Protein Engineering of Microbial Ferritins

by

Hawa Gyamfi

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada

EXAMINING COMMITTEE MEMBERSHIP

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner   NAME: Pelletier Joelle

                                         Title: Professor, Université de Montréal, Montréal, Quebec, Canada

Supervisor   NAME: John F. Honek

                                         Title: Professor

Internal Member   NAME: Tong Leung

                                         Title: Professor

Internal Member   NAME: Elizabeth Meiering

                                         Title: Professor

Internal-external Member   NAME: Roderick A. Slavcev

                                         Title: Associate Professor

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Nanotechnology has become a field of intense interest due to its applications in medicine, cosmetics, energy, and fabrics. The field has evolved to include nanodimensional self-assembling cage proteins which are multisubunit proteins surrounding a hollow cavity. Cage proteins can be engineered with precision, by either chemical, enzymatic or recombinant DNA methods, to accommodate non-native guest molecules both internally or externally. The inside cavity could be engineered to encapsulate inorganic nanoparticles or drugs, enzymes, or to serve as a size constraint reaction vessel for nanoparticle synthesis. However, exterior modification of these proteins has not been extensively explored. The most common initial studies have included the attachment of antibodies for targeting. We show that: (1) The N-termini of the subunits of *Escherichia coli* bacterioferritin (Bfr) can be modified by recombinant DNA methodologies to produce peptide fusions, which are termed "QTag-Bfr" tags, and which can be modified by bacterial transglutaminase to link substituted amines to the exterior surface of Bfr. These modifiable tags may be employed in the future for affinity chromatography applications, as precursors to nanoparticle superlattices and as anticancer drug/imaging agent delivery vehicles. The exact sequence of these tags was found to also control the final quaternary structure of Bfr. (2) The interior surface can be modified with affinity tags to allow for controllable encapsulation of functionalized guest molecules, and we extend previous research by elaborating on the incorporation of green fluorescent protein, gold nanoparticles, and micelles with Bfr and the "QTag-Bfr" constructs. (3) Similar strategies were extended to a recently discovered thermophilic *Archaeoglobus fulgidus* ferritin (AfFtn), and various engineered AfFtn were studied with respect to the encapsulation of a range of guest molecules, yet also provide a more open cavity capable of novel investigations on guest entrapment. (4) An extensive computational docking study was undertaken on the Bfr heme cofactor binding site which led to the identification of potential non-heme replacements which were tested for control of the quaternary structure of *E. coli* Bfr. These studies provide several versatile nanodimensional capsule protein frameworks which could be utilized in the future for the production of new biomaterials, as therapeutics in medicine and expand our knowledge on capsule protein biotemplating and protein engineering.

ACKNOWLEDGMENTS

DEDICATION

I dedicate this work to my dear husband, Adnan Osman, and my lovely kids Osman Adnan and Habibah Adnan.

TABLE OF CONTENTS

LIST OF FIGURES

xv

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.0. Overview

Recent advances in bionanotechnology have included biological materials (biological nanostructures) with relevant nanometer dimensions. Biological nanostructures have defined subunits with a high degree of structural organization. This arrangement allows for the precise modifications of their addressable surfaces; exterior, interior, and the interfaces between the subunits.[1] The applications of these molecules include but are not limited to: materials engineering,[2,3] health [4]and energy production.[5-7]

Here we show how two nanodimensional multisubuit proteins, *Archaeoglubus fulgidus* ferritin (AfFtn) and *Escherichia coli* Bacterioferritin (Bfr) (seen in Figure 1) have been engineered to encapsulate or display functional guest molecules. We also elaborate on the modification strategies as well as their applications in various fields of science. Both AfFtn and Bfr proteins are roughly spherical with 8 nm and 12 nm interior and exterior diameters, respectively. Their spherical shapes are assembled from 24 identical subunits. The subunits self-associate with their C-termini oriented towards the interior surface and the N-termini towards the exterior surface of the protein cages.[8,9] The precise positioning of the N- and C-termini allow for multiple functionalizations.
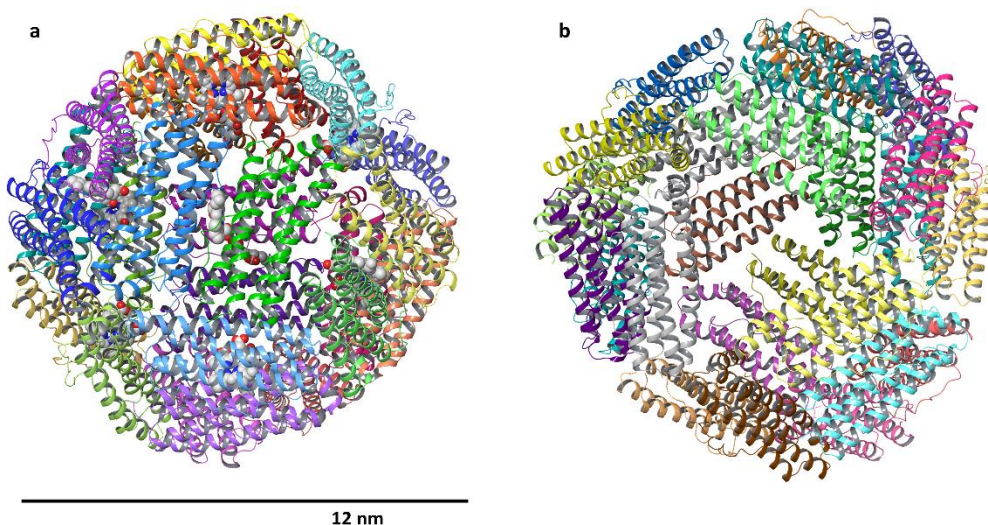
Figure 1. Crystal structures of Bfr and Aftn. (a) Bfr (1Bfr) capsule with 12 heme cofactors (within dimer interfaces). (b) The 12 nm AfFtn (1SQ3) capsule with a 4.5 nm triangular pore. The Figures were generated with Maestro software (Schrödinger, LLC).

