preparation of 1 mg mL^{-1} and protein concentration was calculated by measuring the UV spectrum using an extinction coefficient of 32 400 M^{-1} cm^{-1}. Protein yield was ~40 mg L^{-1} culture.

Self-assembly assessment in solution

Size exclusion chromatography $MgCl_2$-mediated self-assembly was studied by size exclusion chromatography (SEC) using a Superdex 200 26/600 GL column (GE Healthcare). The molecular size of HumAfFt was determined under different conditions by comparing the elution volume with that of standard proteins. Composition of the mobile phase was 25 mM HEPES pH = 7.5 with different $MgCl_2$ concentrations accordingly to the composition of the protein buffer.

Crystallization and crystal structure determination

The purified protein was concentrated to 20 mg mL^{-1} and initial crystallization screening was performed using a Phenix Robot. Crystals were obtained by mixing in a 2 μL hanging drop 1 mM of the purified protein with a solution containing 22% (vol/vol) polyacrylic acid PAA, 0.1 M Tris, 0.02 M $MgCl_2$, pH 7.4, at 25 °C within a week, cryo-protected by increasing the precipitant concentration and flash-frozen in liquid nitrogen. Diffraction data have been collected at ID23-2 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France.

Data were processed with XDS[25] and scaled with Aimless (ccp4 suite) at a final resolution of 2.87 Å. The structure was solved by Molecular Replacement with MolRep (ccp4 suite) using the open pore structure AfFt (pdb code 1S3Q) as the search model. Model Building and refinement were done using Coot[26] and Refmac5, respectively. The final model was analyzed with PROCHECK[27] and Molprobity.[28] The Ramachandran plot showed that 97.8% of residues are in preferred regions, 2.2% in allowed regions and no outlier is observed. The final atomic coordinates and structure factors were deposited with the PDB Data Bank (http://www.rcsb.org) with accession code: 5LS9. Complete data collection and refinement statistics are reported in Table S1.†

Cryo-Electron microscopy

Holey-gold grids were prepared as described by Russo and Passmore[29] from Quantifoil R1.2/1.3 (Quantifoil Micro Tools GmbH, Germany). 3 μL of HumAfFt (12 μM) was applied to the holey-gold grids after plasma cleaning with a mixture of H_2 and O_2. Grids were blotted for 4 seconds and vitrified by rapidly plunging into liquid ethane at −180 °C (ref. 30 and 31) with a Vitrobot (FEI).

Data acquisition was done using a FEI Titan Halo (FEI, Eindhoven) operating at 300 kV. Datasets were imaged with a Volta phase-plate (FEI, Eindhoven)[32] and were collected with

the automated data collection system EPU (FEI, Eindhoven) at a nominal magnification of 59 000× on an FEI Ceta camera (FEI, Eindhoven) with a camera pixel size of 14 μm, corresponding to a calibrated pixel size of 1.49 Å on the specimen scale and with a dose of 50 e^{-} $Å^{-2}$.

Image processing. The particles were picked with the reference-based automated particle picking procedure implemented in RELION 1.3.[33,34] CTF correction was not applied since the data were collected within 200 nm of focus and the first CTF zero crossing was well beyond the achievable resolution of the dataset. Those particles were subjected to 2D classification using RELION with k = 100 classes. Good particles were then subjected to 3D classification using RELION with the number of classes K = 8. Resulting classes were refined with the autorefine procedure in RELION.

Resolution estimation. Reported resolutions are based on the 'gold-standard' protocol with the FSC = 0.143 criterion using soft masks with an 8 pixel soft edge, and were corrected for the effects of the mask on the FSC curve using high-resolution noise substitution.[34]

Maps were visualized using UCSF Chimera.[35]

Cell line generation

The HeLa cell line stably expressing an inducible TagRFP-FUS protein (HeLaTagRFP) was generated by transfection with epB-Puro-TT-RFP-FUS wt plasmid and the piggyBac transposase vector. Plasmid construction is described in ref. 36.

Protein FITC labeling

HumAfFt, AfFt and Olo-transferrin were labeled with fluorescein-isothiocyanide (FITC, ThermoFisher) according to the manufacturer's standard protocol. Briefly, 2 mg mL^{-1} of the purified protein was added with 10-fold molar excess of in protein storage buffer stirring for 2 hours at RT. The non-reacted dye was removed by gel filtration chromatography and the fluorescent dye to protein ratio was determined by UV-spectroscopy. LC-MS spectrometry measurements on HumAfFt-FITC confirmed that >60% of monomers are FITC labeled as reported in Fig. S6.†

Cell cultures and ferritin internalization

HeLa cells were grown at 37 °C in Eagle's MEM supplemented with 10% (v/v) FBS, Glutamax (Invitrogen) and penicillin–streptomycin solution (Sigma). When needed, the cells were induced with doxycycline 0.2 μg mL^{-1}. The internalization assay was performed as follows: after seeding the cells on the relevant substrate depending on the experiment, cells were left one day to attach and then incubated with FITC-ferritin nanoparticles (AfFt-FITC, HumAfFt-FITC or Tf-FITC as specified in each experiment) at the final concentration of 30 μg ml^{-1} for the time indicated (1 h, 3 h or 20 h).

Flow cytometry analysis

For flow cytometry analysis HeLa cells were seeded on multi-well plates. Cells were incubated with FITC-ferritin nanoparticles as described previously, then washed two times with PBS, detached with trypsin-EDTA (Euroclone), washed with PBS and resuspended in BD-FACS flow buffer. Half of each sample was treated with Trypan Blue (TB; Sigma) to quench the FITC signal from membrane-bound nanoparticles that were not internalized. The quenching was performed with 0.04% TB for 10 min on ice. Control cells were treated in the same way but without FITC-ferritin incubation. Internalization of ferritins before and after TB treatments was measured at the BD LSFORTESSA (BD Biosciences, San Jose, CA, USA) equipped with a 488 nm laser and FACSDiva software (BD Biosciences version 6.1.3). Live cells were first gated by forward and side scatter area (FSC-A and SSC-A) plots, then detected in the green channel for FITC expression (530/30 nm filter) and side scatter parameter. The gate for the final detection was set in the control sample. Data were analyzed using FlowJo9.3.4 software (Tree Star, Ashland, OR, USA).

Confocal microscopy of live cells

To visualize ferritin internalization by live cells using a confocal microscope, HeLa TagRFP cells were seeded on a μ-slide 8-well ibiTreat (ibidi) and induced with 0.2 μg mL^{-1} of doxycycline. Cells were then incubated with FITC-ferritin nanoparticles as previously described for 20 h and, before microscopy, cells were washed two times with an imaging medium (DMEM without phenol red, 10% FBS, 10 mM Hepes, Glutamax and penicillin–streptomycin solution) to eliminate the unbound FITC-nanoparticles. The confocal laser-scanning microscope used was an Olympus FV10i platform equipped with a built-in incubator. Images were acquired with a 60×/1.2NA water-immersion objective, LD lasers, 473 nm and 559 nm, and filter sets for FITC and TRITC. Phase-contrast images were acquired simultaneously.

Acknowledgements

We thank Alessandro Rosa and Riccardo De Santis for the HeLa TagRFP cell line. EU H_{2020} Project "X-Probe", Grant No. 637295, to A. B. and M. C. T. is gratefully acknowledged. Flagship Project "Nanomax" to A. B. is acknowledged. Regional project FILAS to B. V. is also acknowledged.

References

1 E. C. Theil, Ferritin protein nanocages use ion channels, catalytic sites, and nucleation channels to manage iron/oxygen chemistry, *Curr. Opin. Chem. Biol.*, 2011, **15**, 304–311.

2 L. Schoonen and J. C. van Hest, Functionalization of protein-based nanocages for drug delivery applications, *Nanoscale*, 2014, **6**, 7124–7141.

3 A. Prastaro, P. Ceci, E. Chiancone, A. Boffi, G. Fabrizi and S. Cacchi, Homocoupling of arylboronic acids and potassium aryltrifluoroborates catalyzed by protein-stabilized

palladium nanoparticles under air in water, *Tetrahedron Lett.*, 2010, **51**, 2550.
4 F. Meng, B. Sana, Y. Li, Y. Liu, S. Lim and X. Chen, Bioengineered tunable memristor based on protein nanocage, *Small*, 2014, **10**, 277–283.
5 M. A. Kostiainen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen and P. Ceci, Electrostatic assembly of binary nanoparticle superlattices using protein cages, *Nat. Nanotechnol.*, 2013, **8**, 52–56.
6 M. Liang, K. Fan, M. Zhou, D. Duan, J. Zheng, D. Yang, J. Feng and X. Yan, H-ferritin-nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14900–14905.
7 L. Vannucci, E. Falvo, C. M. Failla, M. Carbo, M. Fornara, R. Canese, S. Cecchetti, L. Rajsiglova, D. Stakheev, J. Krizan, A. Boffi, G. Carpinelli, V. Morea and P. Ceci, In Vivo Targeting of Cutaneous Melanoma Using an Melanoma Stimulating Hormone-Engineered Human Protein Cage with Fluorophore and Magnetic Resonance Imaging Tracers, *J. Biomed. Nanotechnol.*, 2015, **11**, 81–92.
8 E. Falvo, E. Tremante, R. Fraioli, C. Leonetti, C. Zamparelli, A. Boffi, V. Morea, P. Ceci and P. Giacomini, Antibody-drug conjugates: targeting melanoma with cisplatin encapsulated in protein-cage nanoparticles based on human ferritin, *Nanoscale*, 2013, **5**, 12278–12285.
9 L. Vannucci, E. Falvo, M. Fornara, P. Di Micco, O. Benada, J. Krizan, J. Svoboda, K. Hulikova-Capkova, V. Morea, A. Boffi and P. Ceci, Selective targeting of melanoma by PEG-masked protein-based multifunctional nanoparticles, *Int. J. Nanomed.*, 2012, **7**, 1489–1509.
10 S. Geninatti Crich, M. Cadenazzi, S. Lanzardo, L. Conti, R. Ruiu, D. Alberti, F. Cavallo, J. C. Cutrin and S. Aime, Targeting ferritin receptors for the selective delivery of imaging and therapeutic agents to breast cancer cells, *Nanoscale*, 2015, **7**, 6527–6533.
11 K. Fan, C. Cao, Y. Pan, D. Lu, D. Yang, J. Feng, L. Song, M. Liang and X. Yan, Magnetoferritin nanoparticles for targeting and visualizing tumour tissues, *Nat. Nanotechnol.*, 2012, **7**, 459–464.
12 Z. Zhen, W. Tang, T. Todd and J. Xie, Ferritins as nanoplatforms for imaging and drug delivery, *Expert Opin. Drug Delivery*, 2014, **11**, 1913–1922.
13 K. Mihee, R. Yecheol, S. J. Kyeong, A. Byungcheol, J. Sungmin, K. Heesoo and R. Moonhor, pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly, *Biomacromolecules*, 2011, **12**, 1629–1640.
14 H. Chen, S. Zhang, C. Xu and G. Zhao, Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions, *Chem. Commun.*, 2016, **52**, 7402–7405.
15 S. Kim, J. O. Jeon, E. Jun, J. Jee, H. K. Jung, B. H. Lee, I. S. Kim and S. Kim, Designing Peptide Bunches on Nanocage for Bispecific or Superaffinity Targeting, *Biomacromolecules*, 2016, **17**, 1150–1159.
16 A. M. Sevcenco, M. Paravidino, J. S. Vrouwenvelder, H. T. Wolterbeek, M. C. van Loosdrecht and W. R. Hagen, Phosphate and arsenate removal efficiency by thermostable ferritinenzyme from Pyrococcus furiosus using radioisotopes, *Water Res.*, 2015, **76**, 181–186.
17 B. Sana, E. Johnson and S. Lim, The unique self-assembly/disassembly property of *Archaeoglobus fulgidus* ferritin and its implications on molecular release from the protein cage, *Biochim. Biophys. Acta*, 2015, **1850**, 2544–2551.
18 E. Johnson, D. Cascio, M. R. Sawaya, M. Gingery and I. Schröder, Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilicarchaeon *Archaeoglobus fulgidus*, *Structure*, 2005, **13**, 637–648.
19 B. Sana, E. Johnson, P. Le Magueres, A. Criswell, D. Cascio and S. Lim, The role of nonconserved residues of *Archaeoglobus fulgidus* ferritin on its unique structure and biophysical properties, *J. Biol. Chem.*, 2013, **288**, 32663–32672.
20 Z. Heger, S. Skalickova, O. Zitka, V. Adam and R. Kizek, Apoferritin applications in nanomedicine, *Nanomedicine*, 2014, **9**, 2233–2245.
21 L. Li, C. J. Fang, J. C. Ryan, E. C. Niemi, J. A. Lebrón, P. J. Björkman, H. Arase, F. M. Torti, S. V. Torti, M. C. Nakamura and W. E. Seaman, Binding and uptake of H-ferritin are mediated by human transferrin receptor-1, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 3505–3510.
22 S. Levi, A. Luzzago, G. Cesareni, A. Cozzi, F. Franceschinelli, A. Albertini and P. Arosio, Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferroxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants, *J. Biol. Chem.*, 1988, **263**, 18086–18092.
23 C. Bernacchioni, V. Ghini, C. Pozzi, F. Di Pisa, E. C. Theil and P. Turano, Loop electrostatics modulates the intersubunit interactions in ferritin, *ACS Chem. Biol.*, 2014, **9**, 2517–2525.
24 L. Calisti, I. Benni, M. Cardoso Trabuco, P. Baiocco, B. Ruzicka, A. Boffi, E. Falvo, F. Malatesta and A. Bonamore, Probing bulky ligand entry in engineered archaeal ferritins, *Biochim. Biophys. Acta*, 2016, DOI: 10.1016, [Epub ahead of print].
25 W. Kabsch, Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 133–144.
26 P. Emsley and K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 2126–2132.
27 R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, PROCHECK - a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.*, 1993, **26**, 283–29125.
28 V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallo-

graphy, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 12–21.
29 C. J. Russo and L. A. Passmore, Ultrastable gold substrates for electron cryomicroscopy, *Science*, 2014, **346**, 1377–1380.
30 J. Dubochet, M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall and P. Schultz, Cryo-electron microscopy of vitrified specimens, *Q. Rev. Biophys.*, 1988, **21**, 129–228.
31 T. Wagenknecht, R. Grassucci and J. Frank, Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli*, *J. Mol. Biol.*, 1988, **199**, 137–147.
32 R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko and W. Baumeister, Volta potential phase plate for in-focus phase contrast transmission electron microscopy, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 15635–15640.
33 S. H. W. Scheres, Semi-automated selection of cryo-EM particles in RELION-1.3, *J. Struct. Biol.*, 2015, **189**, 114–122.
34 S. Chen, G. McMullan, A. R. Faruqi, G. N. Murshudov, J. M. Short, S. H. Scheres and R. Henderson, High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy, *Ultramicroscopy*, 2013, **135**, 24–35.
35 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt and E. C. Meng, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
36 M. Morlando, S. Dini Modigliani, G. Torrelli, A. Rosa, V. Di Carlo, E. Caffarelli and I. Bozzoni, FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment, *EMBO J.*, 2012, **31**, 4502.
37 X. Robert and P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res.*, 2014, **42**, 320–324.

Attachment II

Available online at www.sciencedirect.com

ScienceDirect

Biochimie Open 4 (2017) 99–106

http://www.journals.elsevier.com/biochimie-open

Research paper

Subcellular localization of the five members of the human steroid 5α-reductase family

Antonella Scaglione [a,1], Linda Celeste Montemiglio [a,f,1], Giacomo Parisi [a,1], Italia Anna Asteriti [b], Renato Bruni [c], Gabriele Cerutti [a], Claudia Testi [d], Carmelinda Savino [b], Filippo Mancia [e], Patrizia Lavia [b], Beatrice Vallone [a,f,g,*]

[a] Dept. of Biochemical Sciences, Sapienza University of Rome, P.le A.Moro 5, 00185, Rome, Italy
[b] CNR-IBPM, P.le A.Moro 5, 00185, Rome, Italy
[c] Center on Membrane Protein Production and Analysis (COMPPÀ), New York, NY, 10027, USA
[d] Center for Life Nano Science@Sapienza, IIT, V.le Regina Elena 291, Rome, I-00185, Italy
[e] Department of Physiology and Cellular Biophysics, Columbia University, New York, NY, 10032, USA
[f] Istituto Pasteur–Fondazione Cenci Bolognetti, Dept. of Biochemical Sciences, Sapienza University of Rome, Italy
[g] Italian Academy for Advanced Studies in America at Columbia University, USA

Received 1 December 2016; accepted 15 March 2017
Available online 21 March 2017

Abstract

In humans the steroid 5α-reductase (SRD5A) family comprises five integral membrane enzymes that carry out reduction of a double bond in lipidic substrates: Δ^4-3-keto steroids, polyprenol and trans-enoyl CoA. The best-characterized reaction is the conversion of testosterone into the more potent dihydrotestosterone carried out by SRD5A1-2. Some controversy exists on their possible nuclear or endoplasmic reticulum localization.

We report the cloning and transient expression in HeLa cells of the five members of the human steroid 5α-reductase family as both N- and C-terminus green fluorescent protein tagged protein constructs. Following the intrinsic fluorescence of the tag, we have determined that the subcellular localization of these enzymes is in the endoplasmic reticulum, upon expression in HeLa cells. The presence of the tag at either end of the polypeptide chain can affect protein expression and, in the case of trans enoyl-CoA reductase, it induces the formation of protein aggregates.

Keywords: Steroid 5α-reductase; Polyprenol reductase; Trans-enoyl-CoA reductase; Subcellular localization

1. Introduction

In humans the steroid 5α-reductase family (SRD5A) comprises five membrane embedded enzymes that carry out the reduction of a double bond of lipidic substrates. They can be traced back to a common ancestor that split in three subfamilies (SRD5A1-2, SRD5A3 and TECR-TECRL). Members of the three subfamilies are found in all primary eukaryotic species from plant, amoeba and yeast up to vertebrates, indicating that the split arose in early eukaryota [1]. The duplication of an ancestral gene led to substrate diversification and acquisition of a central role in vertebrates in the context of the complex physiological roles of steroids.

SRD5A1 and SRD5A2 carry out the NADPH-dependent reduction of the Δ^4 group (double bond) of C-19 and C-21 steroids into 5α-stereoisomers [2]. This reaction yields dihydrotestosterone, a more potent derivative of testosterone [3]. The first identified role of dihydrotestosterone is to trigger the development of external genitalia in the male fetus, as

* Corresponding author. Dept. of Biochemical Sciences, Sapienza University of Rome, P.le A.Moro 5, 00185, Rome, Italy.
E-mail address: beatrice.vallone@uniroma1.it (B. Vallone).
[1] These authors equally contributed to this work.

http://dx.doi.org/10.1016/j.biopen.2017.03.003

dramatically brought to evidence by the identification of familial incomplete male pseudo-hermaphroditism, a congenital condition due to mutations of the SRD5A2 gene [4]. Due to their activity on testosterone, both SRD5A1 and SRD5A2 can be viewed as pivotal factors in several physio-pathological conditions, from benign prostatic hyperplasia and prostatic cancer to acne and male pattern baldness [5,6]. Two commercial drugs, finasteride and dutasteride, are used in therapy for benign prostatic hyperplasia and, in the case of finasteride, also for male pattern baldness.

The action of SRD5A1 and SRD5A2 on progesterone, testosterone and other related steroids is also a committing step for the production of neurosteroids. These are among the strongest modulators of GABA(A) receptors with anticonvulsant, antidepressant and anxiolytic effects [7].

The expression of SRD5A1 and SRD5A2 was studied in several organs and tissues, with results depending on the methodology (antibody, mRNA analysis, *in situ*, after protein extraction, etc.) and on tissue preparation and origin (for a review see Ref. [8]). Expression of SRD5A1 is consistently reported in the brain, liver and sebaceous glands, whereas SRD5A2 is mainly found in the urogenital tract, genital skin, liver and transiently in the brain.

A third member, SRD5A3, has testosterone reductase activity, and seems to play a role in the onset of hormone refractory prostate cancer [9]. Surprisingly, inactivating mutations of the SRD5A3 gene cause a rare congenital neurodevelopmental disorder. This is possibly related to the fact that SRD5A3 is also endowed with polyprenol reductase activity, thus carrying out one of the earliest steps of protein N-linked glycosylation [10]. SRD5A3 was ubiquitously detected in human healthy tissues and is overexpressed in prostate and breast malignancies [11].

Trans-2-Enoyl-CoA reductase (TER or TECR) comes as a more recent addition to the SRD5A family. This gene, when mutated, causes a rare autosomic syndrome, nonsyndromic mental retardation, and the protein carries out the fourth step of very long-chain fatty acid synthesis and sphingosine degradation [12]. TECR expression seems ubiquitous in man and mouse organs by Northern blot analysis [13], with high expression in the nervous system [14] consistently with its involvement in the synthesis of membrane lipids and with congenital neurological impairment arising from mutations.

A fifth gene of unknown function, called TECR-like (TECRL), was identified in vertebrates through sequence homology with TECR [1].