The well-defined Bfr exterior surface was genetically fused with peptide sequences that were used for enzymatic modification on the Bfr surface whereas the internal C-termini were genetically fused hexahistidine (H_6) tags for nickel (II) mediated affinity interaction with nitriloacetic acid (NTA)-containing guest molecules. AfFtn was similarly engineered with C-terminal H_6 tags for nickel mediated affinity purification and guest encapsulations. AfFtn containing a cysteine mutation was also studied as to its ability to be modified for the fabrication of a new internal affinity site within the cavity.

Lastly, computational modeling was used to identify heme analogs that might be capable of binding the Bfr heme pocket and hence replacing heme as the Bfr cofactor. Some of the docked ligands, from the library, were subsequently obtained from commercial sources and encapsulated into the Bfr protein and their effects on the Bfr quaternary structure were studied. This chapter introduces multisubunit architectures and their applications, followed by a detailed description of the two proteins in the project (AfFtn and Bfr) and outlines the biotemplating strategies.

1.1. Multisubunit Assemblies

Many proteins are composed of an arrangement of protein subunits. The cage protein family is comprised of several types of proteins with the architectures that assemble from homologous protein chains or a combination of different protein chains.[10] Their shapes are mostly spherical with a few other geometries possible such as: tubes (tobacco mosaic virus, TMV[11]), knots (Ketol-acid reductoisomerase, KARI),[12,13] catenane (Bacteriophage HK97 gp6 capsid),[14] icosahedral virus capsids (Cowpea chlorotic mottle virus)[15] and other intricate structures of well-defined exterior and interior surfaces. Spherical cages are found in ferritins,[16] DNA binding proteins from Starved cells (Dps),[17] certain heat shock proteins,[18] hepatitis B virus,[19] and many others (Figure 1.1). The spherical folding pattern physically separates the exterior surface from the interior surface, making them useful for guest encapsulation applications.[20]

Additionally, DNA assemblies have emerged as building blocks for nanodimensional structures. The advances in DNA nanodevices are due to their ease of production as well as the ability to build complex yet predictable and controllable two-dimensional[21] and three-dimensional[22] structures from only four base pairs (adenine A, guanine G, cytosine C, and thymine T). These advances have inspired the engineering of DNA-based synthetic cage-like assemblies to extend their applications.

1.1.1. Natural and Synthetic Assemblies

Unlike synthetic polymeric materials, natural polymeric assemblies have ease of synthesis, assembly, great structural, and functional properties. Even though the properties of natural polymeric materials make them versatile, they may not entirely replace their synthetic partners. For example, the tensile strength of the toughest natural fiber, spider dragline silk (1.1 GPa), is close to that of steel (1.3 GPa),[23] but nowhere close to the tensile strength of carbon nanotubes which is over 100-fold[24] stronger, coupled with its remarkable electrical properties that are not found in natural polymeric materials. The properties of non-natural and natural polymeric materials have inspired the production of artificial capsid-like cage proteins. These capsids are computationally designed from component protein molecules and further assemble *in vitro* into structures that have desirable properties and promising applications in drug delivery.[25] Another form of biomaterial involves the removal of genetic material from viral capsids to produce non-infectious virus-like particles (VLPs), and these have become of intense current interest. The

3

VLPs can be recombinantly produced and capable of self-associating into their native forms without their genetic material, but having similar antigenicity to the infectious form. VLPs mimic their parent virus structures (without the genetic material) and activate the immune system through pattern-recognition receptors on immune cells to induce a potent immune response.[26] For example, the commercial HPV (human papillomavirus) vaccine, Gardasil® (Merck and Co., Inc) is a VLP-based vaccine that is capable of producing antibodies against HPV related infections.

Additionally, the hollow interior of VLP can be used to encapsulate large quantities of non-native materials. Tauze and Coursagate reported the application of VLP in gene therapy. The interior of recombinant HPV16L1 virus-like particles was loaded with GFP genes that were successfully delivered into eukaryotic cells with higher efficiency than liposome and free DNA transfer strategies.[27]
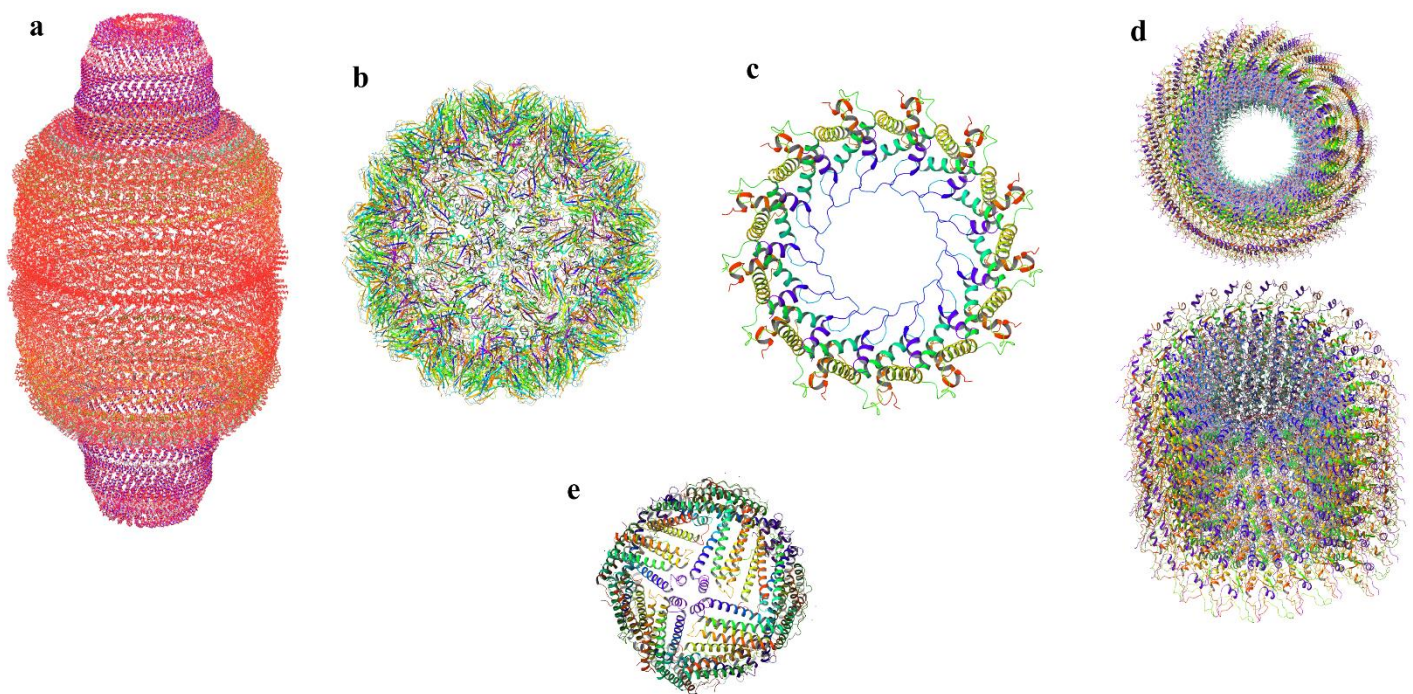
Figure 1.1. Crystal structures of multisubunit biological assemblies. (a) Vault shell (PDB:2QZV) (b) The icosahedral Cowpea Chlorotic mottle virus, CCMV (PDB:3JVO) (c) the Bacteriophage HK97 gp6 capsid catenane (PDB:2IMV) (d) The top and side view of tobacco mosaic virus, TMV (PDB:5A7A) (e) Human H-type ferritin (PDB:2FHA). All images were generated with Maestro software (Schrödinger, LLC).

1.2 Structure-function Relationship of Multisubunit Assemblies

Recent developments in biology and chemistry have shown that the association of multiple protein subunits-into rings, tubes, spheres, and vaults is required for the proper biological function of some biomolecules. For instance, the ferritin cage has been reported to store up to 4500 atoms of iron in vivo,[28] even though the Fe (II) oxidation occurs within the subunit, the oxidized Fe (III) is stored in the interior cavity. Similarly, the process of GroEL-mediated denatured protein folding occurs inside a barrel-shaped multisubunit structure. The repaired protein remains in the barrel until an ATP-induced conformation occurs to release it. These and many more structure-related applications of multisubunit assemblies make them interesting platforms to explore for materials fabrication.

A particular cage protein presents three interfaces for biotemplating: (1) The interior surface, (2) Exterior Surface, and (3) the interfaces between the subunits. The physical separation of these surfaces enables cage proteins to undergo simultaneous modifications. For example, the interior of a cage protein can be loaded with a guest molecule which can be successfully delivered to the desired location via exterior surface targeting. Futhermore, the multiple copies of subunits present multiple copies of amino acid functional groups for modification.[3] This feature allows loading large quantities of cargo molecules on the multisubunit host assembly.

Moreover, the accessibility of the exterior surface facilitates the biotemplating of these assemblies, as illustrated by the engineering of heterogeneous crystals from the electrostatic interaction between a negatively charged cowpea chlorotic mottle virus (CCMV, pI 3.8) and positively charged (pI ~10.5) avidin molecule.[2] Further applications of multisubunit protein assemblies will be discussed under four main subheadings: (1) Synthesis of metallic nanoparticles, (2) Drug delivery and theranosis (3) Energy production and water purification and (4) Materials synthesis

1.2.1 Synthesis of Inorganic Nanoparticles

Ferritins were one of the first cage proteins used in inorganic nanoparticle synthesis. This could be due to the ease of production and its major structural feature, the restricted interior space which helps to constrain the size of the synthesized nanoparticle, and the commercial availability of some ferritins. Cage proteins can template the size of the formed inorganic nanoparticle due to the protein's cavity dimensions. The narrow size distribution imposed by the cage can make the

synthesized nanoparticles useful in drug delivery due to their controlled size which can enhance their permeability and retention (EPR) within tumour tissue.[29] Another advantage of an encapsulated nanoparticle is its decreased toxicity since the protein shell prevents direct contact with cells.[29]