During over 30 years of research on SRD5A enzymes, their purification in an active form has proven elusive. Nevertheless, biochemical studies on the microsomal fraction from liver and prostate led to the characterization of the main enzymatic parameters of SRD5A1, SRD5A2 [15] and, to a lesser extent, SRD5A3 proteins [8,16].

Sequence analysis and biochemical studies of the SRD5A family members did not highlight, to our knowledge, the presence of post-translational modifications, such as specific proteolytic signals, glycosylation, phosphorylation, acylation or isoprenylation. Similarly, no canonical signal for a specific subcellular localization was reported.

The subcellular localization of the SRD5A enzymes is significant to their functions since it might affect binding of reaction products to nuclear steroid receptors in the same cell or diffusion to other cells. We have undertaken a systematic analysis to address their localization. We have cloned all members of the human SRD5A family in vectors for transient expression in mammalian cells, fusing at either C- or N-terminus the enhanced green florescent protein (eGFP). We present a comprehensive assessment of their localization as a guideline for functional analysis in cells and for production and purification for structural and biochemical studies.

2. Material and methods

2.1. Cloning and plasmids

Plasmids containing the cDNA sequence for human SRD5A1, SRD5A2, SRD5A3 and TECR were purchased from the IMAGE Consortium (www.imageconsortium.org), the sequences of the human TECR and TECRL genes were synthesized.

The five genes were cloned into the pTriEx modifed vectors pNYCOMPS_RB_30_GFP and pNYCOMPS_RB_33_GFP, using the ligase independent cloning method described in Ref. [17]. These vectors are designed for mammalian, insect and bacterial protein expression and they add at the N-terminus of the protein coded by the cloned gene a cassette containing sequentially: ten histidines tag, FLAG tag, eGFP and TEV protease cleavage site at the amino terminus of the gene of interest for pNYCOMPS_RB_33_GFP. For the pNYCOMPS_RB_30_GFP the C-term is followed by: TEV protease cleavage site, eGFP, 10 histidine tag and FLAG-tag (see Fig. 1). These plasmids allow detection and expression of the cloned protein upon transient transfection of mammalian or insect cells, or expression in bacteria.

2.2. Cell culture and transfection

HEK293 and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, 2% L-glutamine, and 2% penicillin/streptomycin in humidified atmosphere at 37 °C in 5% CO_2. Confluent cell layers were trypsinized, and replated on poly-L-lysine coated coverslips in DMEM medium, which was replaced by serum-free media the next day.

Transfections were performed using Lipofectamine 2000, all controls were transfected with peGFP-N1 vector. Cells were harvested 48 h after transfection and analyzed by Western blot or immunofluorescence. The transfection rate was calculated after staining the cells on the coverslip with DAPI and counting eGFP-positive cells/DAPI-stained nuclei.

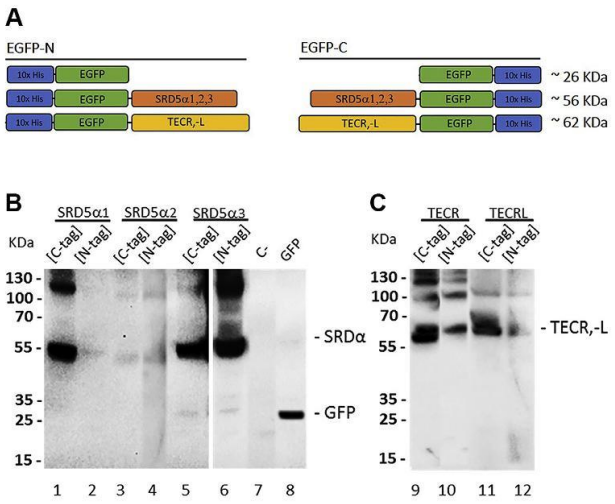


Fig. 1. **Schematic cartoon and expression of human SRD5As, TECR and TECRL eGFP-fusion proteins**. (A) Diagram of SRD5A constructs showing position of eGFP both at N- or C-term; the FLAG tag is omitted. (B) Extract from HeLa cells transfected with human SRD5A1, SRD5A2, SRD5A3 expression plasmids for C-term and N-term eGFP fusions, untransfected and eGFP-transfected cells. Different exposure times are shown (50 s for lanes 1–5, 25 s for lanes 6–8). (C) Extract from HeLa cells transfected with human TECR, TECRL (C-term and N-term eGFP fused). Proteins were visualized using anti-eGFP antibody after 10 s exposure (lanes 9–12).

2.3. Western blot

Total protein lysate was extracted with Normal Salt Buffer (20 mM Hepes, pH 7.0, 200 mM NaCl, 2 mM 2-Mercaptoethanol) supplemented with 1% β-D-Maltopyranoside and protease inhibitor cocktail (Roche). After incubation for 90 min at 4 °C, the lysates were clarified by centrifugation at 12,000×*g* at 4 °C for 20 min. Protein concentration was determined by the Bradford method. Equal amounts of proteins (40 μg) were loaded on 4–12% SDS-PAGE. Samples were electrotransferred onto nitrocellulose membranes. The membranes were blocked with 5% milk in TBS/0.1% Tween 20 (TBST) on an orbital shaker for 1 h and incubated for 1 h with rabbit monoclonal anti-GFP (1:1000, Roche). Following five TBST washes, the membranes were incubated for 1 h with horseradish peroxidase-conjugated anti-rabbit antibody in TBST. Membranes were developed by enhanced chemiluminescence, and detected proteins were identified by their molecular sizes.

2.4. Immunofluorescence and microscopy

Cells grown on coverslips were washed in PBS and fixed at room temperature by: (a) 3.7% PFA/30 mM sucrose in PBS, 10 min, (b) 5 min of membrane permeabilization in 0.1% TritonX-100 in PBS, and (c) 10 min in 0, 1 M Glycine. Blocking and incubations were performed in PBS/0.05% Tween 20/3% BSA. Incubation with primary antibodies was performed for 16 h at 4 °C. Secondary antibodies were incubated for 45 min at room temperature. Cells were counterstained with 4,6-diamidino-2-phenylindole (DAPI, 0.1 μg/ml) and mounted using Vectashield (Vector Laboratories).

Primary antibodies were: mouse anti-Erp57 (1:200, ab13506, Abcam), rabbit anti-Giantin (1:500, ab24586; Abcam), rabbit anti-pericentrin (2 μg/ml, ab4448; Abcam) and mouse anti-α-tubulin (1:2000, B-5-1-2, Sigma–Aldrich). Secondary antibodies were conjugated to Cy3, Texas red, FITC or Alexa647 (Vector Laboratories). Samples were analyzed using a Nikon Eclipse 90i microscope with a Qicam Fast 1394 CCD camera (QImaging). Image acquisition, deconvolution and Extended Depth of Focus on Z-serial optical sections were performed using Nis-Elements AR4.2 (Nikon); images were further processed with Adobe Photoshop CS 8.0.

The ProteoStat® aggresome detection kit was used to detect protein aggregates in HeLa cells transfected for SRD5A protein expression. Samples were analyzed using an Olympus FLUOVIEW FV10i-DOC Confocal Laser Scanning Microscope, equipped with a 60× phase contrast water-immersion objective (N.A.1.2) and an analog detector (PMT), using a

Texas Red filter set for the ProteoStat® dye, and an Midoriishi-Cyan filter set for the eGFP. Image acquisition and analysis were performed using an Image Processing and Analysis in Java software (ImageJ); images were further processed with Adobe Photoshop CS 8.0.

3. Results

3.1. Protein expression

Protein expression was assessed by immunoblot on detergent-solubilized whole lysates of HeLa cells transfected with the five members of the human SRD5A family. The antibody recognized eGFP fused either at the N- or at the C-term. We observed expression for all constructs, albeit with significant quantitative variations, as high as 100-fold (Fig. 1). SRD5A2, tagged at C- or N-term, and SRD5A1 when tagged at the N-term, show faint bands by immunoblot analysis. In parallel, we observed on duplicate slides that their transfection efficiency was only between 5% and 2%, thus explaining the weak signal detected in Western blots. The not transfected control shows no reaction with the anti-GFP antibody as expected. In extract from cells transfected with the vector for eGFP expression, the antibody recognizes a single polypeptide with the correct molecular weight (~27 kDa). Intensely stained bands, of apparent molecular weight corresponding to eGFP-fused SRD5A proteins (~56 kDa), are evident for cells transfected with C-term tagged eGFP-SRD5A1 (Fig. 1; lane 1), and with SRD5A3 eGFP-tagged at either end (Fig. 1; lanes 5 & 6). On the other hand, TECR and TECRL have a higher expected molecular weight (~62 kDa) and consistently their transient expression yielded slower-migrating bands on the immunoblot (Fig. 1; lanes 9–12).

In all samples, the anti-GFP antibody also recognized higher molecular weight species, which we attribute to an oligomeric state of SRD5A membrane proteins, which were not dissociated in SDS PAGE due to their very high hydrophobic character. Indeed, no high molecular weight signals were detected in control samples expressing eGFP not fused to SRD5A proteins.

3.2. Subcellular localization of eGFP tagged SRD5A proteins

The localization of each eGFP-tagged protein was examined relative to relevant markers of the membrane compartments, i.e. the ER resident protein Erp57 and the Golgi compartment resident protein Giantin [18–20]. Due to the fact that preliminary assays depicted diversified patterns for these markers, especially Giantin, in individual cells, we first evaluated the distribution of ERp57 and Giantin in co-staining with cell cycle markers, i.e. the centrosomal protein pericentrin (to identify interphase stages: one spot, G1 phase; two mature spots, late S and G2 phases) and the microtubule constituent α-tubulin (to identify mitosis). This allowed us to exclude that Erp57 and Golgi distribution patterns undergo significant remodeling during the cell cycle and avoid inaccurate evaluation of SRD5As co-localization (Fig. S1).

Fig. 2A shows that N-term and C-term eGFP-tagged SRD5A1, SRD5A2 and SRD5A3 proteins expressed upon transfection reside in the ER, as indicated by co-localization with Erp57 and the overall distribution of the green fluorescence relative to Erp57. No nuclear localization was observed. On the other hand, eGFP alone localizes in the nucleus (Fig. S1). These observations are consistent with the ER localization of SRD5A proteins being genuine and not attributable to the presence of the eGFP tag.

Also for TECR and TECRL members, both the N- and C-term eGFP-tagged proteins, with one exception, showed a pattern and co-localization behavior corresponding to ER membrane insertion (Fig. 2B). A remarkable exception was observed in cells expressing the C-term eGFP-tagged TECR, where only sparse bright dots were detected, moreover the localization of the ER marker Erp57 was less diffuse than in non-transfected cells (Fig. S1). In some cases, there was some limited overlap between Erp57 and C-term eGFP-tagged TECR proteins (individual z-stack scanning are given in Fig. S2). On the other hand, the lack of co-localization with Giantin for TECR and TECRL, in both the C- and N-term eGFP-tagged proteins (Fig. 3A), ruled out the possibility that they might localize to the Golgi.

Staining with ProteoStat® that highlights the presence of aggresomes and protein aggregates [21], showed that the green fluorescent dots observed for C-term eGFP-tagged TECR form foci that are contained within larger aggregates, providing evidence consistent with the fact that the presence of eGFP at the C-term of TECR promotes protein misfolding, at least in part, resulting in aggregation (Fig. 3B).

4. Discussion

The subcellular localization of members of the SRD5A family in humans and also in mouse is controversial, especially as far as SRD5A1 and SRD5A2 are concerned. Russell and Wilson reported indications of a differential organ-specific subcellular localization, nuclear in the prostate and cytoplasmic in the liver, regardless of the enzyme variant (SRD5A1 or SRD5A2) [15,22], or that the localization for both is in the endoplasmic reticulum [23]. Early findings based mainly on cellular component fractionation [24,25], but later also on immunohistochemical studies [26,27], indicate that the SRD5A1 enzyme localization is mainly within nuclear or perinuclear membranes whereas the SRD5A2 protein is found in the cytoplasm, most likely in the ER.

Protein expression in mammalian (human HeLa and HEK293 and simian COS-M6), as well as yeast and insect cells, again yielded contradictory results. Some authors have reported that human SRD5A1 is associated with nuclear fractions upon expression in insect Sf9 cells [28,29] but later papers [30] have instead reported retrieval of enzyme activity from the microsomal fraction. Expression of SRD5A2 and SRD5A1 in HEK293 and in LNGK-9 PCa cells yielded activity in the nuclear and microsomal fractions [31,32]. Upon

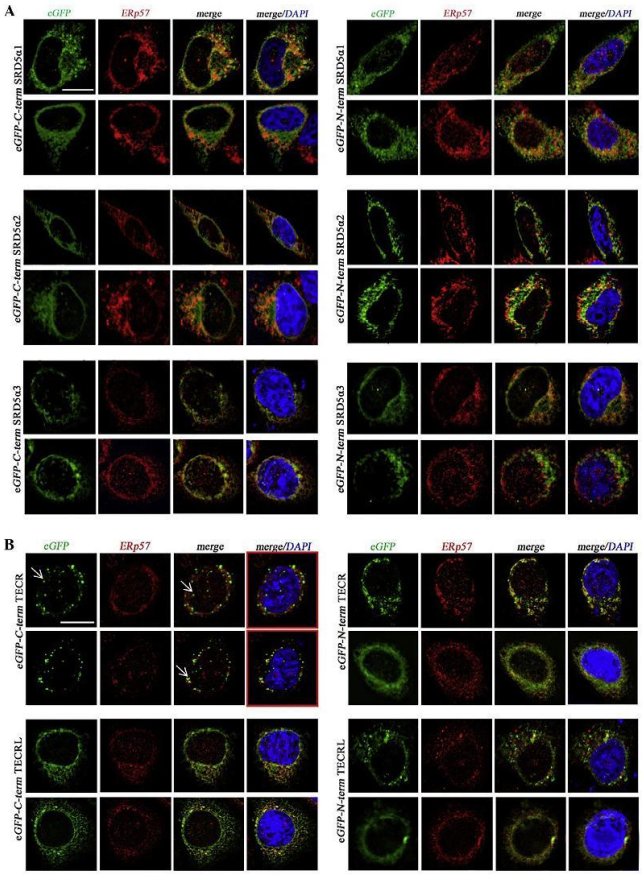


Fig. 2. **Fluorescence of human SRD5As (A) and TECR, TECRL (B) proteins in relation to endogenous ERp57.** SRD5A1, SRD5A2 and SRD5A3 proteins eGFP-tagged at C- or N-term (A), N-term eGFP-tagged TECR and N- and C-term eGFP-tagged TECRL proteins (B) reside in the ER, as indicated by co-localization with Erp57 and overall distribution of green fluorescence relative to Erp57. eGFP C-term TECR (B, left column) shows diffuse bright dots indicated by white arrows, with only limited overlap with ERp57. Deconvolutions plus Extended Depth of Focus (EDF) were performed on Z-stack series (13 frames-0, 6 mm each). The scanning of Z-stacks for red-framed images in B, with eGFP C-term TECR is given in supplementary Fig. 2. Scale bar = 10 μm, Magnification = 100×.

heterologous expression and purification in yeast, SRD5A1 appears to be associated to nuclei, whereas SRD5A2 is in the microsomal fraction [33]. Immunostaining for SRD5A2 showed presence of the enzyme in the cytoplasm throughout most regions of the brain [34]. Finally, it was reported that glioma cells express SRD5A1 in the perinuclear space or in the cytoplasm, depending on culture conditions [35].

As far as the third isoform of testosterone α-reductase (SRD5A3, also known as polyprenol reductase), available data report recovery from the microsomal fraction [8] and transient

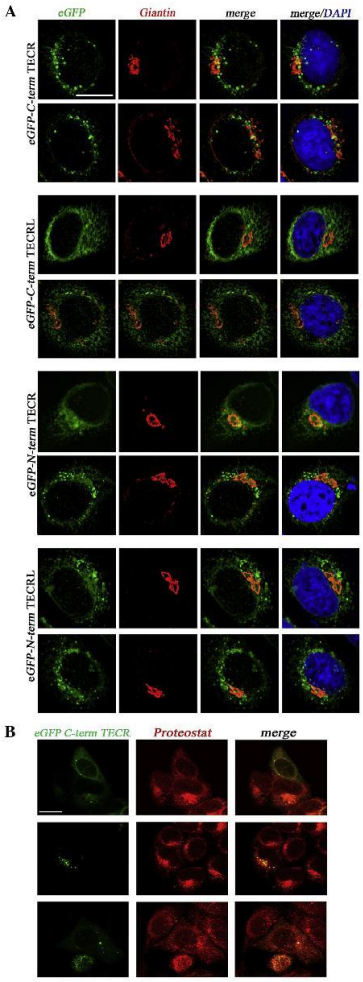


Fig. 3. **Fluorescence analysis of human TECR and TECRL relative to Giantin and fluorescence of eGFP C-term TECR and aggregates/aggresomes in the cytoplasm.** (A) TECR proteins do not localize in the Golgi apparatus, as associated eGFP fluorescence shows a broader distribution around, but not coinciding with, the Giantin signal. Scale bar = 10 μm, Magnification = 100×. (B) Localization of eGFP C-term TECR and aggresomes in HeLa cells after treatment with ProteoStat® dye for 30 min. Fluorescence of C-term eGFP-tagged TECR partially co-localizes with aggregated proteins. Scale bar = 10 μm, Magnification = 60×.

expression in HeLa cells yielded cytoplasmic localization, and no detectable activity [36]. For the TECR enzyme, immunohistochemistry has clearly suggested a localization in the ER [12]. No data is available for the TECRL that has been described only at the gene level.

SRD5A family members share significant sequence similarity, having likely evolved via duplication from an ancestral gene (sequence alignment is shown in Fig. S3). In localization studies, in order to eliminate any potential ambiguities due to antibody specificity, we directly examined each of the five human SRD5A family members by taking advantage of the fluorescent emission of the eGFP fusion tag in transfected HeLa cells. The caveat of aberrant subcellular localization due to the presence of the eGFP tag calls for careful consideration, and we therefore systematically compared the localization of C-term and N-term tagged paired constructs for each SRD5A member in parallel experiments.

The effect of a GFP tag on membrane protein localization and expression can vary, nevertheless this method is reported as a valid tool in subcellular localization studies [37–39]. The positioning on the C- or N-term of the polypeptide chain can affect protein expression and stability and its effect varies depending on the system under analysis. In an extended study on the effects on GFP tagging on subcellular localization [40] C-term tagging is reported to maintain physiological localization, whereas N-term affected it, contrarily to what we observed with human TECR.

The data we obtained with proteins fused with eGFP at the N-term indicate that all members of the SRD5A family reside in the endoplasmic reticulum, when transient expression is carried out in HeLa cells. This is also evident for protein constructs where the fusion with eGFP was at the C-term, with the exception of TECR. TECR was previously localized in the endoplasmic reticulum by immunofluorescence staining in transiently transfected HeLa cells [10]. This, and the patterns obtained here with the C-term chimaera, allowed us to conclude that the presence of eGFP at the C-term affects TECR folding, leading to the accumulation of aggregated protein.

In the SRD5A proteins, N-term eGFP-tagging reduces expression for SRD5A1, TECRL and TECR, and for the latter C-term eGFP-tagging promotes aggregation. On the other hand, for SRD5A3, expression is somewhat higher in the N-term eGFP-tagged form. Thus, even within a family of proteins, the effect of fusion tags on protein expression is not predictable and has to be assessed for each member. Our results indicate that membrane protein tagging with eGFP, which is commonly used to assess protein expression by fluorescence-coupled size-exclusion chromatography (FSEC) [41], can lead to very different expression yields, depending on the tag position, and that higher expression levels, as in the case of C-term tagged TECRL with respect to the N-term tagged form, can underlie unfolding and aberrant subcellular localization.

In summary, the present studies support the conclusion that isozymes 1 and 2 of testosterone 5α-reductase (SRD5A1 and SRD5A2) are both localized in the ER, upon transient expression in HeLa cells, our data are a valuable contribution to clarify the contradictory reports detailed above, although

any ultimate conclusion on the intracellular localization of SRD5As should be validated in other cell lines and complementary experimental settings.

Moreover, the subcellular localization in the ER for SRD5A3 is consistent with the reported microsomal subfractionation. Finally, we show that the previously uncharacterized TECRL gene can be expressed in human cells and that its product is a protein residing in the ER membrane.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Supported by Regione Lazio (*ProTox*) *Prot. FILAS-RU-2014–1020* and by Sapienza Grant C26H154L5A. Support from H2CU Center to BV and GC is acknowledged. This work was also supported by NIH grants U54 GM095315 and P41 GM116799 (Hendrickson, PI) and R01 GM111980 (Mancia, PI) and by the fellowship awarded by the Italian Academy for Advanced Studies in America at Columbia University (Vallone). We thank Dr. Valeria De Turris for access to the Microscopy Facility of CLNS-IIT@Sapienza.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biopen.2017.03.003.

References

[1] V.S. Langlois, D. Zhang, G.M. Cooke, V.L. Trudeau, Evolution of steroid-5alpha-reductases and comparison of their function with 5beta-reductase, Gen. Comp. Endocrinol. 166 (2010) 489–497, http://dx.doi.org/10.1016/j.ygcen.2009.08.004.