Metallic nanoparticle synthesis can be achieved through either delivering metal ions to the cage interior and allowing them to nucleate (or chemically be reduced to their zero-oxidation state) or by clustering biomolecule subunits around the preformed inorganic nanoparticle which are within the dimensions of the cage interior. The first method has been widely used to synthesize ZnSe nanoparticles where Zn^{2+} ions were permitted to diffuse into the interior cavity of horse spleen apoferritin via its iron transport channels and made to react with Se (as selenourea) slowly. The formation of a ZnSe compound led to autonucleation and growth of the ZnSe core within the apoferritin cavity. Polycrystals of ZnSe were formed with 3 nm diameters which were anaerobically heated at 500 ° C for 1 hour to form 7 nm, single cubic ZnSe crystals (Figure 1.2). The synthesized ZnSe quantum dots could be used subsequently as biological labels due to their uniform size distribution and unquenchable fluorescence signal.[30]
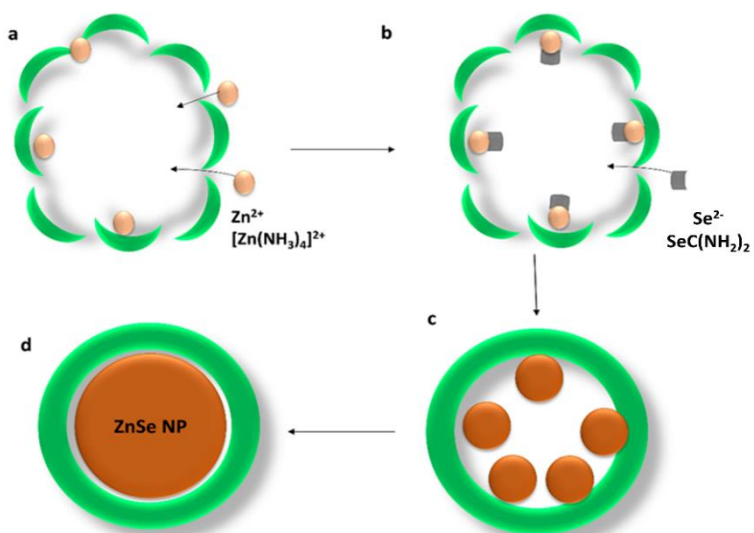
Figure 1.2. Production of ZnSe nanoparticle via ion delivery. (a) The electrostatic localization of Zn^{2+} ions to the HosFtn cavity through the ion channels (upon incubation). (b) subsequent interaction with Se^{2-} ions from selenourea degradation. (c) The reaction between Zn^{2+} and Se^{2-} leads to the formation of 3 nm ZnSe polycrystals in the cavity. (d) Subsequent heating of the 3 nm polycrytals (at 500 ° C for 1 h) resulted in a 7 nm core crystal.This figure was redrawn from Iwahori *et al*.[30]

In addition to the production of monodispersed metallic nanoparticles, the cage architecture mediates or enforces metal crystallization. Douglas and Young have shown that CCMV can mediate crystallization of 15 nm paratungstate ($H_2W_{12}O_{42}^{10-}$). A control reaction in the absence of the CCMV templates did not result in the formation of the paratungstate crystals.[31] Furthermore, the encapsulation of preformed nanoparticles was reported by Zheng and co-workers where recombinant apoferritin was used to deliver gold nanoparticles (AuNP) to SiO surfaces. The C- and N-termini of the ferritin subunits were modified with gold-binding peptides and inorganic ($Ti/SiO_2/Ag$) binding peptides, respectively. The AuNP was encapsulated by the affinity interaction between the C-terminal peptide and the AuNP surface which resulted in the packing of the ferritin subunits around the gold nanoparticle, and encapsulating the AuNP (Figure 1.2.1).

The packing of subunits can be applied to various AuNP sizes provided there is the right kind of affinity interaction occurring between the AuNP and the inner surface of the ferritin subunit. This approach serves as an important strategy for the fabrication of materials with surface plasmon resonance properties.[32] Lastly, virus capsids can be utilized to encapsulate quantum dots by modifying the surface of the quantum dot (QD) with the base sequence required for capsid assembly. Self-assembly of the viral coat proteins resulted in encapsulation of the quantum dot.[33] Alternatively, the viral capsids can also be dissociated by changing the pH and ionic strength of the buffer and then utilized to encapsulate non-native guest molecules. This approach involved removal of the genetic material of Cowpea Chlorotic Mottle Virus, and the resulting VLP was used to encapsulate heterogenous CdSe/ZnS quantum dots which were efficiently internalized by HeLa cells and RAW 246.7 macrophages without any toxic effect.[34] These systems can be useful for imaging and diagnostic applications.
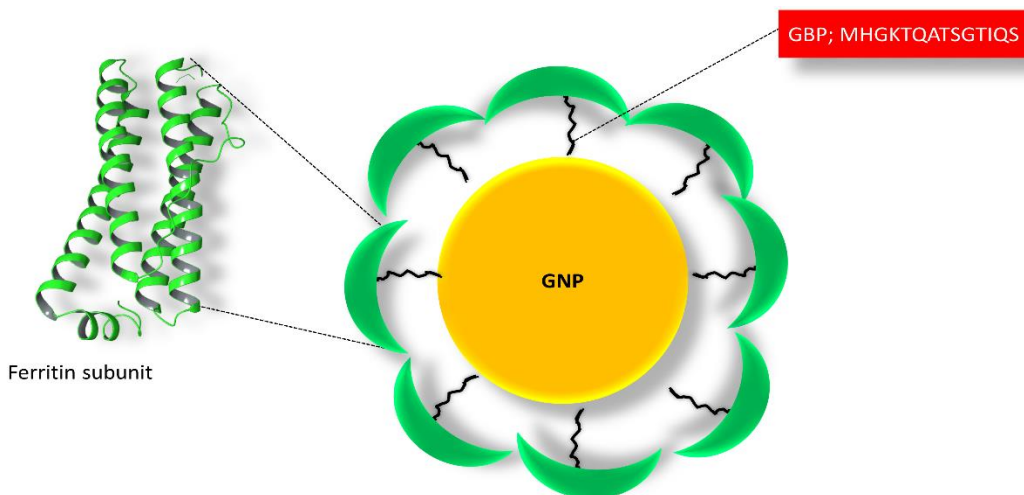
Figure 1.2.1. A representation of gold nanoparticle (GNP) encapsulation via subunit packing. The C-termini of ferritin subunits were modified with gold binding peptide (GBP) which interacted with the GNP surface as demonstrated by Zheng and co-workers.[32] This Figure was generated with Maestro (Schrödinger, LLC) and Microsoft PowerPoint.