[2] R.I. Dorfman, E. Forchielli, Separation of delta 4-5 alpha-hydrogenases from rat liver homogenates, J. Biol. Chem. 223 (1956) 443–448, http://dx.doi.org/10.1155/2012/530121.

[3] T. Saartok, E. Dahlberg, J.A. Gustafsson, Relative binding affinity of anabolic-androgenic steroids: comparison of the binding to the androgen receptors in skeletal muscle and in prostate, as well as to sex hormone-binding globulin, Endocrinology 114 (1984) 2100–2106, http://dx.doi.org/10.1210/endo-114-6-2100.

[4] J. Imperato-McGinley, L. Guerrero, T. Gautier, R.E. Peterson, Steroid 5alpha-reductase deficiency in man: an inherited form of male pseudohermaphroditism, Science 186 (1974) 1213–1215. ISSN: 05476844.

[5] A. Cilotti, G. Danza, M. Serio, Clinical application of 5alpha-reductase inhibitors, J. Endocrinol. Investig. 24 (2001) 199–203.

[6] É. Lévesque, I. Laverdière, L. Lacombe, P. Caron, M. Rouleau, V. Turcotte, B. Têtu, Y. Fradet, C. Guillemette, Importance of 5α-reductase gene polymorphisms on circulating and intraprostatic androgens in prostate cancer, Clin. Cancer Res. 20 (2014) 576–584, http://dx.doi.org/10.1158/1078-0432.CCR-13-1100.

[7] B. Stoffel-Wagner, Neurosteroid biosynthesis in the human brain and its clinical implications, Ann. N. Y. Acad. Sci. 1007 (2003) 64–78, http://dx.doi.org/10.1196/annals.1286.007.

[8] F. Azzouni, A. Godoy, Y. Li, J. Mohler, The 5alpha-reductase isozyme family: a review of basic biology and their role in human diseases, Adv. Urol. 2012 (2012), http://dx.doi.org/10.1155/2012/530121. ID530121.

[9] M. Uemura, K. Tamura, S. Chung, et al., Novel 5alpha-steroid reductase (SRD5A3, type 3) is overexpressed in hormone-refractory prostate cancer, Cancer Sci. 99 (2008) 81–86, http://dx.doi.org/10.1111/j.1349-7006.2007.00656.x.

[10] V. Cantagrel, D.J. Lefeber, B.G. Ng, et al., SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder, Cell 142 (2010) 203–217, http://dx.doi.org/10.1016/j.cell.2010.06.001.

[11] K. Yamana, F. Labrie, V. Luu-The, Human type3 5alpha-reductase is expressed in peripheral tissues at higher levels than types1 and 2 and its activity is potently inhibited by finasteride and dutasteride, Horm. Mol. Biol. Clin. Investig. 2 (2010) 293–299, http://dx.doi.org/10.1515/HMBCI.2010.035.

[12] K. Abe, Y. Ohno, T. Sassa, et al., Mutation for nonsyndromic mental retardation in the trans-2-enoyl-CoA reductase TER gene involved in fatty acid elongation impairs enzyme activity and stability, leading to change in sphingolipid profile, J. Biol. Chem. 288 (2013) 36741–36749, http://dx.doi.org/10.1074/jbc.M113.493221.

[13] Y.A. Moon, J.D. Horton, Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade, J. Biol. Chem. 278 (2003) 7335–7343, http://dx.doi.org/10.1074/jbc.M211 84200.

[14] M. Çalışkan, J.X. Chong, L. Uricchio, R. Anderson, et al., Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13, Hum. Mol. Genet. 20 (2011) 1285–1289, http://dx.doi.org/10.1093/hmg/ddq569.

[15] D.W. Russell, J.D. Wilson, Steroid 5alpha-reductase: two genes/two enzymes, Annu. Rev. Biochem. 63 (1994) 25–61, http://dx.doi.org/10.1146/annurev.bi.63.070194.000325.

[16] K. Normington, D.W. Russell, Tissue distribution and kinetic characteristics of rat steroid 5alpha-reductase isozymes. Evidence for distinct physiological functions, J. Biol. Chem. 267 (1992) 19548–19554. ISSN: 00219258.

[17] R. Bruni, B. Kloss, High-throughput cloning and expression of integral membrane proteins in E.coli, Curr. Protoc. Protein Sci. 74 (2013), http://dx.doi.org/10.1002/0471140864.ps2906s74. Unit 29.6.

[18] N.I. Tarasova, R.H. Stauber, J.K. Choi, et al., Visualization of G protein-coupled receptor trafficking with the aid of the green fluorescent protein. Endocytosis and recycling of cholecystokinin receptor type A, J. Biol. Chem. 272 (1997) 14817–14824, http://dx.doi.org/10.1074/jbc.272.23.14817.

[19] M. Koreishi, T.J. Gniadek, S. Yu, et al., The golgin tether Giantin regulates the secretory pathway by controlling stack organization within Golgi apparatus, PLoS One 8 (2013) e59821, http://dx.doi.org/10.1371/journal.pone.0059821.

[20] A. Chaumet, G.D. Wright, S.H. Seet, et al., Nuclear envelope-associated endosomes deliver surface proteins to the nucleus, Nat. Commun. 6 (2015) 8218, http://dx.doi.org/10.1038/ncomms9218.

[21] Z.J. Guo, L.X. Tao, X.Y. Dong, et al., Characterization of aggregate/aggresome structures formed by polyhedrin of *Bombyx mori* nucleopolyhedrovirus, Sci. Rep. 5 (2015) 14601, http://dx.doi.org/10.1038/srep14601.

[22] R.I. Silver, E.L. Wiley, A.E. Thigpen, et al., Cell type specific expression of steroid 5alpha-reductase 2, J. Urol. 152 (1994) 438–442, http://dx.doi.org/10.1073/pnas.90.20.9359.

[23] A.E. Thigpen, K.M. Cala, D.W. Russell, Characterization of Chinese hamster ovary cell lines expressing human steroid 5alpha-reductase isozymes, J. Biol. Chem. 268 (1993) 17404–17412. ISSN: 00219258.

[24] R.W. Hudson, Studies of nuclear 5alpha-reductase of human hyperplastic prostatic tissue, J. Steroid Biochem. Mol. Biol. 14 (1981) 579–584.

[25] R.J. Moore, J.D. Wilson, Localization of the reduced nicotinamide adenine dinucleotide phosphate: 4-3-ketosteroid 5-oxidoreductase in the nuclear membrane of the rat ventral prostate, J. Biol. Chem. 247 (1972) 958–967. ISSN: 00219258.

[26] B. Houston, G.D. Chisholm, F.K. Habib, Evidence that human prostatic 5alpha-reductase is located exclusively in the nucleus, FEBS Lett. 185 (1985) 231–235, http://dx.doi.org/10.1016/0014-5793(85)80912-1.

[27] J.G. Savory, D. May, T. Reich, et al., 5-alpha-reductase type1 is localized to the outer nuclear membrane, Mol. Cell. Endocrinol. 110 (1995) 137–147.

[28] M.F. Taylor, M. Wang, A.K. Bhattacharyya, et al., Expression of rat steroid 5 alpha-reductase (isozyme-1) in *Spodoptera frugiperda*, SF21, insect cells: expression of rat steroid 5 alpha-reductase, Steroids 62 (1997) 373–378.

[29] S. Délos, C. Iehlé, P.M. Martin, J.P. Raynaud, Inhibition of the activity of 'basic' 5alpha-reductase (type 1) detected in DU 145 cells and expressed in insect cells, J. Steroid Biochem. Mol. Biol. 48 (1994) 347–352, http://dx.doi.org/10.1016/0960-0760(94)90074-4.

[30] C. Iehlé, S. Délos, O. Guirou, R. Tate, J.P. Raynaud, P.M. Martin, Human prostatic steroid 5alpha-reductase isoforms. A comparative study of selective inhibitors, J. Steroid Biochem. Mol. Biol. 54 (1995) 273–279, http://dx.doi.org/10.1016/0960-0760(95)00134-L.

[31] B.U. Panter, J. Jose, R.W. Hartmann, 5alpha-reductase in human embryonic kidney cell line HEK293: evidence for type II enzyme expression and activity, Mol. Cell Biochem. 270 (2005) 201–208, http://dx.doi.org/10.1007/s11010-005-4508-8.

[32] L.N. Thomas, R.C. Douglas, R.S. Rittmaster, C.K. Too, Overexpression of 5alpha-reductase type1 increases sensitivity of prostate cancer cells to low concentrations of testosterone, Prostate 69 (2009) 595–602, http://dx.doi.org/10.1002/pros.20911.

[33] A. Poletti, F. Celotti, M. Motta, L. Martini, Characterization of rat 5alpha-reductases type1 and type2 expressed in Saccharomyces cerevisiae, Biochem. J. 314 (1996) 1047–1052, http://dx.doi.org/10.1042/bj3141047.

[34] G. Pelletier, V. Luu-The, F. Labrie, Immunocytochemical localization of 5alpha-reductase in rat brain, Mol. Cell. Neurosci. 5 (1994) 394–399, http://dx.doi.org/10.1006/mcne.1994.1049.

[35] Y. Tsuruo, K. Ishimura, K. Morita, Influence of serum-free culture conditions on subcellular localization of steroid 5alpha-reductase in rat C6 glioma cells, Brain Res. 801 (1998) 130–136, http://dx.doi.org/10.1016/S0006-8993(98)00555-1.

[36] B. Chávez, L. Ramos, R. García-Becerra, F. Vilchis, Hamster SRD5A3 lacks steroid 5alpha-reductase activity in vitro, Steroids 94 (2015) 41–50, http://dx.doi.org/10.1016/j.steroids.2014.11.005.

[37] A. Limon, J.M. Reyes-Ruiz, F. Eusebi, R. Miledi, Properties of GluR3 receptors tagged with GFP at the amino or carboxyl terminus, Proc. Natl. Acad. Sci. U.S. A 104 (2007) (2007) 15526–15530, http://dx.doi.org/10.1073/pnas.0706773104.

[38] S.I. Hyun, L. Maruri-Avidal, B. Moss, Topology of endoplasmic reticulum-associated cellular and viral proteins determined with split-GFP, Traffic 16 (2015) 787–795, http://dx.doi.org/10.1111/tra.12281.

[39] L. Kallal, J.L. Benovic, Using green fluorescent proteins to study G-protein-coupled receptor localization and trafficking, Trends Pharmacol. Sci. 21 (2000) 175–180, http://dx.doi.org/10.1016/S0165-6147(00)01477-2.

[40] E. Palmer, T. Freeman, Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays, Comp. Funct. Genomics 5 (2004) 342–353, http://dx.doi.org/10.1002/cfg.405.

[41] T. Kawate, W. Gouax, Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins, Structure 14 (2006) 673–681, http://dx.doi.org/10.1016/j.str.2006.01.013.

Attachment III

PLOS ONE

RESEARCH ARTICLE

Engineered ferritin for lanthanide binding

Lorenzo Calisti[1☯], Matilde Cardoso Trabuco[2☯], Alberto Boffi[3,4], Claudia Testi[1], Linda Celeste Montemiglio[1,4], Amédée des Georges[5], Irene Benni[1], Andrea Ilari[4], Bartłomiej Taciak[6,7], Maciej Białasek[6], Tomasz Rygiel[7,8], Magdalena Król[6,7], Paola Baiocco[3]*, Alessandra Bonamore[1]

1 Department of Biochemical Sciences "Alessandro Rossi Fanelli", Sapienza University of Rome, Rome Italy, **2** Molirom srl, Rome, Italy, **3** Center for Life Nano Science @ Sapienza, Istituto Italiano di Tecnologia, Rome, Italy, **4** Institute of Molecular Biology and Pathology, National Research Council, Rome, Italy, **5** The City University of New York Advanced Science Research Center, New York, NY, **6** Faculty of Veterinary Medicine, Warsaw University of Life Sciences, ul. Nowoursynowska, Warszawa, Poland, **7** Cellis Ltd., Warsaw, Poland, **8** Department of Immunology, Center for Biostructure Research, Medical University of Warsaw, Warsaw, Poland

☯ These authors contributed equally to this work.
* paola.baiocco@iit.it

OPEN ACCESS

Citation: Calisti L, Trabuco MC, Boffi A, Testi C, Montemiglio LC, des Georges A, et al. (2018) Engineered ferritin for lanthanide binding. PLoS ONE 13(8): e0201859. https://doi.org/10.1371/journal.pone.0201859

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: June 7, 2018

Accepted: July 23, 2018

Published: August 13, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Final atomic coordinates and structure factors have been deposited on PDB Data Bank (www.rcsb.org) with accession code 5OBA and 5OBB.

Funding: This study has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 312284. The biological part of the study was supported by the grant from National Science Centre (NCN) in Poland no. UMO-2015/18/E/NZ6/00642.EU H2020 Project "X-Probe", Grant N

Abstract

Ferritin H-homopolymers have been extensively used as nanocarriers for diverse applications in the targeted delivery of drugs and imaging agents, due to their unique ability to bind the transferrin receptor (CD71), highly overexpressed in most tumor cells. In order to incorporate novel fluorescence imaging properties, we have fused a lanthanide binding tag (LBT) to the C-terminal end of mouse H-chain ferritin, HFt. The HFt-LBT possesses one high affinity Terbium binding site per each of the 24 subunits provided by six coordinating aminoacid side chains and a tryptophan residue in its close proximity and is thus endowed with strong FRET sensitization properties. Accordingly, the characteristic Terbium emission band at 544 nm for the HFt-LBT Tb(III) complex was detectable upon excitation of the tag enclosed at two order of magnitude higher intensity with respect to the wtHFt protein. X-ray data at 2.9 Å and cryo-EM at 7 Å resolution demonstrated that HFt-LBT is correctly assembled as a 24-mer both in crystal and in solution. On the basis of the intrinsic Tb(III) binding properties of the wt protein, 32 additional Tb(III) binding sites, located within the natural iron binding sites of the protein, were identified besides the 24 Tb(III) ions coordinated to the LBTs. HFt-LBT Tb(III) was demonstrated to be actively uptaken by selected tumor cell lines by confocal microscopy and FACS analysis of their FITC derivatives, although direct fluorescence from Terbium emission could not be singled out with conventional, 295–375 nm, fluorescence excitation.

Introduction

Ferritin is a cage-like protein made of 24 subunits arranged in octahedral 432 symmetry with an outer diameter of roughly 12 nm and an inner diameter of 8 nm [1,2]. The symmetrical positioning of three or four subunits in the protein shell results in the formation of eight

" 637295, to A. Bonamore and M.C.T. is gratefully acknowledged. Flagship project "Nanomax" from MIUR to A. Boffi is also acknowledged. The funder provided support in the form of salaries for authors Trabuco, MC, who is employed at Molirom as a Marie Curie Fellow within the framework of the X-probe EU project, and Taciak B., Rygel T and Krol M., who are employed at Cellis within the framework of the National Science Center of Poland. The companies did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: The funder provided support in the form of salaries for authors Trabuco, MC, who is employed at Molirom as a Marie Curie Fellow within the framework of the X-probe EU project and Taciak B., Rygel T and Krol M., who are employed at Cellis within the framework of the National Science Center of Poland. The companies did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

channels connecting the inner cavity to the outside and allows for the entry and exit of iron and other cations with a relatively broad selectivity [3]. Besides their physiological function, centered around intracellular iron uptake, the nanocage properties of ferritins have been exploited in a number of diverse biotechnological applications as drug delivery vectors [4], scaffolds for vaccine development [5] and tools for bioimaging [6]. In this context, ferritins have been proven particularly useful for the selective targeting to cell populations overexpressing the transferrin receptor (CD71), in particular iron avid tumor cells [7,8]. Ferritin H-homopolymers are in fact endowed with unique properties that confer themselves several advantages over CD71 main ligand, *i.e.* transferrin, and even over targeted antibodies. First of all, ferritins display a 250 $Å^3$ hollow cage capable of hosting tens to hundreds small molecules, either physically "encapsulated" or covalently attached to the inner surface of the cavity. In contrast, transferrin or conventional antibodies can be engineered to covalently attach only a few molecules in selected positions far from the receptor recognition epitopes. Thus, ferritins display a very convenient "drug/protein ratio" with respect to the classical "drug/antibody ratio". As a second instance, it must be considered that ferritins display 24 symmetrically related recognition epitopes for the CD71 receptor instead of the two of classical antibodies or two of transferrins. Overall, the multiplicity of recognition epitopes in ferritins results in a high affinity towards the target receptor (a property often referred to as "multivalence effect" [9]), that, in the case of CD71, rivals with typical antibodies affinities [10].

Among the many ingenious ferritin-based constructs for bioimaging, only a few have been devoted to the creation of smart fluorescent probes, and these include quantum dots gold nanoparticles and fluorescent metal chelators [11–13]. However, advanced optical imaging techniques need an expanded color palette of bright fluorescent probes for biological visualization in order to enable real-time cellular imaging with high spatial resolution for close-up view into subcellular compartments and for providing key information on intracellular activities and macromolecular dynamics.

In this framework, fluorescent probes based on trivalent lanthanide ions are becoming widespread due to their unique photophysical properties [14–16]. Lanthanide f-orbitals with their high quantum yields are capable of efficiently radiating most of the absorbed energy, although their small absorption cross sections pose limits to their use. To improve the luminescent signals, small organic fluorophores that absorb in the UV region and transfer the absorbed light to the lanthanide atom are thus currently used in complex with the metal ion. Complexes of lanthanides are characterized by narrow band emission spectra, large Stokes shift (150–300 nm), and excited state lifetimes within the range of micro to milliseconds. By exploiting the microsecond fluorescence of lanthanides, time-resolved spectroscopy allows for the elimination of the short living background signals whose lifetimes are usually not more than 10–15 ns and the enhancement of the sensitivity for recording the delayed signal [17]. Moreover, the inherently low extinction coefficient of lanthanide ions, due to the forbidden character of the electronic transitions, can be overcome by Förster resonance energy transfer (FRET) from an appropriately placed (within 5–6 Å distance) sensitizer-fluorophore onto the emitting level of the lanthanide, a phenomenon often referred to as "antenna effect" [18]. Among lanthanides, Tb(III) and Eu(III) are the most interesting due to their more intense microsecond fluorescence in the visible region [19, 20]. In a first approach, lanthanide fluorescent chelators were directly incorporated within the ferritin cavity using the "encapsulation" method, which entails pH induced ferritin dissociation with subsequent trapping of the payload [12]. However, at least in the case of fluorescent probes, this method leads to a random and inhomogeneous distribution of the small organic molecules within the inner cavity, resulting in possible aggregation-induced quenching effects and/or leakage outside the cavity through the iron channels. Among the various methods for the incorporation of lanthanide

ions into biomolecules, a straightforward and generalizable approach has been proposed in recent years that integrates a lanthanide binding sequence as a protein co-expression tag *via* molecular biology strategies. On the basis of known properties of calcium binding loops, recent design and engineering studies have resulted in the development of short polypeptides comprising 20 or fewer encoded aminoacids that are capable of tight and selective binding to lanthanides. These peptide sequences, identified by Imperiali's group as "lanthanide-binding tags" (LBT), show low-nanomolar affinities for the target ions and are selective for lanthanides over other common metal ions [21–24]. LBTs thus represent a most convenient option for lanthanide protein labeling in that they can be directly encoded within a recombinant protein expression construct. Improved LBTs sequences have been developed that entails genetic encoding of polypeptide sequences at the N-terminal or C-terminal of a specific carrier protein or even insertion into specific protein loops [25]. The probe nature of these protein tags has been demonstrated by their use for luminescence-based visualization on gels, as magnetic-field paramagnetic alignment agents in protein NMR experiments [26, 27] in fluorescence microscopy [28], and as partners in luminescence resonance energy transfer (LRET) studies [29].

Among the number of proteins physiologically involved in metal binding that can be used for Terbium binding, ferritins may appear as one of the most suitable. In fact, native apoferritins have been reported to naturally bind Tb(III) within the iron binding sites [30]. In particular, it was demonstrated that mammalian apoferritins could bind more than one Tb(III) ion per subunit, corresponding respectively to ferroxidase site, threefold channel and nucleation centre [31]. Moreover, upon excitation between 280–295 nm Tb(III) ferritin complex showed characteristic emission bands at 490 ($^5D_4$-$^7F_6$ Tb electronic transition) and at 544 nm ($^5D_4$-$^7F_5$ Tb electronic transition) due to a FRET sensitization effect provided by aromatic aminoacids [31b]. However, the distance between Terbium ions and aromatic moieties in native ferritin isoforms made FRET efficiency very poor and suboptimal for any type of fluorescence/luminescence based measurement. Moreover, the original Terbium binding sites of the wild type protein were shown to exhibit variable affinities for Tb(III), with estimated Kd ranging from 2 to 666 μM at pH 6.4 [31a].

In the present paper, a high Tb(III) affinity LBT sequence has been genetically fused at the C-terminal end of the heavy chain of mouse ferritin. The tag has been designed to be located inside the 250 $Å^3$ wide inner cavity such that the lanthanide ions diffusing through the surface pores could eventually bind to the LBT sequence. The construct would thus act both as carrier targeted to CD71 receptors and as a FRET sensitizer. Mouse ferritin was used in view of the identical sequence within the CD71 binding region as the human ferritin sequence [32] and because of obviously more favourable immunogenic profile for forthcoming in vivo study in mouse.

Methods

Protein design, expression and purification

A synthetic gene encoding for mouse H chain ferritin (HFt) fused with a lanthanide binding peptide (LBT) was designed, synthesized, and optimized for *Escherichia coli* codon usage by Geneart (Geneart AG). The last five C-terminal aminoacids of H chain ferritin were replaced by the GSG spacer sequence, followed by the LBT sequence YIDTNNDGWIEGDELLA [24]. As such, the LBT sequence is designed to be located within the inner cavity, as a prolongation of the inward directed C-terminal region (or E helix in ferritin secondary structure nomenclature). The resulting HFt-LBT construct, was subcloned into a pET22-b vector (Novagene)

between NdeI/XhoI restriction sites. A stop codon was inserted before the His-tag region to avoid transcription of the unwanted tag.

HFt-LBT was overexpressed in *Escherichia coli* BL21 cells upon induction with 1 mM IPTG (Isopropyl-β-D-1-thiogalactopyranoside) at OD_{600} = 0.6. Cells were harvested by centrifugation 16 hours post induction at 37˚C.

Cells harvested from 1 L culture were resuspended in 20 mM HEPES buffer, pH 7.5, containing 300 mM NaCl, 1 mM TCEP (tris(2-carboxiethyl)phosphine), and a cOmplete Mini Protease Inhibitor Cocktail Tablet (Roche). Cells were disrupted by sonication and the soluble fraction was purified by heat treatment at 78˚C for 10 minutes. Denatured *E. coli* proteins were removed by centrifugation at 12000 rpm at 4˚C for 1 hour. The soluble protein was further purified by ammonium sulfate precipitation. The precipitated fraction at 70% ammonium sulfate was resuspended in 20 mM HEPES, 150 mM NaCl, pH 7.5 and dialyzed versus the same buffer. As final purification step, the protein was loaded onto a HiLoad 26/600 Superdex 200 pg column previously equilibrated in the same buffer using an ÄKTA-Prime system (GE Healthcare) (S1 Fig). Purified protein (S2 Fig) was concentrated to obtain the final protein preparation of 1 mg/mL and protein concentration was calculated by measuring the UV spectrum using an extinction coefficient of 26000 $M^{-1}cm^{-1}$. A small amount of higher molecular weight aggregates was observed in all ferritins preparations (less than 5%), and attributed to the presence of intermolecular disulphide bridges, possibly involving the surface exposed Cys103 (S1 Fig and inset). As indicated Protein yield was ~50 mg/L culture.

The expected molecular weight of 22662 Da was confirmed by MALDI-TOF Mass Spectrometry as reported in Supplementary Data (S3 Fig).

MALDI-TOF mass spectrometry

40 μl of protein sample were desalted on a C8 Empore Disk (3M, Minneapolis, MN) homemade stage tip and resuspended in 3 μl formic acid 1%. 1 μl was spotted on a MALDI sample plate and allowed to air dry. Recrystallized sinapinic acid (SA matrix from Thermo Fisher Scientific) was prepared at a concentration of 5 mg/ml in 50:50 acetonitrile/water (0.1% Formic Acid) and spotted directly prior to insertion into the mass spectrometer.

Matrix-assisted laser desorption ionization (MALDI) mass spectra were acquired on 4800 MALDI-TOF/ TOF mass spectrometer (Applied Biosystems, Foster City, CA) equipped with a nitrogen laser operated at 336 nm laser. Acquisitions were performed in linear mode averaging 2500 laser shots in a random, uniform pattern. Ions were accelerated with a 20 kV pulse, with a delayed extraction period of 860 ns. Spectra were generated by averaging between 500 and 2000 laser pulses in a mass range from 4 kDa to 50 kDa. Laser intensity was set to optimize the signal-to-noise ratio and the resolution of mass peaks of the analyte. All spectra were externally calibrated and processed via Data Explorer software (version 4.7).

Fluorescence spectroscopy

Fluorescence spectra and titrations were performed using FluoroMax 4 (Horiba) spectrofluorimeter with a Haake D8 refrigerated bath at 20˚C. Emission spectra were recorded between 450 and 560 nm, in order to include the luminescent maxima of Tb(III) (490 and 545 nm). The excitation wavelength was chosen at 295 nm to minimize the overlap of second order diffraction (570 nm) with the Tb(III) emission at 545 nm. Emission spectra were taken with excitation and emission band passes of 4 and 8 nm and corrected for the blank contribution and the instrument response at 295 nm in a quartz cell of 1 cm pathlength. Emission spectra were normalized to 1 at 545 nm.

Fluorescence static spectra were performed using 1 μM (monomer) apoHFt-LBT and wild type apoHFt as a control in 100 mM MES buffer pH 6.4. A 50 mM $TbCl_3$ anhydrous powder (Sigma Aldrich) stock solution was also prepared in MES buffer at pH 6.4. Under these conditions, precipitation of Terbium hydroxides, easily formed around neutrality, is avoided. Fluorescence spectra of the protein Tb(III) complexes were recorded after 30 min incubation and after addition of 150 μM $TbCl_3$ in buffer solution in order to saturate all possible Tb(III) binding sites both in HFt-LBT and HFt. Before recording spectra, protein solutions were exchanged with buffer (Terbium free) by centrifugal ultrafiltration on vivaspin MWCO 100 kDa concentrators (Sartorius) in order to remove unbound and weakly bound metal ions (5 exchange steps, 1×10 concentration each). Protein concentration was measured again and adjusted to the final concentration with buffer (1 μM monomer). Samples used for crystallographic and cryoEM measurements were also prepared according to this procedure.

Fluorescence titrations of HFt-LBT were carried out by adding incremental amounts (1.8 μL or multiples) of a 0.5 mM $TbCl_3$ stock solution in 0.1 M MES buffer pH 6.4 to 3 mL of 1μM (monomer) protein in the same buffer under stirring. Emission spectra were recorded in 1 cm pathlength cuvette upon excitation at 295 nm at 25˚C, 30 minutes after addition of $TbCl_3$ solution aliquots. Fluorescence intensity of HFt-LBT Tb(III) complex as a function of the Tb (III)/HFt-LBT ratio has been reported. Fluorescent intensity was recorded at 545 nm corrected for the dilution factor and normalized to the emission maximum.

Crystallization and X-ray structure determination

Crystals of wild type HFt-LBT and HFt-LBT in complex with Tb(III) were obtained by mixing in a 2 μl hanging drop the purified protein at 15 mg/ml with a solution containing 1.8/2.0 M ammonium sulphate and 0.1 M Tris, pH8.5, at 25˚ C within a week, cryo-protected by extensively washing the crystals in sodium malonate and flash-frozen in liquid nitrogen. Diffraction data were collected at λ = 1.0 Å and 1.4 Å respectively at XRD1 beamline at the Elettra Synchrotron, Trieste, Italy.

Data were processed with XDS [33] and scaled with Aimless [34] at final resolution of 2.85 Å and 2.65 Å, respectively. The structures were solved by Molecular Replacement with Phaser [35, 36] using the structure of mouse H-chain modified ferritin (pdb code 3WNW) without waters and ligands as a starting model. Model building and refinement were done using Coot [37] and Refmac5 [34]. Anomalous difference electron density map was calculated from the HFt-LBT Tb(III) crystal diffraction data, collected at the Tb emission peak (λ = 1.4 Å). The map has been generated using the Fast Fourier Transform Program belonging to the CCP4 suite [34]. The final model was analyzed with PROCHECK [38] and Molprobity [39]. Ramachandran Plot showed that more than 98% residues were in preferred regions and no outlier was observed in both structures. The validation of metal binding sites was performed using CheckMyMetal web server [40]. Final atomic coordinates and structure factors of apoHFt-LBT and HFt-LBT with Tb(III) were deposited in the PDB Data Bank (www.rcsb.org) with accession code 5OBA and 5OBB, respectively. Complete data collection and refinement statistics are reported in Supplementary Data (S1 Table) and validation reports were available for the review process.

Cryo-electron microscopy

Holey-gold grids from Quantifoil R1.2/1.3 (Quantifoil Micro Tools GmbH) were prepared as described [41]. Grids surfaces were treated with plasma cleaning using a mixture of H_2 and O_2; then, 3 μl of a solution of 20 mM HEPES and 150 mM NaCl, pH 7.5, containing 1 μM (24-mer concentration) HFt-LBT Tb(III) complex was applied. After 30 s waiting time, grids were

blotted (3 s) at 100% humidity with filter paper and vitrified by rapidly plunging into liquid ethane at −180˚C with a Vitrobot Mark IV (FEI) [42, 43].

Data acquisition of 350 micrographs was performed using a FEI Titan Halo (FEI, Eindhoven) operating at 300 kV, while the specimen was maintained at liquid nitrogen temperatures. Data sets were collected with an automated data collection system [44] on a K2 Summit direct detector camera (Gatan, Pleasanton) operating in super-resolution mode, with a calibrated pixel size of 1.15 Å on the object scale and a magnification of 59000×. Images were typically recorded with a defocus range between −0.7 and −3.0 μm and using a dose of electrons on the specimen plane ranging between 10 and 20 electrons/$Å^2$.

Data analysis was carried out using RELION 2.0 [45], while motion correction was performed using MotionCor2 [46]. A number of particles (91947) was picked and extracted from the original micrographs with the reference-based automated particle picking procedure implemented in RELION [44]; after the extraction, particles were 2D classified using 100 classes. Particles (48047) belonging to good 2D classes were selected and subjected to 3D classification, using as reference model the mouse H-ferritin structure (PDB code 3WNW), ultimately yielding 8 classes (S4 Fig). A selection of 35625 particles were subjected to another round of 3D classification. The resulting good class (7345 particles) was refined with the 3D Autorefine procedure in RELION and then subjected to the Post Processing step [47].

The final Post Processing 3D map resolution (equal to 7.1 Å) was estimated with the Fourier shell correlation (FSC) = 0.143 criterion, based on the 'gold-standard' protocol [48], using soft masks with a 4 pixel soft edge, and it has been corrected for the effects of the mask on the FSC curve using high-resolution noise substitution (S5 Fig) [49]. The final map was visualized using UCSF Chimera [50].

Protein FITC labeling

HFt-LBT and HFt were labeled with Fluorescein-isothiocyanide (Sigma Aldrich) according to the manufacturer's standard protocol. Briefly, 2 mg/ml (94 μM monomer) of purified protein was equilibrated in sodium carbonate buffer pH 9.0. 50 μl of 1 mg/ml (2.6 mM) FITC in DMSO were added to 1 ml of protein solution and the reaction mix was incubated under stirring for 2 hours at room temperature. The excess dye was removed by gel filtration chromatography and dye/protein ratio was determined by UV-spectroscopy. Both proteins were labeled with 0.95 FITC molecules per each subunit.

Cell cultures and ferritins internalization

Human prostate cancer cell line DU-145 (ATCC HTB81), human colorectal cancer cell line HCT-116 (ATCC CCL-247), human breast cancer cell line MDA-MB-231 (ATCC HTB-26) and human ovarian cancer cell line SKOV-3 (ATCC HTB-77) were cultured in DMEM medium containing 10% FBS, 100 μg/ml streptomycin and 100 U/ml penicillin G in a humidified 37˚C incubator. The internalization assay was performed as follow: cells were detached using trypsin, then washed twice with PBS and resuspended in non-supplemented DMEM medium containing FITC-ferritin nanoparticles (HFt-LBT or mouse HFt as a control) at the final concentration of 0.5 mg/ml for 1 hour in a humidified 37˚C incubator. After incubation cells were washed three times with PBS and subjected for confocal microscopy and flow cytometry analysis.

Confocal microscopy

Following internalization step described above, cells were seeded into 8-well Nunc Lab-Tek Chambered Coverglass with 200 μl DMEM medium containing 10% FBS, 100 μg/ml

streptomycin and 100 U/ml penicillin G per well. Chambers with cells were then incubated on ice until microscopic visualization. Images were acquired using an inverted confocal microscope IX70 FV 500 (Olympus), with 488 nm laser, 20x objective lens and emission filter 505–560 nm. Image processing was performed using ImageJ software (National Institutes of Health, https://imagej.nih.gov/ij/) [51].

Flow cytometry analysis

Cells were incubated with FITC-ferritin nanoparticles as described previously, then washed three times with PBS, and resuspended in FACS buffer (2% FBS, 1 mM EDTA). Internalization of ferritins before and after treatments was measured at the BD FACS Aria III equipped with a 488 nm laser. Cells were first gated by forward and side scatter area (FSC-A and SSC-A) plot and for singlets population (FSC-H and SSC-A), then detected in the channel for FITC expression (530/30nm filter) and side scatter parameter. The gate for the final detection was set according to the gate set on the control sample. Data were analyzed using BD FACSDIVA and FlowJo softwares.

Results

Fluorescence spectroscopy

Static emission spectra were recorded for Tb(III) saturated HFt-LBT and wild type mouse HFt after removal of unbound Tb(III) by centrifugal ultrafiltration (see Methods, section 2.3), upon excitation at 295 nm. As reported in Fig 1, the intensity of the Terbium emission peak at

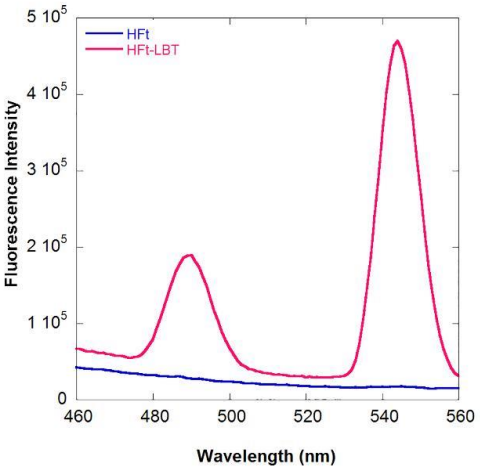


Fig 1. Fluorescence spectra of HFt-LBT Tb(III) complex. Fluorescence spectra of HFt-LBT Tb(III) (red line) and wild type mouse HFtTb(III) complexes (blue line) at the same protein concentration (1 μM monomer). The main emission band due to the $^5D_4$-$^7F_5$ (Tb d-f orbitals) electronic transition at 544 nm is accompanied by the weaker $^5D_4$-$^7F_6$ transition. Spectra were recorded after protein saturation with $TbCl_3$ in 0.1 M MES buffer pH 6.4 (see Methods).

https://doi.org/10.1371/journal.pone.0201859.g001

544 nm for HFt-LBT was at least two orders of magnitude higher with respect to HFt, indicating that the presence of the LBT sensitizes the Terbium signal due to the presence of a tryptophan residue within the tag at an efficient energy transfer distance in HFt-LBT.

Fluorescence titration analysis was thus carried out on HFt-LBT by adding free Tb(III) ions to the apo protein (Fig 2A and 2B and Methods, section 2.3).

Intriguingly, the titration endpoint was reached at 1.7 equivalent amounts of Tb(III) per subunit instead of the predicted 1 equivalent based on the presence of one LBT moiety per monomer. Analogous titrations carried out on wild type mouse HFt, demonstrated a negligible fluorescence contribution provided by Terbium bound to the metal binding sites of the native protein (*i.e.* ferroxidase site, threefold channel and nucleation centre). Thus, the presence of the tryptophan residue within the LBT, appears to contribute for most of the observed fluorescence signal. The discrepancy between the observed and expected titration endpoint might consist in the additional contribution of the LBT tryptophan residue to the enhanced emission not only of the Tb(III) atom within the tag itself but also to other Tb(III) atoms present in the ferroxidase site or in the threefold channels. The demonstration of such effect may rely only on the generation of multiple mutants on the ferroxidase center and threefold channel that are unable to bind Tb(III). More interestingly, positioning of additional tryptophan residues next to these binding sites by mutagenesis could provide further enhancement of fluorescence/luminescence Tb(III) yields besides the LBT tag and they will be evaluated in the future. Further experimental work will also be needed in order to single out individual dissociation constants from each Tb(III) binding site both at equilibrium and in kinetic measurements. At present, the thermodynamics of this complex system can only rely on the published Kd relative to Tb(III) binding to the LBT of 57 nM [22, 23] and those relative to Tb(III) binding sites on wild type horse spleen heteropolymers, ranging from 2 to 666 μM [30]. Again, analysis of combined mutants will be necessary in order to solve the complex thermodynamics of the system. As a last comment, it is worth considering the possible effect of added iron II in solution on the stability of the HFt-LBT Tb(III) complex in view of the reported reversible competition among the diverse binding sites. A control experiment, depicted in S10 Fig, demonstrated that

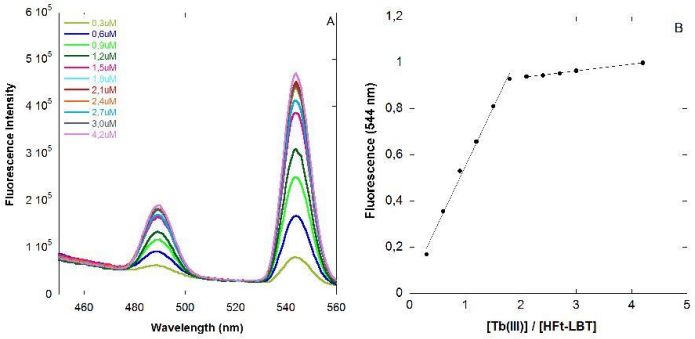


Fig 2. Fluorescence titration of HFt-LBT with Tb(III). A) Fluorescence titration of HFt-LBT (1μM monomer) with incremental concentration of Tb(III) in 0.1 M MES buffer pH 6.4. Emission spectra were recorded in 1 cm pathlength cuvette upon excitation at 295 nm. B) Fluorescence intensity of HFt-LBT Tb(III) complex as a function of the Tb(III)/HFt-LBT ratio. Fluorescent intensity was recorded at 545 nm, corrected for dilution factor and normalized to the emission maximum.

https://doi.org/10.1371/journal.pone.0201859.g002

addition of stoichiometric amounts of iron II (1 iron atom per subunit) brought about a small decrease (2–3%) in the observed HFt-LBT Tb(III) fluorescent signal, accounting for possible displacement of Tb(III) ions from accessible sites, most likely the three-fold channels. No further decrease was observed over 48 h incubation and increasing concentration of added iron resulted in formation of insoluble iron oxide aggregates.

HFt-LBT and HFt-LBT-Tb(III): Structural analysis

The structures of wild type apoHFt-LBT and HFt-LBT Tb(III) were determined by X-ray crystallography at 2.85 Å and 2.65 Å resolution, respectively. They both crystallized in I222 space group with 24 identical subunits in the asymmetric unit (ASU) with a solvent content of 64.7%. The overall X-ray-structures confirmed that the presence of the LBT tail does not affect the overall protein scaffold, which thoroughly matches the native H-chain mouse ferritin (pdb code 3WNW) with a *rmsd* value of 0.1 Å. However, the model building process allowed to position residues just from 3 to 176 since the LBT loop was not visible, probably because of the intrinsic high flexibility of the C-terminus region. When looking at the H-chain mouse ferritin structure, each ferroxidase center and 3-fold axes displayed one and two Mg(II) ions, respectively, due to the crystallization buffer solution. In a similar way, in each ferroxidase center and 3-fold axes of wild type apoHFt-LBT structure, a significant electron density peak was indeed observed and ascribed to iron ions (refined with occupancy $<100\%$), since, although no salt was in the crystallization solution, the X-ray fluorescence scan displayed the presence of residual iron atoms. As a matter of fact, residual iron (5 atoms per ferritin 24-mer) is consistently present in recombinant ferritin H-homopolymers, most likely due to iron uptake within bacterial environment. In addition, in 3WNW structure, one iron ion was modelled in each 4-fold axis while in HFt-LBT structure, one water molecule was positioned and successfully refined.

When comparing the wild type apoHFt-LBT and HFt-LBT Tb(III), the LBT loop was still not visible while a very large electron density peak appeared in each ferroxidase center and in the eight 3-fold axes. In the X-ray emission scan, ranging from 4.0 to 21.0 KeV, the typical X-ray emission lines of Terbium were clearly identified in the HFt-LBT Tb(III) crystals (S6 Fig). Accordingly, the crystals of wild type apoHFt-LBT lacked these emission energies.

As shown in Fig 3A, in HFt-LBT Tb(III) structure, 32 Terbium ions were positioned in each three-fold axis and in each ferroxidase site, depicted as yellow and magenta spheres, respectively. The position of Tb(III) ions was clearly assigned by comparison of the 2Fo-Fc electron density maps of the HFt-LBT Tb(III) and the wild type apoHFt-LBT was used as a control. An additional demonstration of the Terbium assignments was obtained by anomalous difference electron density map, which clearly displayed the positions attributed to Terbium (S7 Fig). In the ferroxidase center, each Terbium ion was successfully refined with 70–75% occupancy and was located in a trigonal planar coordination to OE1-Gln141 and to OE1 and OE2-Glu62, and to OE2-Glu27, in a range of 2.6–3.2 Å of distance (Fig 3B). In the 3-fold center, the Terbium ion was successfully refined with 100% occupancy and it was tetrahedrally coordinated to OE1-Glu134 of the three subunits (at 2.2–2.4 Å distance) (Fig 3C). In each 4-fold channel, one water molecule was positioned and successfully refined as well as in the native protein.