1.2.2. Theragnosis and Drug Delivery

Biological assemblies containing therapeutic agents have promising applications in medicine. Some formulations are currently employed in medicine while others are being investigated in clinical trials.[25b, 29] The continuous exploitation of proteins for medicinal applications despite the existing liposomal delivery vehicles is due to their numerous advantages, not limited to: biocompatibility and ease of metabolism; non-toxic; well-defined size that enhances permeability; ease of modification and production; as well as the availability of structural and functional information for many proteins.[35,36] Clinical applications of caged proteins arise due to their generally high stability, biocompatibility and their ability to dissociate and associate and encapsulate therapeutic molecules. The hollowed interior serves as a useful space for drug and imaging cargo and their surfaces could be used for targeting. Examples of these are not limited to the *in vivo* targeting of arteriosclerosis-associated macrophages with tumor targeting LyP-1 peptide on the surface of heat shock protein which was loaded with fluorescent dyes. This arrangement allowed for site-specific imaging and diagnosis of atherosclerotic plaques.[37] Also, Falbo *et al*. have demonstrated the potential of ferritin-based

nanoparticles in the drug delivery area, wherein the surface of human ferritin was modified with melanoma targeting antibodies, and the interior was loaded with the anticancer agent cisplatin. The Ep1 monoclonal antibody specifically targetted the CSPG4 melanoma antigen to deliver the cisplatin to C30 melanoma cell lines. A proper stoichiometry between antibody, ferritin, and cisplatin generated a drug with remarkable anti-tumor effects.[4b] Lastly, the iron core in ferritin has been used as an MRI agent due to its intrinsic quantum mechanical properties.[38] The mineral core is capable of accelerating transverse relaxation of water protons that leads to black or dark contrast in tissues.[36a] The super magnetic ferric oxide mineral was used as a reporter for delivery of replicative-effective adenovirus (Adv) vector to mouse brain. The presence of this vector resulted in the expression of ferritin subunits, which were used to sequester cellular Fe (II) in the brain tissue and stored as the Fe (III) oxide   MRI contrasting agent.[39] The medical applications of multisubunit biological assemblies are diverse with untapped applications. Their structural properties could enable the fabrication of multifunctional theranostic molecules beyond the double duties illustrated here. There is the potential for building an all-in-one theranostic molecule with integrated diagnosis, drug monitoring, targeted delivery, and controlled drug release.[36a]

1.2.3. Energy Production and Water Purification

1.2.3.1. Energy Production

Hydrogen gas production is another application of metallic nanoparticle synthesis. Two groups of researchers have reported the conversion of atomic hydrogen to molecular hydrogen employing cage proteins. In 2005, Varpness *et al*. mimicked the hydrogenase enzyme with a small heat shock protein (Hsp), the interior of Hsp was loaded with up to 1000 platinum atoms. Hydrogen gas was produced by photoreduction of methyl viologen by $Ru(bpy)_3^{2+}$ which then transferred electrons to the Pt mineral surface and caused proton reduction. They observed a comparable rate of proton reduction to that produced by an active hydrogenase enzyme.[5] Recently Clark and co-workers presented another route for hydrogen gas production using the bacterioferritin heme cofactor. The native heme was replaced with a zinc analog, zinc-protoporphyrin IX, and platinum was encapsulated within the bacterioferritin cavity. The light activation of triethanolamine resulted in electron release which was mediated by methyl viologen (as illustrated by Varpness *et al*.) and proton reduction to form H_2 was observed.[6]

1.2.3.2. Water Purification

There is an increased release of phosphorus in municipal wastewater since most synthetic detergents are phosphorous-based. Most of the phosphorus is converted to phosphate in wastewater, which has detrimental effects on aquatic life and increased water purification cost. The primary method of phosphate removal is chemical precipitation. Chemical precipitation uses coagulants to precipitate the phosphate ions, lowering phosphate levels, but does not reach the 0.01 mg/L ecological level. Also, this process is expensive, requiring a higher concentration of coagulants and leads to increased sludge volume.[40]

An alternative approach to phosphate removal is the biological phosphate removal which uses organisms capable of soaking up these ions. The idea that ferritin cages can be used in water purification stems from the presence of phosphate ion in the Fe (III) mineral core. Aitken *et al.* observed up to a 1:1 ratio of Fe (III) to phosphate in the bacterioferritin mineral core. Phosphate ions serve as counter-ions for maintaining the positive charge on Fe^{3+} and also are believed to enhance the rate of core formation (by 5-fold) by an unknown mechanism.[41] Ferritin-mediated phosphate absorption was illustrated by the thermophilic *Pyrococcus furiosus* ferritin, PfFtn. The PfFtn mineral core completely removed phosphate ions in solution with a gram of PfFtn removing up to 11 mg of phosphate from solution, unlike the chemical precipitation method. Another advantage of this approach is that the PfFtn cage could be regenerated to save on the purification cost. [42] A commercial example is the "intelligent enzyme solution" from EUCODIS Bioscience and BiAqua limited, who launched a pilot-scale *Pyrococcus furiosus* ferritin water treatment system on September 30, 2013. The company has stated that they have produced 1 m^3 of the protein of this particular ferritin for this purpose. http://www.eucodis.com/

1.2.4. Materials Synthesis

There are increasing demands for flexible electronics for health monitoring and other materials applications. This demand has led to the successful fabrication of inorganic materials with unique electrical, optical, and plasmonic properties. However, these materials suffer from sustainability and biocompatibility issues. Hence biomolecules have emerged as naturally occurring, biocompatible, and sustainable alternatives. Biomolecules are flexible and may find

applications in bendable materials such as paper-like displays, photovoltaic cells, light-emitting diodes, and theranostic devices.[43-45]

The self-assembling property of multisubunit proteins allows them to exist in different shapes and sizes for one, two, and three-dimensional materials. For example, the cylindrical shaped virus particles have been employed in the fabrication of thin and flexible electronic devices. Genetically programmed self-assembly was used to immobilize $BaTiO_3$ nanocrystals on the filamentous bacteriophage M13 coat proteins and resulted in a high electrical output (up to 300 nA and 6 V) energy harvesting device.[7]