With the aim to assess the 24-mer assembly of HFt-LBT and the degree of homogeneity of the sample in solution in a near native condition, HFt-LBT Tb(III) was analysed by cryo-electron microscopy.

A qualitative analysis of the collected micrographs showed the high level of homogeneity and monodispersion of the HFt-LBT sample (S8 Fig). The processed imaged particles yielded

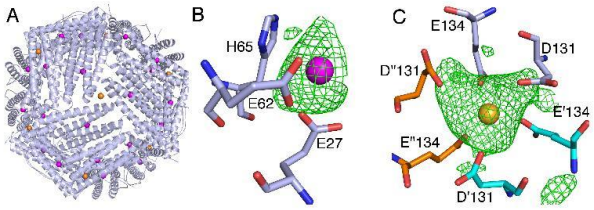


Fig 3. Terbium binding sites from X-ray crystallography. A) An overall view of the 24-meric shell of HFt-LBT Tb (III) showing the positions of Tb(III), displayed as spheres, is shown. In the close-up view, the omit map contoured at 3 σ as a green mesh is shown for the B) ferroxidase center where Tb(III) is depicted in magenta and for the C) 3-fold axes where Tb(III) is depicted as a yellow sphere and the residues Glu131 and Glu134 of three different monomers are depicted as sticks in light blue, cyan and orange, respectively.

https://doi.org/10.1371/journal.pone.0201859.g003

a ≈7 Å resolution 3D map of the protein that overall corresponds to the high-resolution structure obtained by X-ray crystallography, indicating a perfect matching of the subunit assembly and the helix axes (Fig 4).

Nevertheless, differently from the X-ray map, the cryo-EM structure highlights the presence of significant L-shaped electron densities that protrude from the last C-term residue of the E helix of the X-ray HFt-LBT Tb(III) model to the internal cavity of the protein and that can be attributed to the lanthanide binding tags. These fragments, covering a distance of about 14 Å, correspond to segments of 6–7 residues each; they can also be seen in selected 3D classes (S4 Fig).

Thus, cryo-EM data clearly indicate that the E-helix is in its "flip" conformation in solution and presence of possible "flop" conformation, with the E-helix oriented outside the cavity, is not observed in any of the 3D classes [52]. Nevertheless, the considerable heterogeneity of the observed signals did not allow a reliable model reconstruction and only a partial (i.e. 6–7 aminoacids on a total of 17) segment of the tail was identified.

Ferritins uptake by tumor cell lines

After the demonstration that HFt-LBT Tb(III) maintained the overall structure of the wild type mouse HFt, binding and internalization of the construct was analyzed in selected cancer cell lines: in fact, it is known that HFt is recognized and internalized by CD71 receptor (also known as TfR1), which is overexpressed in many types of tumor cells but not in normal cells and healthy tissues [53]. Experiments on cells treated with the same amount (23.5 μM monomer, corresponding to 0.5 mg/ml) of HFt-LBT Tb(III) or wild type HFt were undertaken in order to study uptake efficiency by cancer cells by flow cytometry and confocal microscopy. As a baseline for FITC fluorescence, control cells untreated with FITC-ferritins were used. In order to exclude any signal generated from outside particles sticking on the cell membrane (due to unspecific binding or remaining from the washing steps), Trypan blue quenching was performed before FACS acquisition. In S9 Fig, FACS acquisition plots analysis is summarized, shown as the percentage of cells internalizing the nanoparticles: these data highlighted that HFt-LBT Tb(III) nanoparticles are efficiently taken up by all cells of each cell line. Confocal representative images of entire field of view of live cells incubated with HFt-LBT Tb(III) are shown in Fig 5A. Images confirmed the high extent of both HFt-LBT Tb(III) and wild type mouse HFt internalization and highlighted similar cellular distribution in the cytoplasm and in the perinuclear space.

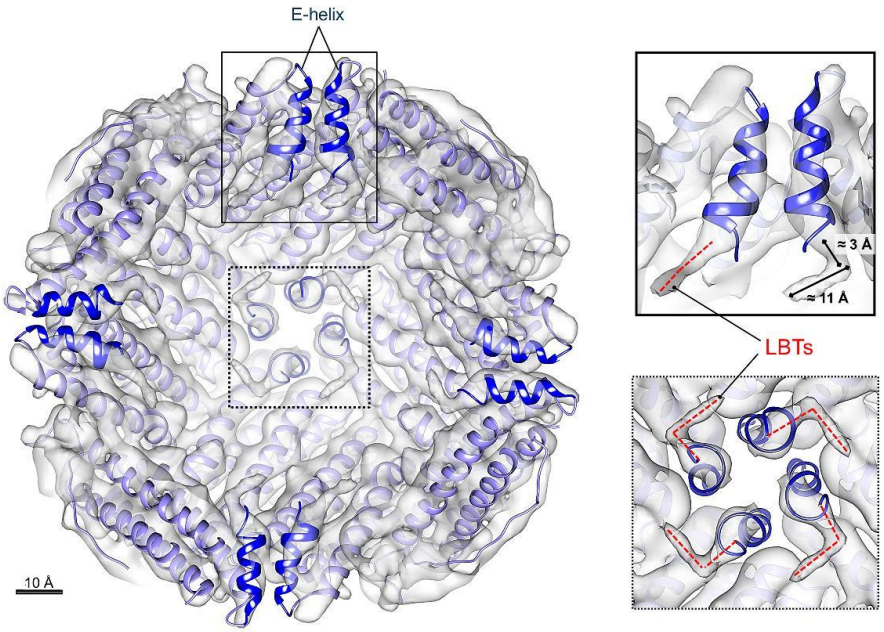


Fig 4. Cryo-electron microscopy structure of HFt-LBT. Internal view, sliced at the protein center, of the density map of HFt-LBT Tb(III) (resolution = 7.1 Å). The crystallographic model of HFt-LBT Tb(III) (purple ribbons) was rigid-body fitted into the Cryo-EM map (gray solid) using USCF Chimera [50]. The E helices at the C-terminal part of each monomer are highlighted in blue. LBT tails fused to the E helices are enlarged on the right panels showing the top view (dashed line box) and the side-view (continuous line box). Well-defined electron density extends over the end of the X-ray model as due to the presence of the terminal LBT tails and it approximately covers 14 Å distance, that corresponds to the length of 6–7 amino acids (main chain). Scale bar = 10 Å.

https://doi.org/10.1371/journal.pone.0201859.g004

Discussion

In the present work, an engineered ferritin construct, bearing a lanthanide binding tag on the C-terminal end of each subunit of mouse H ferritin, was designed in order to build a biomolecular nano-system endowed with strong FRET sensitization properties. The peptide tag (LBT) was designed according to the findings of Martin *et al.* [24] that demonstrated efficient lanthanide chelating properties of a peptide sequence derived from Ca^{2+} binding sites from Troponin C EF hand motif. The LBT has six metal-binding residues that form a coordination sphere around a lanthanide (III) ion and is characterized by high affinity and strong FRET effect from the tag encoded tryptophan residue to the Terbium or Europium ions. Spectroscopic and crystallographic studies showed that the inner coordination sphere of Tb(III) bound to this sequence was free from water molecules [23], a key feature for luminescence experiments, given the quenching effect of O-H vibration [54]. Accordingly, the HFt-LBT construct exhibited high affinity Tb(III) binding as demonstrated by fluorescence spectroscopy measurements

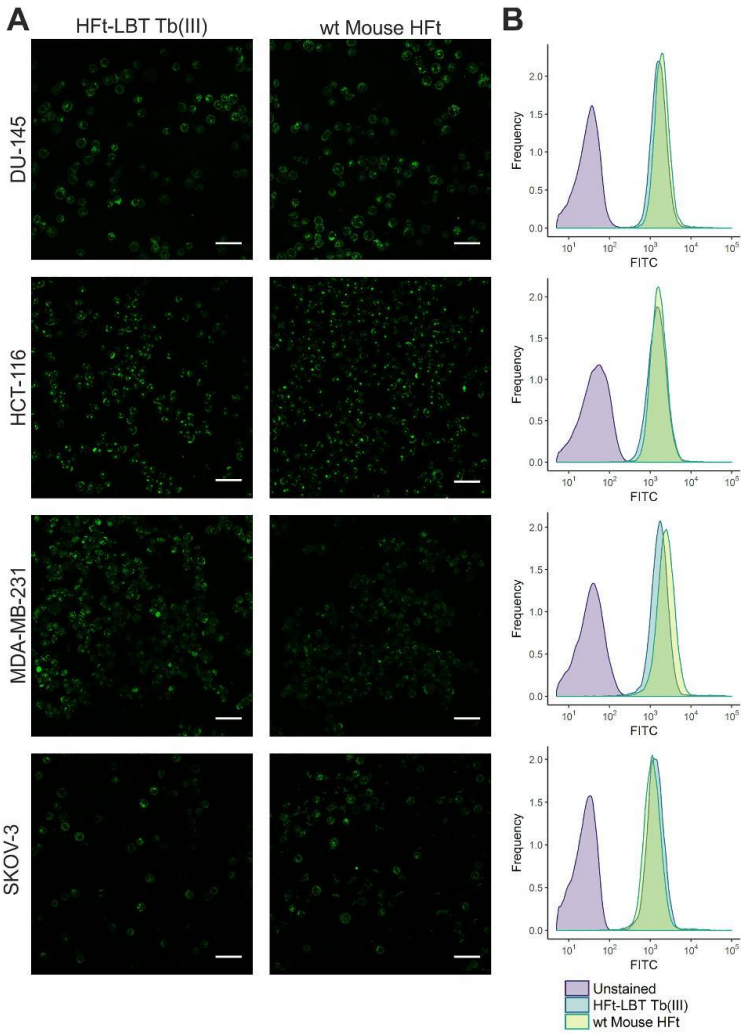

A
HFt-LBT Tb(III)
wt Mouse HFt
DU-145
HCT-116
MDA-MB-231
SKOV-3
B
Frequency
FITC
Unstained
HFt-LBT Tb(III)
wt Mouse HFt

Fig 5. Confocal microscopy images and flow cytometry analysis of HFt-LBT Tb(III) uptake by selected tumor cell lines. DU-145, HCT-116, MDA-MB-231 and SKOV-3 cancer cells were incubated with either HFt-LBT Tb(III) or wild type mouse HFt (0.5 mg/ml) for 60 min. A) Images acquired by confocal microscopy showing side by side comparison of cellular distribution of HFt-LBT Tb(III) and HFt conjugated with FITC. Scale bar = 50 μm. B) Flow cytometry analysis of HFt-LBT Tb(III) and HFt cellular uptake.

https://doi.org/10.1371/journal.pone.0201859.g005

in solution. In these measurements, excitation of the tryptophan residue present in the center of the LBT provided an excellent FRET exchange to the Terbium atom with a sustained narrow emission at the 544 nm line typical of Terbium excited state decay (Fig 1). The thermodynamics of the Terbium binding process was however complex due to the well-known presence of additional metal binding sites to the ferritin 24-mer [30–32]. In particular, the observed fluorescence signal accounted for 1.7 Tb(III) ions per subunit indicates that the presence of the tryptophan residue within the tag is able to act as an antenna system not only for the Tb(III) ion bound within the LBT tag but also for a number of extra Tb(III) atoms bound to the natural ferritin binding sites. Control experiments carried out by titrating wild type mouse HFt with Tb(III) demonstrated negligible fluorescence contributions in the Tb(III) emission regions (S1 Fig). Such finding, in agreement with previous reports [30–31] and with current X-ray crystallography findings, confirms that Tb(III) is bound to the canonical iron binding sites in the threefold channels and ferroxidase sites but is only capable of receiving extremely weak energy transfer from the only Trp residue (namely, W94) present in the wild type protein.

Analysis of the three-dimensional structure of the HFt-LBT Tb(III) complex both by X-ray crystallography and cryo-EM demonstrated that the presence of the C-terminal tag does not affect the overall assembly of the protein and that the genetically fused tags point to the interior cavity. The peptide arm connecting the lanthanide binding loop to the C-terminal sequence is, as expected, flexible around the GSG spacer peptide connecting the C-terminal end to the LBT sequence, and does not allow for a complete resolution of the local structure [55]. Interestingly, cryo-EM 3D map reconstruction allowed the identification of low resolution but definite L-shaped density patterns relative to the first few (i.e. 6–7 on a total of 17) aminoacids of the tag. However, the possible multiple orientations of the loop region precluded the observation of the Tb(III) complex. X-ray data further demonstrated that Terbium ion is efficiently complexed at the threefold axis channels (8 channels per 24-mer) by side chains of E131 and D134 aminoacids from each of the three adjacent subunits. Further Tb(III) binding occurs stoichiometrically at the ferroxidase site (24 sites per 24-mer), coordinated by carboxyl residues E27 and E62 with minor participation of H65. In contrast, LBT bound Tb(III) atoms were not visible neither by X-ray diffraction nor by cryo-EM analysis. In summary, HFt-LBT construct is capable of: *i)* high affinity binding of 24 Tb(III) atoms, one per each lanthanide binding tag, *ii)* intermediate affinity binding of 20 Tb(III) atoms at the ferroxidase binding site (out of 24 available, as demonstrated by 75% occupancy in X-ray structure), and *iii)* lower affinity binding of 8 Tb(III) atoms at the entrance of the threefold channels. Hence, within the whole 24-mer a total of 56 Terbium atoms can be hosted, if we consider full occupancy of the ferroxidase sites. A full thermodynamic description concerning the dissection of the Tb sites thus appears extremely complex at this stage and will be evaluated in future on selected mutants. In particular, single and double mutations on the ferroxidase site and/or on the residues lining the threefold channels key residues will be needed. In the present investigation on confocal microscope, however, attempts to image live cells by direct excitation at 290–375 nm of the tryptophan residue in the lanthanide binding loop after ferritin uptake yielded very poor results due to the substantial fluorescence background originating from the pool of cytoplasmic protein, within the typical Terbium emission interval. Clearly, a time resolved fluorescence microscopy approach or a red shifted antenna system will be needed to expand the

scope of the present work. We thus reported measurements in which HFt-LBT Tb(III) or wild type mouse HFt have been labeled with common FITC. Results demonstrate that the HFt-LBT Tb(III) complex is very efficiently uptaken by all four human tumor cell lines selected, in agreement with the identification of CD71 as the receptor of ferritin in humans [56]. In particular, DU145 (from a central nervous system metastasis, of primary prostate adenocarcinoma origin), MDA-MB-231 (from invasive ductal carcinoma), Hepatoma cell line and SK-OV-3 (highly resistant ovarian cancer cell line) were demonstrated for the first time to display a high uptake of both wild type mouseHFt and HFt-LBT Tb(III): this fact is remarkable, since these cancer cell lines overexpress CD71 receptor and are subject of cancer therapy studies focused on this receptor [54–60].

Conclusions

In conclusion, we have shown that a ferritin nanocage can be engineered by addition of appropriate metal binding tags inside the cavity in order to provide specific additional metal sites in topologically selected positions. In this framework, the key physical property of the tag, namely chelating antenna system for Terbium, can be coupled to the CD71 receptor recognition properties of the ferritin protein within a construct that can be rationally designed and manipulated. We propose this approach as an alternative to the *quasi* random metal clusters (or small organic fluorophore) insertion into the ferritin cavity that has been commonly used by free diffusion of metal ions (or organic molecules) through the open pores on the surface of the macromolecule or by disassembly-reassembly of the 24-mer structure ("encapsulation" procedure) [8]. The importance of a guided allocation of metal sites inside the cavity is considered essential in order to proceed with rational positioning of antenna systems close to lanthanide sites with the aim of designing the best geometry for efficient FRET. Rational design of such metal binding sites would foster more advanced applications such as the construction of up-converting nanoparticles or ultra-bright fluorescent organic polymers for single molecule detection. The present construct thus represents a first step towards the engineering of the ferritin cavity in order to generate a more efficient fluorescence/luminescence intracellular tracking system according to a structurally guided, rational design approach.

Supporting information

S1 Fig. SEC profile of HFt-LBT. Molecular weight determination by gel filtration was carried out by comparing the elution volume of HFt-LBT with the values obtained for several known calibration standards. HFt-LBT, loaded onto a HiLoad 26/600 Superdex 200 column, elutes as a single pick at 70 ml, the elution volume expected for the 24meric state of the protein, preceded by two small peaks corresponding to higher molecular weight aggregates (58 and 48 min). The inset shows the native page of HFt-LBT (lane 2) and HFt-LBT treated with of 1 mM TCEP (lane 3). Small amount of dimeric (dimers of 24-mers) or trimeric (trimer of 24-mers) can be detected both in the SEC profiles and in the corresponding native page. The higher MW multimers have been attributed to the presence of intermolecular disulfide bridges, as their presence is greatly decreased by addition of TCEP.
(TIF)

S2 Fig. SDS-PAGE of purified HFt-LBT. Line 1: Precision Plus Protein Unstained Protein Standards (BioRad); line 2: purified HFt-LBT.
(TIF)

S3 Fig. MALDI-TOF mass spectrum of HFt LBT. The identified molecular mass of 22531 Da corresponds to theoretical protein mass (22662 Da) subtracted of the contribution of a

methionine (131 Da).
(TIF)

S4 Fig. Representative 3D classes averages. Classes 3 and 5 contain noisy contributions; other classes are estimated high quality for 3D reconstructions obtained with RELION 2.0. Even at such low resolution, it is possible to notice non trivial structures inside the cavity.
(TIF)

S5 Fig. Fourier shell correlation (FSC) of the final 3D map reconstruction (see Fig 4). The resolution calculation is based on the 'gold-standard' protocol that assures the independence of the half-set reconstructions. Dashed purple line: FSC = 0.143. Final resolution: 7.1 Å.
(TIF)

S6 Fig. X-ray emission spectrum of HFt-LBT-Tb. L emission lines of terbium at 7.5, 8.2 and 8.5 keV are displayed. Additional lines at 6.3 and 7.0 keV are ascribed to iron ions.
(TIF)

S7 Fig. Anomalous difference map peaks of terbium. The structure is represented as cartoon, the residues surrounding the terbium ion are represented as sticks, terbium is represented as spheres and the map, contoured at 3 σ, is represented as a blue mesh. A) Terbium at the ferroxidase center B) Terbium at the 3-fold axes.
(TIF)

S8 Fig. Sample micrograph of HFt-LBT Tb(III). The sample was highly homogeneous and monodisperse. Scale bar = 100 nm.
(TIF)

S9 Fig. Representative FACS images: FITC labelled HFt-LBT Tb(III) and wt mouse HFt were loaded within the four cell lines: HCT, MDA-MB, DU-145 and S-KOV3. V500: violet-excitable dye engineered to improve brightness and reduce spectral overlap into the FITC channel.
(TIF)

S10 Fig. Effect of iron II addition to HFt-LBT Tb(III) adduct. Fluorescence spectra of HFt-LBT Tb(III) (red line) and HFt-LBT Tb(III) containing 10 μM ammonium iron(II) sulfate hexahydrate (blue line). Samples were at the same protein concentration (1 μM monomer). Spectra in the presence of iron were recorded after 2, 4, 8, 16 and 24 h at 25˚C in 0.1 M MES buffer pH 6.4 and did not show detectable time dependent changes. For clarity, only the spectrum after 24 h is shown. Addition of higher amount of iron resulted in hazyness and precipitation, most likely due to the formation of insoluble iron hydroxide precipitates.
(TIF)

S1 Table. Data collection and refinement statistics.
(PDF)

Acknowledgments

The authors are grateful to the ELETTRA XRD-1 beamline staff for their help and assistance during the data collection and fluorescence emission experiments. This study has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n˚ 312284. The biological part of the study was supported by the grant from National Science Centre (NCN) in Poland no. UMO-2015/18/E/NZ6/00642. EU H2020 Project "X-Probe", Grant N˚ 637295, to A. Bonamore and M.C.T. is gratefully acknowledged. Flagship project "Nanomax" from MIUR to A. Boffi is also acknowledged.

Author Contributions

Conceptualization: Alberto Boffi.

Data curation: Lorenzo Calisti, Matilde Cardoso Trabuco, Claudia Testi, Linda Celeste Montemiglio, Irene Benni, Andrea Ilari, Bartłomiej Taciak, Maciej Białasek, Paola Baiocco, Alessandra Bonamore.

Formal analysis: Claudia Testi, Linda Celeste Montemiglio, Amédée des Georges, Andrea Ilari, Bartłomiej Taciak, Maciej Białasek.

Funding acquisition: Alberto Boffi, Alessandra Bonamore.

Methodology: Matilde Cardoso Trabuco.

Project administration: Alberto Boffi.

Supervision: Alberto Boffi, Tomasz Rygiel, Magdalena Król, Alessandra Bonamore.

Validation: Matilde Cardoso Trabuco, Claudia Testi, Linda Celeste Montemiglio, Amédée des Georges, Andrea Ilari, Bartłomiej Taciak, Maciej Białasek, Paola Baiocco.

Writing – original draft: Lorenzo Calisti, Alberto Boffi, Alessandra Bonamore.

Writing – review & editing: Alberto Boffi, Claudia Testi, Linda Celeste Montemiglio, Amédée des Georges, Andrea Ilari, Bartłomiej Taciak, Maciej Białasek, Tomasz Rygiel, Magdalena Król, Paola Baiocco.