Alternatively, spherical assemblies can be immobilized onto two-dimensional electrodes as seen in the construction of a ferritin-based indium oxide electrode. The electrode was functionalized with horse spleen ferritin through a poly (L-lysine) and poly(L-arginine) linker on the indium oxide surface and interactions resulted in the formation of a monolayer of ferritin-coated electrode having 9-13 x10^{11} ferritin molecules per cm^{-2} of the electrode. This ferritin-coated electrode has potential application for *in situ* voltammetric measurements.[46] Moreover, a one-dimensional array was reported by Biswas *et al*., wherein they fabricated up to 2.5-micron long tubes from the GroEL chaperonin without affecting its native functionality, that of a protein folding chaperone. The rim of the GroEL barrel (Figure 1.2.2) was modified with cysteine residues which interacted with spiropyran (SP), in the presence of Mg^{2+} and light, to spontaneously form the merocyanine isomer. The merocyanine (MC) induced self-association of the micrometer long GroELSP/MC fibers. These fibers were capable of aiding in the refolding of denatured α-lactalbumin,[47] showing that biological multi-assemblies are very robust and can withstand different degrees of modification without removing their native functionalities.
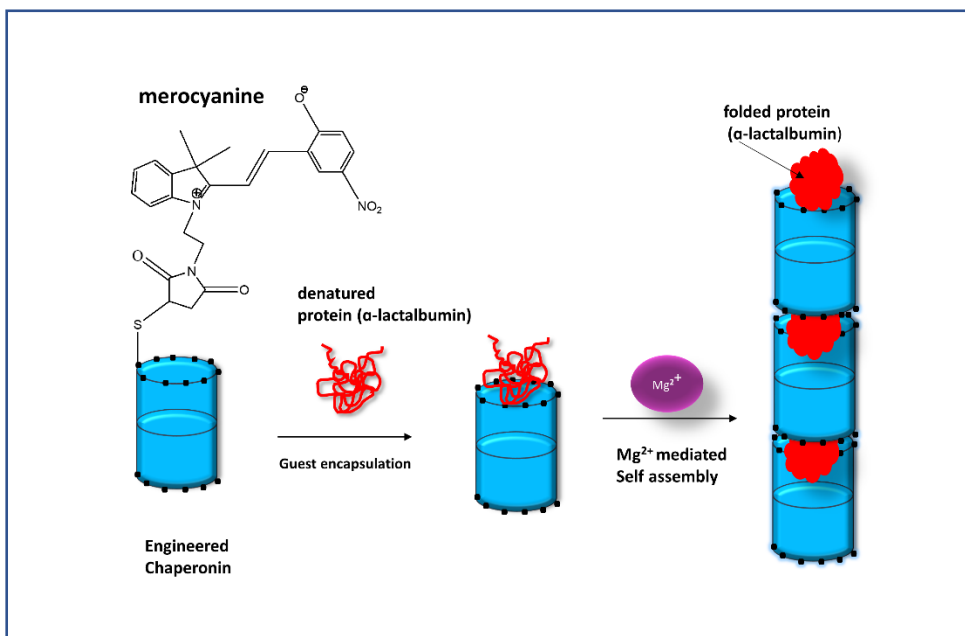
Figure 1.2.2. Formation of one-dimensional tubular assembly from GroEL-merocyanine conjugate. The assembled complex possesses chaperonin activity. Figure was generated with Chemdraw and Microsoft PowerPoint based on the report by Biswas *et al.*[47]

1.2.6. Catalytic Confinement

The process of encapsulating active enzymes within capsule proteins might be a useful mimic of the crowded intracellular cellular conditions found within cells. It has been suggested that the encapsulated enzyme may have an increased rate of catalysis due to an increased collision between the confined enzyme and the substrate to form the enzyme-substrate complex.[48,25b] For example, key enzymes (methylglyoxal synthase, glycerol dehydrogenase, and alcohol dehydrogenase) involved in the biosynthesis of the compound R-1,2-propanediol had a higher catalytic turnover for the compartmentalized enzymes which were confined to a specific sub-volume while all other metabolites could freely diffuse. The compartmentalization resulted in a higher catalytic turnover being measured compared to the non-compartmentalized pathway.[49]

Conversely, excessive enzyme loading in the cavity may prevent substrate turnover due to the constrained interior space of the host protein, and this may lead to decreased turnover.[50] Hence it necessary to control the enzyme-host ratios in order to make room for the substrate. The

controlled entrapment of enzymes/catalysts in a caged assembly enables the investigation of the rates of diffusion of reactants and products. When a single horseradish peroxidase (HRP) enzyme was encapsulated within the CCMV capsid, the pores of CCMV allowed diffusion of the dihydrorhodamine 6G substrate into the core. The observed diffusion times of fluorescent products from entrapped HRP were different from those detected for free HRP due to pH-dependent swelling of the capsid, which is believed to affect substrate mobility in and out of the capsid.

1.3. Cage Proteins

Cage proteins can be defined as three-dimensional (3D) assemblies of protein subunits that form an internal cavity when these subunits self-associate. These topologies are diverse in terms of dimensions (9 -750 nm),[17,52] structure, overall fold, stability, and electrostatics.[10,13] Despite the structural variations, their subunits assemble into structures that have hollow interior cavities which can be useful for the encapsulation of a range of guest molecules, targeted guest delivery, constrained reaction vessels, imaging, and diagnostics.[53,29-30,36a,54] The encapsulated guests are protected from the external environment, similar to how viral capsids protect their genetic material. These shells are capable of withstanding very harsh conditions, yet dynamic enough to release their encapsulated cargos. Their dynamic nature is mediated by protein-protein interactions between the subunits. The exterior motifs are useful in targeting whereas the interior motifs accommodate the guest molecules.

Interesting cage structures and their possible applications have been reported (Figure 1), but this work focuses on the ferritin family. The ferritin superfamily and their structure-function relations will be described. Lastly, two types of ferritins, *Archaeoglobus fulgidus* ferritin and *E. coli* bacterioferritin (Bfr), which were used in this project, will be described in detail.