References

1. Arosio P, Elia L and Poli M, Ferritin, cellular iron storage and regulation *IUBMB Life*. 2017; 69(6) 414–422 https://doi.org/10.1002/iub.1621 PMID: 28349628
2. Theil EC, Tosha T and Behera RK, Solving Biology's Iron Chemistry Problem with Ferritin Protein Nanocages *Acc. Chem. Res*. 2016; 49 784–791
3. Calisti L, Benni I, Cardoso Trabuco M, Baiocco P, Ruzicka B, Boffi A, Falvo E, Malatesta F and Bonamore A, Probing bulky ligand entry in engineered archaeal ferritins *Biochim. Biophys. Acta*. 2017; 1861 (2) 450–456
4. Zhen Z, Tang W, Chen H, Lin X, Todd T, Wang G, Cowger T, Chen X and Xie, RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors *ACS Nano* 2013; 7 4830–4837 https://doi.org/10.1021/nn305791q PMID: 23718215
5. Kanekiyo M, Wei C-J, Yassine HM, Mc Tamney PM, Boyington JC, Whittle JR, Rao SS, Kong WP, Wang L and Nabel GJ, Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies *Nature* 2013; 499 102–106 https://doi.org/10.1038/nature12202 PMID: 23698367
6. Valero E, Fiorini S, Tambalo S, Busquier H, Callejas-Fernández J, Marzola P, Gálvez N and Domínguez-Vera JM, In vivo long-term magnetic resonance imaging activity of ferritin-based magnetic nanoparticles versus a standard contrast agent *J.Med.Chem*. 2014; 57 (13) 5686–5992
7. Fracasso G, Falvo E, Colotti G, Fazi F, Ingegnere T, Amalfitano A, Doglietto GB, Alfieri S, Boffi A, Morea V, Conti G, Tremante E, Giacomini P, Arcovito A and Ceci P, Selective delivery of doxorubicin by novel stimuli-sensitive nano-ferritins overcomes tumor refractoriness *J. Control Release* 2016; 239 10–18
8. Falvo E, Tremante E, Arcovito A, Papi M, Elad N, Boffi A, Morea V, Conti G, Toffoli G, Fracasso G, Giacomini P and Ceci P, Improved Doxorubicin Encapsulation and Pharmacokinetics of Ferritin-Fusion Protein Nanocarriers Bearing Proline, Serine, and Alanine Elements *Biomacromolecules* 2016; 17(2) 514–522 https://doi.org/10.1021/acs.biomac.5b01446 PMID: 26686226
9. Wang Z, Gao H, Zhang Y, Liu G, Functional ferritin nanoparticles for biomedical applications *Front Chem Sci Eng*.; 2017 11(4) 633–646 https://doi.org/10.1007/s11705-017-1620-8 PMID: 29503759
10. Fracasso G, Falvo E, Colotti G, Fazi F, Ingegnere T, Amalfitano A, Doglietto GB, Alfieri S, Boffi A, Morea V, Conti G, Tremante E, Giacomini P, Arcovito A, Ceci P, Selective delivery of doxorubicin by novel stimuli-sensitive nano-ferritins overcomes tumor refractoriness *J Control Release* 2016; 239 10–18. https://doi.org/10.1016/j.jconrel.2016.08.010 PMID: 27524282

11. Bain J and Staniland SS, Bioinspired nanoreactors for the biomineralisation of metallic-based nanoparticles for nanomedicine *Phys. Chem. Chem.Phys.* 2015; 17(24) 15508–15521
12. Tian L, Dai Z, Liu X, Song B, Ye Z and Yuan J, Ratiometric Time-Gated Luminescence Probe for Nitric Oxide Based on an Apoferritin-Assembled Lanthanide Complex-Rhodamine Luminescence Resonance Energy Transfer System *Anal. Chem.* 2015; 87(21) 10878–10885 https://doi.org/10.1021/acs.analchem.5b02347 PMID: 26462065
13. Maity B, Abe S and Ueno T, Observation of gold sub-nanocluster nucleation within a crystalline protein cage Nat.Commun. 2017; 8 14820 https://doi.org/10.1038/ncomms14820 PMID: 28300064
14. Handl HL and Gillies R, Lanthanide-based luminescent assays for ligand-receptor interactions *J. Life Sci.* 2005; 77 361–371 https://doi.org/10.1016/j.lfs.2005.01.009 PMID: 15894006
15. Hemmila I and Laitala VJ, Progress in lanthanides as luminescent probes *Fluoresc.* 2005; 15 529–542
16. Bunzli J, Benefiting from the unique properties of lanthanide ions *Acc. Chem. Res.* 2006; 39 53–61
17. Rajapakse HE, Reddy DR, Mohandessi S, Butlin NG and Miller LW, Luminescent terbium protein labels for time-resolved microscopy and screening *Angew. Chem. Int.* 2009; 48 4990–4992
18. Gudgin Dickson EF, Pollak A and Diamandis EP, Time-resolved detection of lanthanide luminescence for ultrasensitive bioanalytical assays *Photochem.Photobiol. B* 1995; 27(1) 3–19
19. Selvin PR, Principles and biophysical applications of lanthanide-based probes *Annu. Rev. Biophys. Biomembr.* 2002; 31 275–302
20. Bunzli JC, Luminescent lanthanide probes as diagnostic and therapeutic tools *Met. Ions. Biol. Syst.* 2004; 42 39–75
21. Franz KJ, Nitz M and Imperiali B, Lanthanide-binding tags as versatile protein coexpression probes *Chembiochem.* 2003; 4(4) 265–271 https://doi.org/10.1002/cbic.200390046 PMID: 12672105
22. Nitz M, Franz KJ, Maglathlin RL and Imperiali B, A powerful combinatorial screen to identify high-affinity terbium(III)-binding peptides *Chembiochem.* 2003; 4(4) 272–276 https://doi.org/10.1002/cbic.200390047 PMID: 12672106
23. Nitz M, Sherawat M, Franz KJ, Peisach E, Allen KN and Imperiali B, Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide *Angew. Chem. Int. Ed. Engl.* 2004; 43(28) 3682–3685
24. Martin LJ, Hähnke MJ, Nitz M, Wöhnert J, Nicholas R, Silvaggi, Allen KR, Schwalbe H, Imperiali B, Double-Lanthanide-Binding Tags: Design, Photophysical Properties, and NMR Applications *J. Am. Chem. Soc.* 2007; 129(22) 7106–7113
25. Adam W Barb, Tienhuei Grace Ho, Heather Flanagan-Steet, James H Prestegard, Lanthanide binding and IgG affinity construct: Potential applications in solution NMR, MRI, and luminescence microscopy *Protein Sci.* 2012; 21(10) 1456–1466 https://doi.org/10.1002/pro.2133 PMID: 22851279
26. Wöhnert J, Franz KJ, Nitz MI, Imperiali B and Schwalbe H, Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings *J. Am.Chem. Soc.* 2003; 125 13338–13339
27. Su XC, Huber T, Dixon NE and Otting G Site-specific labelling of proteins with a rigid lanthanide-binding tag *Chembiochem.* 2006; 71 599–1604
28. Goda N, Tenno T, Inomata K, Iwaya N, Sasaki Y, Shirakawa M and Hiroaki H, LBT/PTD dual tagged vector for purification, cellular protein delivery and visualization in living cells *Biochim.Biophys. Acta* 2007; 1773(2) 141–6 https://doi.org/10.1016/j.bbamcr.2006.11.016 PMID: 17207544
29. Sculimbrene BR and Imperiali B, Lanthanide-binding tags as luminescent probes for studying protein interactions *J. Am. Chem. Soc.* 2006; 128 7346–7352
30. Stefanini S, Chiancone E, Antonini E and Finazzi-Agro A, Binding of Terbium to apoferritin: a fluorescence study *Arch. Biochem. Biophys.* 1983; 222 (2) 430–434 PMID: 6847196
31. a) Treffry A and Harrison PM, Spectroscopic studies on the binding of iron, terbium, and zinc by apoferritin *J. Inorg. Biochem.* 1984; 21(1) 9–20 b) Bou-Abdallah F, Arosio P, Levi S, Janus-Chandler C and Chasteen ND, Defining metal ion inhibitor interactions with recombinant human H- and L-chain ferritins and site-directed variants: an isothermal titration calorimetry study *J. Biol. Inorg. Chem.* 2003; 8(4) 489–497 PMID: 6726222
32. de Turris V, Cardoso Trabuco M, Peruzzi G, Boffi A, Testi C, Vallone B, Montemiglio LC, De Georges A, Calisti L, Benni I, Bonamore A, Baiocco P, Humanized archaeal ferritin as a tool for cell targeted delivery, Nanoscale 2017; 9, 647–655 https://doi.org/10.1039/c6nr07129e PMID: 27942679
33. Kabsch W, Integration, scaling, space-group assignment and post-refinement *Acta Crystallogr. D Biol. Crystallogr.* 2010; 66, 133–44
34. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, Mc Nicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A

and Wilson KS, Overview of the CCP4 suite and current developments *Acta Crystallogr. D Biol. Crystallogr.* 2011; 67(4) 235–242

35. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC and Zwart PH, PHENIX: a comprehensive Python-based system for macromolecular structure solution *Acta Cryst. D* 2010; 66 213–221
36. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ, Phaser crystallographic software *J. Appl. Crystallogr.* 2007; 40 658–674 https://doi.org/10.1107/S0021889807021206 PMID: 19461840
37. Emsley P and Cowtan K, Coot: model-building tools for molecular graphics *ActaCrystallogr. D Biol Crystallogr.* 2004; 60 2126–2132
38. Laskowski RA, MacArthur MW, Moss DS and Thornton JM, PROCHECK—a program to check the stereochemical quality of protein structures *J. App. Cryst.* 1993; 26 283–29125
39. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC, MolProbity: all-atom structure validation for macromolecular crystallography *ActaCrystallogr. D Biol.Crystallogr.* 2010; 66 (1) 12–21
40. Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB and Minor W, Check My Metal: a macromolecular metal-binding validation tool *Acta Crystallogr. D Struct. Biol.* 2017; 73(3) 223–233
41. Russo CJ and Passmore LA, Ultrastable gold substrates for electron cryomicroscopy *Science* 2014; 346 1377–1380 https://doi.org/10.1126/science.1259530 PMID: 25504723
42. Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowall AW and Schultz P, Cryo-electron microscopy of vitrified specimens *Q. Rev. Biophys.* 1998; 21 129–228
43. Wagenknecht T, Grassucci R and Frank J, Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from Escherichia coli *J. Mol. Biol.* 1988; 199 137–147
44. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS and Carragher B, Automated molecular microscopy: the new Leginon system *Struct. Biol.* 2005; 151 41–60
45. Scheres SH, RELION: Implementation of a Bayesian approach to Cryo-EM structure determination J. Struct. *Biol.* 2012; 180 (3) 519–530 https://doi.org/10.1016/j.jsb.2012.09.006 PMID: 23000701
46. Zheng SQ, Palovcak E, Armache JP, Cheng Y and Agard DA, Anisotropic correction of beam-induced motion for improved single-particle electron Cryo-microscopy *Nature Methods* 2016; 14 331–332
47. Scheres SH, Semi-automated selection of cryo-EM particles in RELION-1.3 *J. Struct. Biol.* 2015; 189 114–122 https://doi.org/10.1016/j.jsb.2014.11.010 PMID: 25486611
48. Rosenthal PB and Henderson R, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy *J. Mol. Biol.* 2003; 333 721–745 PMID: 14568533
49. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SH and Henderson R, High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryo-microscopy *Ultramicroscopy* 2013; 135 24–35 https://doi.org/10.1016/j.ultramic.2013.06.004 PMID: 23872039
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM and Meng EC, UCSF Chimera—a visualization system for exploratory research and analysis *J. Comput. Chem.* 2004; 251605–1612 https://doi.org/10.1002/jcc.20084 PMID: 15264254
51. Schneider CA, Rasband WS and Eliceiri KW, Image to ImageJ: 25 years of image analysis *Nature Methods* 2012; 9 671–675 PMID: 22930834
52. Luzzago A, Cesareni G, Isolation of point mutations that affect the folding of the H chain of human ferritin in E.coli. *EMBO J.* 1989; 8(2) 569–76 PMID: 2656256
53. Habashy HO, Powe DG, Staka CM, Rakha EA, Ball G, Green AR, Aleskandarany M, Paish EC, Douglas Macmillan R, Nicholson RI, Ellis IO and Gee JM, Transferrin receptor (CD71) is a marker of poor prognosis in breast cancer and can predict response to tamoxifen *Breast Cancer Res Treat.* 2010; 119 (2) 283–93 https://doi.org/10.1007/s10549-009-0345-x PMID: 19238537
54. Horrocks WD Jr and Sudnick DR, Lanthanide ion probes of structure in biology. Laser-induced luminescence decay constants provide a direct measure of the number of metal-coordinated water molecules *J. Am. Chem. Soc.* 1979; 101 334–340
55. Barthelmes D, Granz M, Barthelmes K, Allen KN, Imperiali B, Prisner T and Schwalbe H, Encoded loop-lanthanide-binding tags for long-range distance measurements in proteins by NMR and EPR spectroscopy *J. Biomol. NMR* 2015; 63 275–282
56. Li L, Fang CJ, Ryan JC, Niemi EC, Lebrón JA, Bjorkman PJ, Arase H, Torti FM, Torti SV, Nakamura MC and Seaman WE, Binding and uptake of H-ferritin are mediated by human transferrin receptor-1.

Proc Natl Acad Sci U S A 2010; 107 3505–3510 https://doi.org/10.1073/pnas.0913192107 PMID: 20133674

57. Barth BM, Sharma R, Altinoğlu EI, Morgan TT, Shanmugavelandy SS, Kaiser JM, McGovern C, Matters GL, Smith JP, Kester M and Adair JH, Bioconjugation of calcium phosphosilicate composite nanoparticles for selective targeting of human breast and pancreatic cancers in vivo *ACS Nano*. 2010; 4 (3) 1279–87 https://doi.org/10.1021/nn901297q PMID: 20180585

58. Tang X, Liang Y, Zhu Y, Xie C, Yao A, Chen L, Jiang Q, Liu T, Wang X, Qian Y, Wei J, Ni W, Dai J, Jiang Z and Hou W, Anti-transferrin receptor-modified amphotericin B-loaded PLA-PEG nanoparticles cure Candidal meningitis and reduce drug toxicity *Int J Nanomedicine* 2015; 10 6 227–41 https://doi.org/10.2147/IJN.S88289

59. Gong Y, Gallis BM, Goodlett DR, Yang Y, Lu H, Lacoste E, Lai H and Sasaki T, Effects of transferrin conjugates of artemisinin and artemisinin dimer on breast cancer cell lines *Anticancer Res*. 2013; 33 (1) 123–32 PMID: 23267137

60. Sangar F, Schreurs AS, Umaña-Diaz C, Clapéron A, Desbois-Mouthon C,Calmel C, Mauger O, Zaanan A, Miquel C, Fléjou JF and Praz F, Involvement of small ArfGAP1 (SMAP1), a novel Arf6-specific GTPase-activating protein, in microsatellite instability oncogenesis *Oncogene* 2014; 33 (21) 758–67

Attachment IV

Cryo-EM structure of the human Ferritin-Transferrin Receptor 1 complex

Linda Celeste Montemiglio[a,c,h,*], Claudia Testi[b,*], Pierpaolo Ceci[c], Elisabetta Falvo[c], Martina Pitea[a], Carmelinda Savino[c], Alessandro Arcovito[d,e], Giovanna Peruzzi[b], Paola Baiocco[b], Filippo Mancia[f], Alberto Boffi[a], Amédée des Georges[g,**], Beatrice Vallone[a,h,**].

[a] Department of Biochemical Sciences "Alessandro Rossi Fanelli", Sapienza University of Rome, P.le A. Moro 5, 00185 Rome, Italy.

[b] Center for Life Nano Science @ Sapienza, Istituto Italiano di Tecnologia, V.le Regina Elena 291, 00161 Rome, Italy.

[c] Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro 5, 00185 Rome, Italy.

[d] Istituto di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Largo F. Vito 1, 00168 Rome, Italy

[e] Fondazione Policlinico Universitario Agostino Gemelli- IRCCS , Largo F. Vito 1, 00168 Rome, Italy

[f] Department of Physiology and Cellular Biophysics, Russ Berrie Pavilion, Columbia University Medical Center, 1150 St Nicholas Ave, 10032 New York, NY, USA

[g] Structural Biology Initiative, Advanced Science Research Center, CUNY, 85 St Nicholas Terrace, 10031 New York, NY, USA

[h] Istituto Pasteur-Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, P. le A. Moro 5, 00185 Rome, Italy

Keywords: Transferrin receptor; Ferritin; Cryo electron microscopy; single-particle analysis.

* These authors contributed equally to this work

** To whom correspondence should be addressed: beatrice.vallone@uniroma1.it and amedee.desgeorges@asrc.cuny.edu

Abstract

Human transferrin receptor 1 (CD71) guarantees an adequate supply of iron to mammalian cells. This is accomplished by ligand-induced activation of the cellular endocytic machinery upon binding of iron-loaded transferrin and ferritin to CD71, resulting in internalization of the respective complexes [1,2]. CD71 is also exploited by viruses [3-7] and by the malaria parasite [8] as a route for cell invasion. The structural features of CD71 interaction with transferrin and pathogenic hosts have been extensively studied and have been clarified recently by the identification of the respective specific recognition epitopes [9-11]. Here, we provide the molecular basis of the interaction between CD71 and human ferritin by reporting the structure of the complex at 3.9 Å resolution obtained by single-particle cryo-electron microscopy, and by validating our structural findings in cellular context. The contact surfaces between the H-chain ferritin homopolymer and the CD71 epitope largely overlap with the arenaviruses and *Plasmodium vivax* binding regions in the apical part of the receptor ectodomain. Our structural data account for the transferrin-independent binding of ferritin to CD71 and suggest that select pathogens may have adapted to enter cells by mimicking the ferritin access gate.

Main text

Human transferrin receptor 1 (CD71 or hTfR1) is a promiscuous and ubiquitously expressed cell entry carrier whose primary function is to import iron in response to variations in intracellular

concentration of this essential element. Iron uptake is mediated by the internalization of the transferrin-iron complex through receptor-mediated endocytosis *via* a clathrin-dependent pathway [1]. Once the iron cargo is delivered, the receptor is recycled back to the cell surface and apo-transferrin is released into the bloodstream [2]. CD71 has been also shown to mediate the uptake of heavy chain ferritin (H-Ft) from serum as an alternative or additional source of bioavailable iron [12]. CD71 is also a preferred entry carrier for human pathogenic arenaviruses [5,7,11,13] and hepatitis C virus [6] as well as feline and canine specific parvoviruses [4]. Viral systems recognize epitopes on the host-encoded CD71 receptor through their surface spike glycoproteins, triggering the internalization of the complex. Recently, *Plasmodium vivax*, the most common malaria parasite, was demonstrated to access reticulocyte cytoplasm by recognizing the same CD71 receptor epitope as arenaviruses [8,10].

CD71 is a homodimeric type II transmembrane protein composed of a small cytoplasmic domain, a single-pass transmembrane region, and a complex extracellular domain. Each monomer of the ectodomain is subdivided in a protease-like domain in contact with the cell membrane, a helical domain comprising the dimer contact regions, and an apical domain (**Fig. 1a**) [14]. The ectodomain displays ligand binding sites for diverse proteins: its basal portion (formed by the protease-like and the helical domains) binds transferrin (Tf) [9,15] and the dimer interface region forms a complex with the Hereditary Hemochromatosis factor (HFE) [16] (**Fig. 1b**). The upper part of the apical domain has been shown to interact with arenaviruses [11,13] and with the *Plasmodium vivax* invasion protein *Pv*RBP2b [10] (**Fig. 1b**).

In this framework, a key missing piece of information concerns the structural basis of the interaction between CD71 and H-Ft. Experimental evidence was provided for a scarce competition between ferritin and transferrin for CD71 binding, thus pointing out the possibility of the existence of different epitopes for the two protein ligands [12,17]. Recently, an exposed loop region in the H-Ft subunit was identified that, transplanted in an archaeal ferritin, originally unable to recognize the human CD71 receptor, was sufficient to induce binding of this chimeric protein to the receptor [18].

The importance of the CD71/H-Ft interaction is dictated by the emerging physiological and pathological significance of the circulating ferritin and its scavenger receptor [19,20]. Moreover, nano-sized H-Ft homopolymers have moved to the center stage of nanomedicine research as theranostic agents [21], due to their unique cargo capabilities for small therapeutic molecules or isotopic tracers coupled to selectivity towards CD71 [22,23]. CD71 is overexpressed in the most common cancer cell types, further highlighting the interest for this receptor as a privileged target for the selective delivery of cytotoxic drugs coupled to transferrin, ferritin or monoclonal antibody drug conjugates [24-26].

We used single-particle cryo-electron microscopy to solve the structure of H-chain ferritin bound to human CD71 to 3.9 Å resolution, unveiling the structural determinants that govern their recognition (**Fig. 2, Extended Data Fig. 1-2-3 and Table 1**). H-Ft binds the CD71 receptor in a "virus-like" fashion, covering an overall area of approximately 1900 $Å^2$. As depicted in Fig. 2, CD71 interacts with H-Ft through four specific contact regions on the apical domain that comprise: *i)* a motif of six amino acids, from R208 to L212 and N215 on the βII-2 strand; *ii)* residues E343, K344 and N348 on the αII-2 helix. We refer to these residues as "common contacts" on CD71, since they represent the key structural determinants for binding of arenavirus (MACV) and plasmodial *Pv*RBP2b proteins. Additionally, "exclusive contact" regions on CD71 for H-Ft comprise *iii)* the βI-1-βII-1 loop and the βII-1 strand (S195, E197, S199 and I202) and *iv)* the βII-8 strand (K374). The H-Ft binding counterpart regions are: *i)* the external BC loop (R79, F81, Q83, K86, K87); *ii)* the N-terminus of the A-helix (Q14, D15, E17-A19, N21, R22, N25); *iii)* the C-terminus of the C helix (E116, K119, D123) (**Fig. 2**).

The H-Ft "exclusive contacts" pairwise interactions are shown in Fig. 2b and 2c, where: *i)* CD71 βII-1 strand contacts H-Ft *via* the A helix; *ii)* βI-1-βII-1 loop interacts with the H-Ft C helix; *iii)* K374 on βII-8 of CD71 electrostatically interacts with Q14 at the beginning of H-Ft A helix.

The "common contacts" with CD71, shared between pathogens and H-Ft, are shown in Fig. 2d and are established by the βII-2 strand of CD71 with the H-Ft A helix, BC loop and C helix and

with T5 at the H-Ft N-Terminus. Further common interactions are established between the CD71 αII-2 and H-Ft A helix, the BC loop and K87 on the BC loop of a flanking H-Ft monomer (Fig. 2e; the detailed contacts are given in Extended Data Table 2).

Notably, the amino acids involved in "common contacts" coincide with amino acids on CD71 that lead to gain or loss of interaction with pathogen binding proteins upon mutations [10,27]. These cluster in the βII-2 and αII-2 regions, which appear to be essential on CD71 for binding with various partners. Therefore, in order to identify the determinants on H-Ft for binding to CD71 we selected and mutated H-Ft residues involved in "common contacts" which are not conserved in ferritins shown to be unable to bind human CD71, *i.e.* human Light-chain ferritin [12] (Hum L-chain) and *Archeoglobus fulgidus* ferritin [18] (AfFt) (**Extended Data Fig. 4**). We produced three multiple mutants of residues peculiar of human H-Ft: *i)* mutant A lacking the polar residues at the N-terminal of the A helix (Q14A/D15A/R22A), *ii)* mutant B lacking F81 and Q83 on the external BC-loop (F81A/Q83A), and *iii)* mutant C that combines A and B mutations (Q14A/D15A/R22A/F81A/Q83A) (**Extended Data Fig. 4**). Surface Plasmon Resonance (SPR) measurements using wild-type or mutant H-Fts as analytes and CD71 as ligand showed that the binding affinity for the receptor is reduced of about two orders of magnitude in mutants A and B as compared to the wild type, and abolished in mutant C (**Fig. 3a**). In particular a critical drop of the k_{on} value is increased across mutants B and C, suggesting that mutations at the BC loop have a dominant role in impairing the CD71/H-Ft interaction, likely due to the loss of contact of F81 and Q83 with Y211 on CD71 βII-2 and Q83 with N348 on CD71 αII-2. Consistently, FACS (**Fig. 3b, Extended Data Fig. 5**) and confocal microscopy (**Fig. 3c**) measurements on HeLa cells uptake of the three H-Ft mutants show reduced (mutant A) or negligible (mutants B and C) internalization with respect to the wild type.

In summary, these results assess the critical relevance of ferritin external BC-loop in CD71 binding, recognizing the "common contact" residues Y211 and N348, crucial for binding of viruses

and plasmodial proteins, which mutations severely hamper binding and suppress internalization, however they also highlight the relevance of Q14, D15 and R22 at the N-terminus of the A helix.

Therefore, the significant binding capability of the Humanized *Archeoglobus fulgidus* ferritin (Hum-AfFt) where the human H-Ft BC loop had been transplanted, is likely due not only to the presence of F81 and Q83 on the loop, but also to R22 and D15 on the A helix, serendipitously present in this archea ferritin. On the other hand, the Hum L-Ft which lacks F81, Q14 and R22 is unable to bind the CD71 receptor [12] **(Extended Data Fig. 4)**.

In conclusion, we identified the specific sites on CD71 to be hooked by ferritin for physiological access to cell through the CD71 "iron door". We provide structural evidence that transferrin and ferritin exploit alternative epitopes on the same receptor, allowing differential regulation of iron, as in the case of the HFE [12,16].

Notably, pathogens have adapted to exploit the ferritin epitope to unlock cellular barriers by mimicking this physiological interaction with the CD71 apical domain. In this framework, changes due to Single Nucleotide Polimorphisms (SNPs) within the CD71 apical domain may account for permissivity of virus or parasite entry, but cannot be considered neutral with respect to ferritin uptake. Along this line, CD71 species-specific variants must conveniently be matched to H-Ft co-evolved variants in order to conserve the physiological functions of serum circulating ferritin.

Our work provides a sound structural basis to elaborate on the possibility of developing alternative ferritin-like anti-viral or anti-parasite therapeutic ligand, be it an antibody or a peptidomimetic capable of blocking the "common contacts" epitope on CD71 residue, and to further engineering ferritins as theranostic agents.

Methods

Cloning, expression and purification of Human H-Ft and CD71.

The genes encoding the human apo-H-chain ferritin (H-Ft) and mutants A (Q14A/D15A/R22A) and B (F81A/Q83A) were designed, synthesized, optimized for *Escherichia coli* codon usage and cloned in pET11b vector by Geneart (Geneart AG). Mutant C was produced by using the Quick Change Lightning Kit (Agilent Technologies, Santa Clara, California) according to the manufacturer's instructions and using mutant A as a template. All apo-H-Ft variants were expressed and purified as previously reported [25].

The gene encoding the ectodomain of human CD71 (residues 121-760) was extracted by pcr from the plasmid pAcGP67A-TfR [28] (Addgene, Cambridge, MA) and BamHI/EcoRI inserted using the Gibson cloning method and fused at the 3' of the Kozak sequence of the pα−H mammalian expression vector modified by the addition of the hydrophobic leader peptide from the baculovirus protein gp67 (pα−H BiP). An octa-histidine tag was also placed at the C-terminus of the CD71 gene. The expression plasmid pα−H BiP/TfR1 was transiently transfected in HEK 293 using polyethylenimine (PEI) as transfection agent. Cells were grown in FreeStyle 293 expression media (ThermoFisher Scientific, Hampton, USA) supplemented with 1% of fetal bovine serum (FBS) at 37°C in a humidified atmosphere of 5% CO_2. After 96 hours, cells were harvested and CD71 was purified from supernatants using Ni- or Co-affinity chromatography. Supernatant was filtered and incubated with the resin after addition of 50 mM sodium phosphate buffer, pH 8.0, together with 200 mM NaCl and 20 mM imidazole; 250 mM imidazole was used to elute CD71. The protein sample was stored at -80°C in 50 mM sodium phosphate, 200 mM NaCl, pH 8. Quality and quantity of purified protein was evaluated by SDS-PAGE and UV/visible spectra using the theoretical $\varepsilon_{280\,nm}$ 96260 $M^{-1}cm^{-1}$.

CD71/H-Ft complex preparation.

In vitro, incubation of H-Ft and CD71 at different stoichiometric ratios results in protein aggregation due to the presence of multiple binding sites on both H-Ft and CD71, forming insoluble precipitates. We managed to isolate a soluble sample of CD71/H-Ft complex by means of a pull-down experiment [12]. 250 μg 8xHis-tag fused CD71 was incubated with 100 μL of TALON resin (TALON Superflow Metal Affinity Resin, Ge Healthcare, UK) in 25 mM Tris·HCl, 150 mM NaCl, 1% PEG 8000 and 10 mM Imidazole, pH 7.2 (buffer A), for 60 minutes at 4°C, under rotation. After several washes with buffer A, CD71-conjugated beads were incubated with 950 μg of H-Ft for 90 minutes at 4°C under rotation. The beads were washed increasing imidazole concentration in buffer A up to 30 mM, and the complex was eluted in 350 μL using 290 mM imidazole in buffer A. As a control experiment, CD71-unconjugated beads were also incubated with 950 μg of H-Ft for 90 minutes, following the same procedure adopted to isolate the complex. The pull-down assay final samples were analyzed by SDS/PAGE (**Extended Data Fig. 1a**).

Grids preparation for cryo-electron microscopy.

Purified CD71/H-Ft complex eluted from Talon resin at 0.2 mg/mL was directly used for grid preparation and two datasets were collected using the same batch of grids to obtain the final map. Holey-gold R0.9/1 (dataset 1) and holey-carbon R1.2/1.3 (dataset 2) grids (Quantifoil Micro Tools GmbH) covered by 2 nm film of carbon were prepared as described [29]. Grid surfaces were treated with plasma cleaning using a mixture of Ar and O_2 for 60 s before applying 3 μl of sample. The screening of several blotting conditions revealed that the time between sample application to the grid and plunge into ethane affects the number of particles per field and their distribution. The dataset 1 grid was prepared with 135 seconds wait time and the dataset 2 grid with a 90 second wait time. Grids were then blotted for 1 second (100% humidity, 4°C, force 4) with filter paper and vitrified by rapidly plunging into liquid ethane at −180 °C using a Vitrobot Mark IV (FEI, Hillsboro) [30].

Data collection.

The first dataset (dataset 1, 690 micrographs, Extended Data Fig. 3) was imaged using a FEI Titan Halo (ThermoFisher Scientific, Eindhoven) operating at 300 kV acceleration voltage, while the specimen was maintained at liquid nitrogen temperature using a Gatan 626 side entry cryo-holder (Gatan, Pleasanton). Images were recorded using the automated acquisition program Leginon [31]. We used a Gatan K2 Summit direct-detector camera (Gatan, Pleasanton) operating in counting mode, with a calibrated pixel size of 1.15 Å on the object scale. Images were typically recorded with a defocus range between −0.7 and −3.0 µm. Movies were acquired with a total exposure time of 12 sec (60 frames/image), with exposure rate of 7.8 electrons/pix/sec.

A second dataset (dataset 2, 573 micrographs, **Extended Data Fig. 3**) was imaged using a 300 kV Titan Krios (ThermoFisher Scientific, Eindhoven). The dataset was collected automatically using EPU (ThermoFisher Scientific, Eindhoven). Images were recorded on a Gatan K2 Summit direct-detector camera (Gatan, Pleasanton) equipped with a Gatan Bioquantum LS/967 energy filter and operating in super-resolution mode, using a calibrated pixel size of 1.33 Å on the object scale. Images were typically recorded with a defocus range between −1.0 and −3.0 µm. Movies were acquired in electron counting mode, the total exposure time was set to 12 sec (40 frames/image), with exposure rate of 6.2 e^-/pix/sec.

Image processing.

The main steps of data analysis are schematized in the workflow shown in **Extended Data Figure 3**.

Micrograph frames collected in both datasets (**Extended Data Fig. 1b**) were aligned for beam-induced motion correction and drift with MotionCor2 [32]; global contrast transfer function was calculated using Gctf [33]. Micrographs with resolution limits ≤ 6Å (dataset 1) or ≤ 5Å (dataset 2) were kept for further processing.

All subsequent data analysis was carried out using RELION 2.0 [34]. More than 1100 particles for both datasets were manually picked to produce a reference for the automated particle picking procedure implemented in RELION [35]. A total of 140,567 and 194,501 particles were automatically

picked respectively from dataset 1 and dataset 2 and extracted from the original micrographs. After the extraction, particles were initially classified in 2D using K=100 classes. A second round of 2D classification was performed on each dataset with K= 25 classes. The best 2D class averages clearly showed the CD71/H-Ft complex in different orientations, revealing some extent of sample heterogeneity due to the presence of alternative populations endowed with different stoichiometry (**Extended Data Fig. 1c**). A set of 27,690 particles, belonging to good 2D classes of dataset 1, was selected and subjected to a first round of 3D classification, using 8 classes and without imposing any symmetry; the cryo-EM apo-ferritin map (EMDB code EMD-2788 [29]) was used as reference model, filtered at 60 Å. This procedure gave only one class (8,860 particles) where a 1:1=H-Ft:CD71 complex was clearly displayed. The resulting map was used as reference (filtered at 60 Å) to run a second round of 3D classification, which resulted in two better-resolved classes (total 17,370 particles). These classes (see **Extended Data Fig. 3**) were selected for further refinement using the 3D Autorefine procedure, applying a spherical mask of 290 Å diameter. This resulted in a map at 8.2 Å resolution.

The reconstructed map of dataset 1 (filtered at 40 Å) was used as reference model for the 3D classification of particles selected from 2D classification of dataset 2 (73,700 particles). Good classes (**Extended Data Fig. 3**) were selected for further refinement using the 3D Autorefine procedure (25,870 particles, spherical mask of 290 Å diameter), which yielded a map at 5.5 Å resolution.

With the aim to further increase the resolution of the contacting region between the two molecules, particles used for the 3D Refinement of both datasets were joined (total 43,240 particles) and subjected to one round of 3D Refinement, imposing O symmetry (*i.e.*, ferritin point-group symmetry) and applying a smaller mask diameter of 180 Å, to only include the ferritin molecule. This yielded a 4.8 Å map of ferritin re-centered to its center of symmetry (**Extended Data Fig. 3**). We used the corresponding particle alignment parameters to perform symmetry expansion of this dataset using RELION (*relion_particle_symmetry_expand*). This allowed us to subdivide ferritin

particles in individual subunits in order to identify by classification all ferritin subunits bound to a transferrin receptor [36]. The dataset was thus artificially expanded according to the pseudo-symmetric O point group and enlarged 24-fold, resulting in 1,037,760 particles, that were 3D Classified using C1 symmetry, no image alignment and a mask generated from the complex with a single receptor bound. This classification round identified all subunits bearing a receptor bound. All four classes generated were subjected to a second round of 3D Classification, this time with local image alignment, and only two of them allowed the identification of subclasses of ferritin bound to the receptor. Therefore, three H-Ft bound subclasses were combined (total 53,878 particles) and refined, resulting in a map with a global resolution of 4.4 Å. Resolution measured with a mask only including ferritin and the receptor contact sites was estimated to be 4.2 Å resolution.

The refined particles were exported into cisTEM to perform per-particle defocus refinement procedure implemented in cisTEM [37]. The resulting final map resolution of H-Ft and CD71 contact region was improved up to 3.9 Å, clearly showing secondary structures and bulky side chains of interacting residues, which allowed more precise model building. Statistic information of the 5.5 Å and of the 3.9 Å density maps is reported in Extended Data Table 1.

Resolution estimation.

The overall resolution of the two maps at 5.5 Å and 4.2 Å obtained with RELION and the one at 3.9 Å obtained with cisTEM (**Extended Data Fig. 2a, 2c and 2e-f**) was estimated with the Fourier shell correlation (FSC) = 0.143 criterion, based on the 'gold-standard' protocol [38] using a mask around the complex density. The input map was corrected for the modulation transfer function (MTF) of K2 detector and sharpened using negative temperature B factors as calculated by RELION and cisTEM (**Extended Data Table 1**). Local resolution was estimated for the map at 5.5 Å and 4.2 Å using ResMap [39] (**Extended Data Fig. 2b and 2d**). The electron density maps obtained at 5.5 Å and at 3.9 Å resolution were sharpened using autosharpen in Phenix [40]. UCSF Chimera[41] and PyMOL (http://pymol.sourceforge.net) were used for graphical visualizations.

Model building and refinement.

The program Chimera was initially employed to rigid body fit the crystal structures of H-Ft (PDB code 3AJO [42]) and CD71 (PDB code 3KAS [11]) into the sharpened electron density map obtained at 5.5 Å resolution, which was further refined as rigid body with Phenix *real_space_refinement*. In this map, both CD71 and H-Ft were visible at the level of their secondary structures. The resulting CD71/H-Ft model was then rigid body fitted into the sharpened 3.9 Å density map using UCSF Chimera. Model building of H-Ft and of the CD71 contact region was performed manually using Coot [43]. Given the high quality of the reconstructed map, the side chains of residues exposed at the contacting interface were modeled into the electron density (**Extended Data Fig. 2g**).

The model of the CD71/H-Ft complex was refined using the *real_space_refinement* routine in minimization global mode (Phenix) against the overall map at 3.9 Å, imposing secondary structure and Ramachandran restraints. Final visual inspection was performed in COOT to manually correct Ramachandran outliers. The final model was validated using MolProbity [44] and EMRinger [45] (**Extended Data Table 1**). All figures were produced using Pymol and UCSF Chimera.

CD71/H-Ft SPR assay.

The interactions between the immobilized N-terminal His-tagged transferrin receptor CD71 (ligand) and H-Ft-based constructs (analytes) were measured by surface plasmon resonance (SPR) technique on a Biacore X100 instrument (Biacore, Uppsala, Sweden) according to the procedure previously reported [26]. Briefly, CD71 was immobilized on a Sensor Chip nitrilotriacetic acid (NTA) (GE Healthcare Europe GmbH) according to the manufacturer's instructions. The optimal experimental setup was determined and CD71 was injected at 22 μg/ml for 60 s. Analyte concentration was in the range of 1000–62.5 μg/ml. The sensor chip surface was regenerated using fresh histidine tagged protein after every cycle of the assay. The SPR assay was performed at 25 °C, at flow rate= 30 μl/min; the association and dissociation phases were monitored for 180 and 600 s, respectively. Analytes were dissolved in degassed 10 mM Phosphate-buffered saline (PBS) at pH 7.4. In this condition H-Ft retains the 24-mer assembly. To regenerate the chip, complete dissociation of the

complexes was achieved by the addition of 10 mM HEPES, 150 mM NaCl, 350 mM EDTA and 0.005% (vol/vol) surfactant P20 (pH 8.3) for 30 sec before starting a new cycle. The k_{on} and k_{off} rates as well as the dissociation constant (K_D) were estimated using the Biacore X100 evaluation software according to a 1:1 binding model or alternatively a heterogeneous analyte binding model (**Extended Data Tables 3 and 4**). All experimental data shown in **Figure 3a** are reported at analyte concentration of 1 mg/ml. Fits are reported as black lines corresponding to a heterogeneous analyte binding model for wild type, Mutant A and Mutant B respectively, and to a simple 1:1 kinetic model for Mutant C.

Protein FITC labeling.

HFt, AfFt, Mutant A, Mutant B and Mutant C were labeled with fluorescein-isothiocyanide (FITC, ThermoFisher) according to the manufacturer's standard protocol. Briefly, 2 mg/mL of the purified protein was added with 10-fold molar excess of FITC in protein storage buffer, stirring for 2 h at RT. The non-reacted dye was removed by gel filtration chromatography and the fluoresourscent dye to protein ratio was determined by UV-spectroscopy. LC-MS spectrometry measurements all FITC-conjugated proteins confirmed that > 40% of monomers are FITC labeled.

Cell cultures and ferritins internalization.

HeLa cells were grown at 37 °C in Dulbecco's Modified Eagle's Medium with Glutamax (DMEM, Gibco) supplemented with 10% FBS (Gibco) and 1% Penstrep (100 U/ml penicillin and 100 µg/ml streptomycin solution; Gibco). DMEM without phenol red (Sigma) supplemented with Glutamax (Invitrogen), 10% FBS and 1% Penstrep (incubation medium) was used for apo-ferritins internalization assays by FACS and confocal microscopy.

Flow cytometry analysis.

HeLa cells were seeded on 6-well plates and left one day prior performing FACS experiments. Upon growing medium removal and rinse with PBS, confluent cells were incubated with FITC-ferritin nanoparticles (AfFt, HFt, Mutant A, B and C as specified in each experiment) at the final concentration of 30 µg/ml for 2 hours 30 minutes. Cell were then washed twice with PBS, detached

with trypsin-EDTA (Euroclone), rinsed with PBS and resuspended in BD-FACS flow buffer. Control cells were treated in the same conditions without ferritins. Internalization of ferritins was evaluated with sample acquisition at the BD LSRFortessa (BD Biosciences, San Jose, CA, USA) equipped with a 488 nm laser and FACSDiva software (BD Biosciences version 6.1.3). Live cells were first gated by forward and side scatter area (FSC-A and SSC-A) plots, then detected in the green channel for FITC expression (530/30 nm filter) and side scatter parameter. The gate for the specific signal was set based on the control sample. Data were analyzed using FlowJo 9.3.4 software (Tree Star, Ashland, OR, USA).

Confocal microscopy of live cells.

For apo-ferritins internalization by live imaging on a confocal microscope, HeLa cells were seeded on a μ-slide 8-well ibiTreat (ibidi) and left one day to grow. After the removal of the medium, cells were washed with PBS and incubated for 20 hours with 30 μg/ml FITC-ferritin nanoparticles (H-Ft, Mutant A, B and C). Prior imaging, cells were washed twice with PBS to eliminate the unbound FITC-ferritins and then replaced in the incubation medium (suitable for confocal imaging purposes). The confocal laser-scanning microscope used was an Olympus FV10i platform equipped with a built-in incubator. Images were acquired with a 60×/1.2NA water-immersion objective, LD lasers and filter sets for FITC. FITC and phase-contrast channels were acquired simultaneously.