1.3.1. Ferritins

1.3.2. The Ferritin Superfamily and Classes

The ferritin cages are subdivided into: the classical and universal ferritins (maxi ferritins) of Prokarya and Eukarya[16]; the prokaryotic Dps ferritins (mini ferritins)[17] and the heme-containing bacterioferritins in bacteria.[55] Maxi ferritins are assembled from 24 subunits with interior and exterior diameters of 8 and 12 nm, respectively, whereas the mini ferritins have 12 subunits with 8 nm and 6 nm exterior and interior diameters (Figure1.1), respectively.[56] The cages are also referred to as iron storage containers due to their ability to store essential Fe^{3+} in a stable, bioavailable form. A maxi ferritin cage can store up to 4500 atoms of iron.[28] The bioavailability of natural iron in most habitats makes it useful for catalytic reactions in many organisms, requiring organisms to regulate the cellular levels of iron. Fe^{3+} precipitates under physiological conditions due its limited solubility[57] so ferritins have evolved to store this ion in its Fe^{3+} oxidation state in biocompatible and bioavailable forms to prevent issues associated with iron toxicity. Ferritin nomenclature is mostly derived from the organism whose ferritin is being studied, such as human ferritin,[16] horse spleen ferritin,[58] *E. coli* ferritins,[59] *Arhaeoglobus fulgidus* ferritin,[8] *Pyrococcus furiosus* ferritin,[60] *Pseudomonas aeruginosa* bacterioferritin,[61] bullfrog M-type ferritin,[62] and many more.
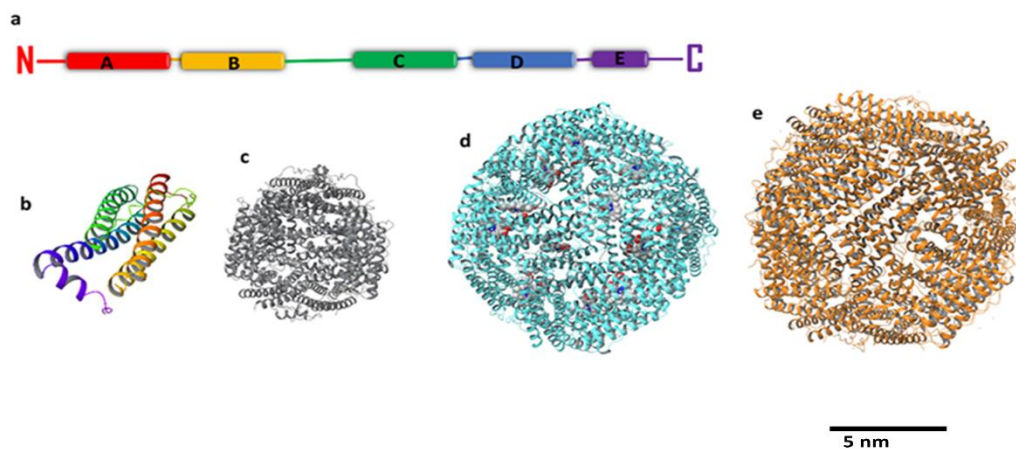
Figure 1.3. Composition, structure, and size comparison of the ferritin family members. The four alpha-helical bundle that makes up individual subunits (a, b). The Dsp mini ferritin (2IY4), *E. coli* bacterioferritin with its heme cofactors in spheres (1BFR) and the human H-type ferritin (2FHA), (c-e) respectively.The images were generated with Maestro (Schrodinger, LLC) software.

1.3.3. Structural Features of Ferritin

Even though ferritins are approximately spherical in overall shape, they have very low sequence similarity (20%) among their family members [9,16-17,59,61,63]. There are various types of intersubunit associations that contribute to the overall 24-mer or 12-mer quaternary structures. The individual subunits from the 24-mer and 12-mer ferritins are all α-helical, similar to that of methane monooxygenase, ribonucleotide reductase, archaeoferritin, bacterioferritin, DNA binding protein from starved cells (Dps), and many others from the four α- helix bundle structure family.[56] Each subunit has four (A-D) tightly packed left-handed alpha helices. Successive helices are connected by a series of short loop sequences except that the A-B pair of helices are connected to the C-D helices via a long loop sequence (that connects helix B to C). Inaddition, a fifth helix, the E helix, which is the shortest helix of the subunit, is located at an angle relative to the arrangement of the C-terminal sequence of the D helix(Figure 1.3). The individual helices interact via hydrogen bonding.