Data availability.

All data supporting the findings of this study are available within this paper. The cryo-EM maps of CD71/H-Ft complex at 5.5 Å and at 3.9 Å and coordinates generated and analyzed in the current study have been deposited in the Electron Microscopy Data Bank and in the Protein Data Bank under accession code EMD-0046 (PDB 6GSR) and EMD-0140 (PDB 6H5I), respectively.

References

1 Gammella, E., Buratti, P., Cairo, G. & Recalcati, S. The transferrin receptor: the cellular iron gate. *Metallomics* **9(10)**, 1367-1375 (2017).

2 Mayle, K. M., Le, A. M. & Kamei, D. T. The intracellular trafficking pathway of transferrin. *Biochim Biophys Acta* **1820(3)**, 264-281 (2012).

3 Demogines, A., Abraham, J., Choe, H., Farzan, M., Sawyer, S.L. Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol* **11(5)**, e1001571 (2013).

4 Hafenstein, S. *et al.* Asymmetric binding of transferrin receptor to parvovirus capsids. *Proc Natl Acad Sci U S A* **104(16)**, 6585-6589 (2007).

5 Kerr, S. A. *et al.* Computational and Functional Analysis of the Virus-Receptor Interface Reveals Host Range Trade-Offs in New World Arenaviruses. *J Virol* **89(22)**, 11643-11653 (2015).

6 Martin, D. N. & Uprichard, S. L. Identification of transferrin receptor 1 as a hepatitis C virus entry factor. *Proc Natl Acad Sci U S A* **110(26)**, 10777-10782 (2013).

7 Radoshitzky, S. R. *et al.* Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature* **446**, 92–96 (2007).

8 Gruszczyk, J., Kanjee, U., Chan, L.J., Menant, S., Malleret, B., Lim, N.T.Y., Schmidt, C.Q., Mok, Y.F., Lin, K.M., Pearson, R.D., Rangel, G., Smith, B.J., Call, M.J., Weekes, M.P., Griffin, M.D.W., Murphy, J.M., Abraham, J., Sriprawat, K., Menezes, M.J., Ferreira, M.U., Russell, B., Renia, L., Duraisingh, M.T., Tham, W.H.,. Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax. *Science* **359(6371)**, 48-55 (2018).

9 Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. & Walz, T. Structure of the human transferrin receptor-transferrin complex. *Cell* **116(4)**, 565-576 (2004).

10 Gruszczyk, J., Huang, R.K., Chan, L.J., Menant, S., Hong, C., Murphy, J.M., Mok, Y.F., Griffin, M.D.W., Pearson, R.D., Wong, W., Cowman, A.F., Yu, Z., Tham, W.H. Cryo-EM

structure of an essential Plasmodium vivax invasion complex. *Nature* **559(7712)**, 135-139 (2018).

11 Radoshitzky, S. R. *et al.* Machupo virus glycoprotein determinants for human transferrin receptor 1 binding and cell entry. *Plos One* **6**, e21398 (2011).

12 Li, L. *et al.* Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proc Natl Acad Sci U S A* **107**, 3505-3510, doi:10.1073/pnas.0913192107 (2010).

13 Abraham, J., Corbett, K.D., Farzan, M., Choe, H., Harrison, S.C. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat Struct Mol Biol* **17**, 438–444 (2010).

14 Lawrence, C. M., Ray, S., Babyonyshev, M., Galluser, R., Borhani, D.W., Harrison, S.C. Crystal structure of the ectodomain of human transferrin receptor. *Science* **286(5440)**, 779-782 (1999).

15 Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. & Walz, T. Single particle reconstruction of the human apo-transferrin-transferrin receptor complex. *J Struct Biol* **152(3)**, 204-210 (2005).

16 Lebrón, J. A. *et al.* Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with Transferrin Receptor. *Cell* **93**, 111-123, doi:10.1016/S0092-8674(00)81151-4 (1998).

17 Sakamoto, S., Kawabata ,H., Masuda, T., Uchiyama, T., Mizumoto, C., Ohmori, K., Koeffler, H.P., Kadowaki, N., Takaori-Kondo, 2. H-Ferritin Is preferentially incorporated by human erythroid cells through transferrin receptor 1 in a threshold-dependent manner. *Plos One* **10(10)**, e0139915 (2015).

18 de Turris, V. *et al.* Humanized archaeal ferritin as a tool for cell targeted delivery. *Nanoscale* **9(2)**, 647-655 (2017).

19 Greene, C. J. *et al.* Transferrin receptor 1 upregulation in primary tumor and downregulation in benign kidney is associated with progression and mortality in renal cell carcinoma patients. *Oncotarget* **8(63)**, 107052-107075 (2017).

20 Rosager, A. M. *et al.* Transferrin receptor-1 and ferritin heavy and light chains in astrocytic brain tumors: Expression and prognostic value. *Plos One* **12(8)**, e0182954 (2017).

21 Truffi, M. *et al.* Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacol Res* **107**, 57-65 (2016).

22 Abe, S., Maity, B., Ueno, T. Design of a confined environment using protein cages and crystals for the development of biohybrid materials. *Chem Commun (Camb)* **52(39)**, 6496-6512 (2016).

23 Fan, K., Cao, C., Pan, Y., Lu, D., Yang, D., Feng, J., Song, L., Liang, M., Yan, X. Magnetoferritin nanoparticles for targeting and visualizing tumour tissues. *Nat Nanotechnol* **7(7)**, 459-464 (2012).

24 Daniels-Wells, T. R. & Penichet, M. L. Transferrin receptor 1: a target for antibody-mediated cancer therapy. *Immunotherapy* **8(9)**, 991-994 (2016).

25 Falvo, E., Tremante, E., Arcovito, A., Papi, M., Elad, N., Boffi, A., Morea, V., Conti, G., & Toffoli, G., Fracasso, G., Giacomini, P., Ceci, P. Improved doxorubicin encapsulation and pharmacokinetics of ferritin-fusion protein nanocarriers bearing proline, serine, and alanine elements. *Biomacromolecules* **17(2)**, 514-522 (2016).

26 Fracasso, G. *et al.* Selective delivery of doxorubicin by novel stimuli-sensitive nano-ferritins overcomes tumor refractoriness. *J Control Release* **239**, 10-18 (2016).

27 Radoshitzky, S. R., Kuhn, J.H., Spiropoulou, C.F., Albariño, G.C., Nguyen, D.P., Salazar-Bravo, J., Dorfman, T., Lee, A.S., Wang, E., Ross, S.R., Choe, H., Farzan, M.,. Receptor determinants of zoonotic transmission of New World hemorrhagic fever arenaviruses. *Proc Natl Acad Sci U S A* **105(7)**, 2664-2669 (2008).

28 Dalton, A. C. & Barton, W. A. Over-expression of secreted proteins from mammalian cell lines. *Protein Sci* **23(5)**, 517-525, doi: 10.1002/pro.2439 (2014).

29 Russo, C. J., and Passmore, L.A. Ultrastable gold substrates for electron cryomicroscopy *Science* **346(6215)**, 1377-1380 (2014).

30 Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W., Schultz, P. . Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21(2)**, 129–228 (1998).

31 Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B. Automated molecular microscopy: the new Leginon system *Struct Biol* **151(1)**, 41-60 (2005).

32 Zheng, S. Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., Agard, D.A. MotionCor2: anisotropic correction of beam-induced motion for improved single-particle electron cryo-microscopy *Nat Methods* **14(4)**, 331–332 (2017).

33 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193(1)**, 1-12 (2016).

34 Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180(3)**, 519-530 (2012).

35 Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol* **189(2)**, 114-122 (2015).

36 Zhou, M., Li, Y., Hu, Q., Bai, X.C., Huang, W., Yan, C., Scheres, S.H., Shi, Y. . Atomic structure of the apoptosome: mechanism of cytochrome c-and dATP-mediated activation of Apaf-1. *Genes Dev* **29(22)**, 2349-2361 (2015).

37 Grant, T., Rohou, A., Grigorieff, N. . cisTEM, user-friendly software for single-particle image processing. *Elife* **7**, e35383 (2018).

38 Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H., Henderson, R. High-resolution noise substitution to measure overfitting and validate

resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35 (2013).

39 Kucukelbir, A., Sigworth, F. J., Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11(1)**, 63-65 (2014).

40 Adams, P. D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66(Pt 2)**, 213-221 (2010).

41 Pettersen, E. F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25(13)**, 1605-1612 (2004).

42 Masuda, T., Goto, F., Yoshihara, T., Mikami, B. The universal mechanism for iron translocation to the ferroxidase site in ferritin, which is mediated by the well conserved transit site. *Biochem Biophys Res Commun* **400**, 94-99 (2010).

43 Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. . Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66(Pt 4)**, 486-501 (2010).

44 Chen, V. B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66(Pt 1)**, 12-21 (2010).

45 Barad, B. A., Echols, N., Wang, R.Y., Cheng, Y., DiMaio, F., Adams, P.D., Fraser, J.S. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* **12(10)**, 943-946 (2015).

Acknowledgements

We thank the H2CU, Honors Center of Italian University for support to BV, CT and LCM, the AIC (Associazione Italiana Cristallografia) for support to LCM and, the Italian Academy at for Advanced Studies at Columbia University for support to BV. We are grateful to Dr. L. Shapiro (Columbia University) for the kind gift of the pα−H BiP vector, modified from the original pα−H optimized for protein secretion by Dr. D.J. Leahy (Texas University at Austin). We are grateful to TW and the BioImaging Facility at CUNY ASRC (NY, USA) and to synchrotron ESRF (beamlines CM01, Grenoble, France). This research was supported by grants from the MIUR flagship Project "Nanomax" (AB) and Associazione Italiana per la Ricerca sul Cancro (AIRC) I.G. Grant 16776 (PC).

Author Contributions:

AB, AdG and BV conceived the work and designed the project. BV performed cloning of CD71. LCM, CT and BV expressed and purified CD71. PC, EF and MP designed mutations and expressed and purified human apo-ferritin and mutants. LCM optimized the biochemical methods for complex formation and isolation. LCM and CT prepared specimen for Cryo-EM and performed data collection. LCM, CT and AdG performed data analysis, generating maps. LCM, CT and CS performed model building, refinement and validation. CT and GP conceived and performed HeLa cells ferritin uptake, collected and analyzed confocal and FACS data. PB performed the FITC labeling of ferritins. AA collected, analyzed and interpreted SPR binding data. LCM, CT, BV and AB interpreted data. LCM, CT and CS produced figures. LCM, CT, FM, AB, BV, AdG wrote and finalized the manuscript.

Author Information

Competing interests: The authors declare no competing interests.

Correspondence and requests for materials should be addressed to beatrice.vallone@uniroma1.it and amedee.desgeorges@asrc.cuny.edu.

Figures and Figure Legends

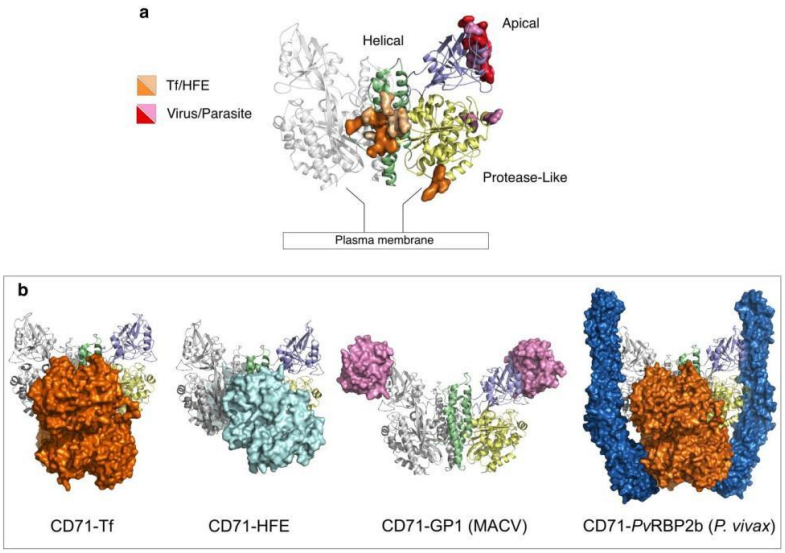
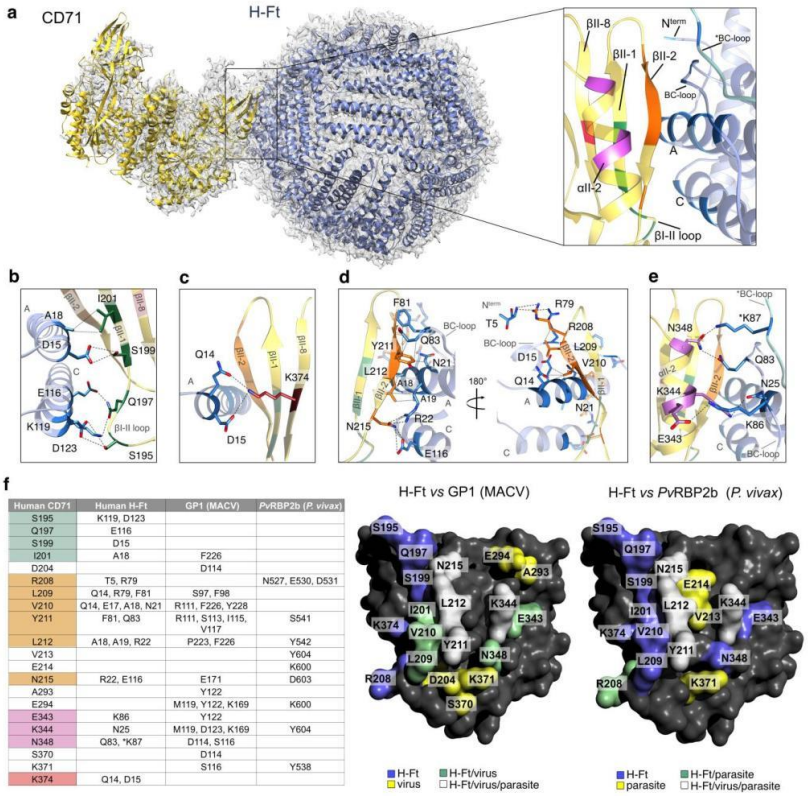


Figure 1. CD71 receptor: ligand recognition epitopes and binding modes. CD71 homodimer is shown in ribbon representation (pdb 3KAS [11]). One monomer is in light grey, the other is colored variably to the receptor domains (apical, light blue; protease-like, yellow; helical, green). **a.** CD71 residues identified as recognition epitopes for Tf/HFE and viruses/parasite are represented as orange/wheat and red/pink surfaces, respectively. **b.** CD71 receptor is shown bound to Tf (orange surface, pdb 1SUV [9]), HFE (cyan surface, pdb 1DE4 [16]), GP1 protein of MACV (pink surface, pdb 3KAS), and Tf (orange surface) and *Pv*RBP2b from *P. vivax* (blue surface, pdb 6D04 [10]).



Human CD71	Human H-Ft	GP1 (MACV)	*Pv*RBP2b (*P. vivax*)
S195	K119, D123		
Q197	E116		
S199	D15		
I201	A18	F226	
D204		D114	
R208	T5, R79		N527, E530, D531
L209	Q14, R79, F81	S97, F98	
V210	Q14, E17, A18, N21	R111, F226, Y228	
Y211	F81, Q83	R111, S113, I115, V117	S541
L212	A18, A19, R22	P223, F226	Y542
V213			Y604
E214			K600
N215	R22, E116	E171	D603
A293		Y122	
E294		M119, Y122, K169	K600
E343	K86	Y122	
K344	N25	M119, D123, K169	Y604
N348	Q83, *K87	D114, S116	
S370		D114	
K371		S116	Y538
K374	Q14, D15		

Figure 2. Structure of the human CD71/H-Ft complex. a. Atomic model of the complex of human CD71 receptor (yellow) and human H-Ft (light blue) fitted in the cryo-EM map at a global resolution of 3.9 Å (grey mesh) are shown. On the side, close-up view of the contact region. The interacting residues of CD71 are highlighted by different colors depending on the specific secondary structural element of the apical domain to which they belong. Interacting residues of H-Ft are colored in light blue. **b-e.** CD71 and H-Ft interacting regions. Panels b and c show the "exclusive contacts"; panels d and e show the "common contacts". Contacting residues within 5 Å

distance are shown in sticks, labeled and colored according to the color code used for secondary structure in panel a (right side). Dashed black lines indicate electrostatic interactions. Dotted grey lines represent hydrophobic contacts. **c.** Table: summary of the interfacing residues at CD71 apical domain contacted by the human H-Ft, GP1 of Machupo virus [11] and *Pv*RBP2b of *P. vivax* [10]. CD71 residues contacted by H-Ft are colored according to the color code used in panel a (right side). On the right, surface representation of the human CD71 apical domain (dark grey). Ligand specific and overlapping binding residues are mapped out following the color code shown at the bottom.

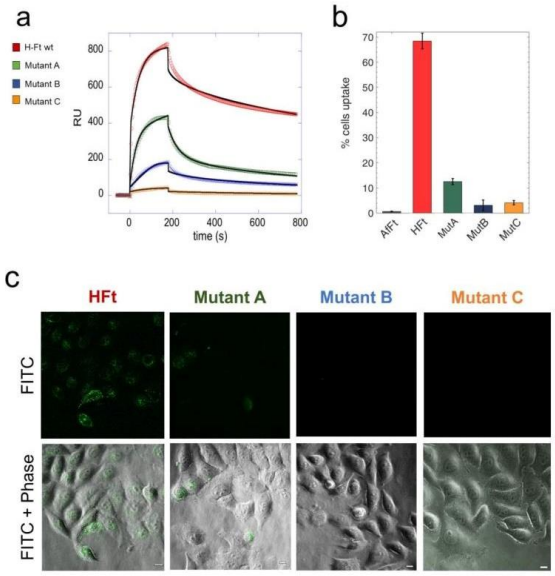


Figure 3. Characterization of human H-Ft wild type and mutants binding to CD71. a. SPR sensograms of the interaction between the immobilized his-tagged CD71 receptor and H-Fts (wild-type and mutants), used as analytes. Fits are reported as black lines. **b.** Ferritins uptake in HeLa cells has been quantified by flow cytometry. The percentage of cells internalizing AfFt-FITC (here used as negative control [18]), H-Ft-FITC, MutA-FITC, MutB-FITC and MutC-FITC is shown as mean ± s.e.m. for *n=3* independent experiments. **c.** Ferritins internalizations in HeLa cells observed at a 60x confocal microscope, here shown as single FITC and overlay images with phase-contrast. Scale bars: 10 µm.

Acknowledgments

This thesis has been carried out in Rome, in the laboratories of the Department of Biochemical Sciences in Sapienza University and of the Italian Institute of Technology Center for Life Nanoscience (IIT CLNS@Sapienza).
A part of this thesis has also been performed in New York, in Prof. Mancia Laboratory of Columbia University, Department of Physiology and Cellular Biophysics, and in Prof. Des Georges Laboratory in Advanced Science Research Center of CUNY University.

These acknowledgments will probably be very repetitive, but (luckily!) during my PhD path I have mostly met inspirational and passionate people.
Firstly, I truly and deeply thank my tutor Prof. Alberto Boffi for always believing in me and in everything I do, for always being so enthusiastic and inspiring about almost everything. Thanks!!!
A profound and warm acknowledge to Prof. Beatrice Vallone for continuously encouraging me in good and bad times, for the great scientific and personal opportunity to enjoy all those wonderful days in New York, and for being a true mentor.
Also, a huge thanks also to my former master's degree tutor, Prof. Giancarlo Ruocco, for all the many counsels (scientific and personal) he has given and he continues to give me everyday, with his sole presence. I owe him a lot!

Then, I really thank prof. Amedee des Georges and Filippo Mancia for the great opportunity of spending several months in their laboratories in New York; here, I always felt home. In particular, thanks to Amedee for teaching me all I know about cryo-EM and for the patience in all the thousand emails and skype calls during the data analysis!
My strong gratitude goes also to Prof. Francesco Malatesta, coordinator of my Ph.D. program, for his supportive and continuous interest in me and in all my results.

Finally, I would like to thank all the other people who helped me during this three years path: dr. Paola Baiocco, dr. Giovanna Peruzzi, dr. Simone de Panfilis, dr. Danya Ben-Hail (des Georges lab), dr. Gabriele Cerutti, dr. Matilde C. Trabuco, dr. Alessandra Bonamore, prof. Silvia di Angelantonio, dr. Carmelinda Savino. Thanks to all the wonderful people in IIT laboratories, to make it a real multidisciplinary and exciting place to work in.
Last, but really not the least!, a profound and heartfelt thanks to Linda C. Montemiglio, who helped me beyond any "post-doc" duty; she is not only a wonderful scientific guide, but also and mostly a great friend and personal counselor. Thank you for all the times spent together at your and my workstations, on skype, in New York and Grenoble, I'll never forget them.

Grazie!

Claudia