Also, there are 3-fold and 4-fold axes of symmetry in the overall ferritin three-dimensional structure where the apices of the rhombic dodecahedron touch in the 432-point group symmetry (Figure 1.3.1a). The 3- and 4-fold subunit arrangements result in the formation of 0.3-0.5 nm diameter channels that penetrate the shell (Figure 1.3.1b). These channels are useful for iron and small molecule transport, and the close packing of subunits gate them from unwanted reducing agents.[58] The arrangement of subunits results in the formation of a hollow interior that is used for iron storage. The amount of stored iron is dependent on cellular availability. Hence the cage may exist with or without an iron core yet maintain its three-dimensional structure.
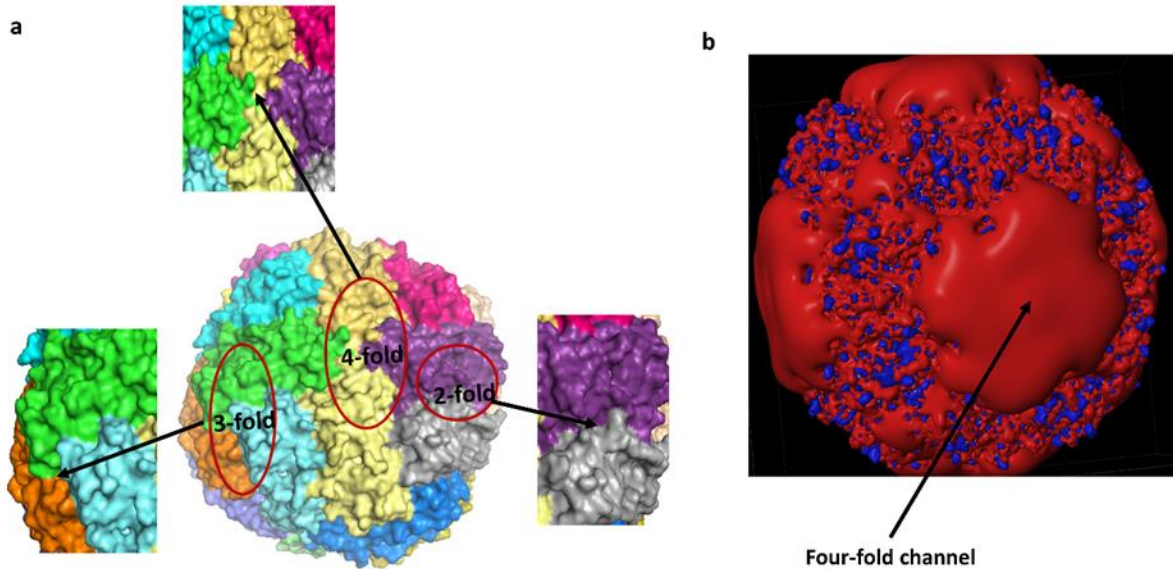
Figure 1.3.1. The surface of a Bfr (1Bfr) showing the symmetry-related axes. (a) The 3 and 4-fold channels are used for Fe^{2+} and small molecule transport. pyMOL image (b) The electrostatic potential on the exterior surface of Bfr with pores at 4-fold axes (red pores) for transport in and out of the cage (Source: J. Honek).

Some members of the classical ferritins are homopolymeric while others are heteropolymeric.[64] A particular ferritin cage could be composed of a mixture of heavy, light and middle weight chains having masses of 22.8 kDa, 20 kDa, and 21 kDa, respectively.[64] The 24-mer assembly of eukaryotic ferritins is usually a combination of these homologous chains. An example is ferritin from horse spleen (HoSF) which is self-assembled from variable amounts of H and L-type chains. Alternatively, homopolymeric eukaryotic ferritins have been genetically engineered to possess only one type of these chains such as the human H-type ferritin (HuHF) and the bullfrog M-type ferritin.[64]

On the other hand, the mini ferritins have tetrahedral (23-point group) symmetry and form a hollow assembly of 12 identical subunits. The interior and exterior dimensions of the mini ferritin cage are roughly 4.6 nm and 9 nm, respectively. The subunit composition is similar to many other ferritins; there are four alpha helices (A-D) tightly connected with short loops except

that they lack the short C-terminal E helix. This E helix is involved in symmetry related interactions to generate the four-fold channels in ferritins. As a result, there are no 4-fold channels in mini ferritins, but the C-termini of the D helices engage other 3-fold interactions with lower symmetry that results in a more compact cage architecture, capable of withstanding the highly negative charge distribution and a cluster of acidic residues on the interior surface and 3-fold axes, respectively.[17,65] Figure 1.3 shows the size comparison of the three classes.

1.3.4. Ferritin Self-assembly

Five interacting subunits surround each ferritin subunit. The flexible loop regions of two separate subunits give rise to a two-fold symmetry arrangement known as dimers. Other oligomeric forms of subunit association such as trimers, tetramers, as well as octamers have been reported,[66,58] but the exact intermediates involved in the native 24-mer arrangement has not been established. Thus, it is necessary to investigate the species involved in the self-assembly process, from monomers to intact 24-mers. Analysis of sedimentation velocity and circular dichroism analyses of pH-dependent reassembly showed that self-assembly of horse spleen ferritin (HoSF) proceeds from a dimer through to tetramers and octamers.[66] However, dimers and trimers could efficiently produce 80 % of the 24-mers in solution while tetramers alone did not form intact 24-mers.[58] This appears to suggest an alternative route for 24-mer assembly. In order to identify the quaternary structures that lead to 24-mers, Yang and coworkers used "virtual alanine scanning" to identify key residues at the 2-fold, 3-fold and 4-fold axes of *E. coli* bacterioferritin. Nine residues at the 2,3 and 4-fold axis were designed and further analyzed via size exclusion, temperature dependent CD, dynamic light scattering, and native PAGE for 24-mer formation. Their results showed that mutation of two 3-fold residues (R61A and Y114 A) and one 2 -fold residue (R30A) completely abolished the formation of intact 24-mers leading to the formation of only dimers in solution.[67] Recently, Huard *et al.* used Cu(II)-induced self-assembly to study the mechanism of 24-mer formation (for human H-ferritin) where they induced dimer formation from monomeric subunits in response to Cu(II) via the reverse metal-templated interface redesign (rMeTIR) strategy and found Y39E, N74E, and P88A mutants form monomers that were converted to dimers upon addition of Cu(II) and ultimately 24-mer shells.[68] The Cu(II) mediated 24-mer formation, and other research suggests that dimers are the most stable intermediates that spontaneously associate into 24-mers in solution. As a result, the residues in the loop region that

connect the A-B helix to C-D play a vital role in the assembly process. The mutation of aspartic acid residue to lysine in the loop (Asp80K in the middle) that connects the dimers in bullfrog M-type ferritin (BfMFtn) is reported to decrease the rate of dimer formation and its stability by abolishing the salt bridge between Asp80 and Lys82. Thus, electrostatic interactions within the dimer interface play a crucial role in the 24-mer assembly process.[64]

Thermally stable ferritins from hyperthermophilic organisms (optimum growth $\geq$ 80 °C), such as *Pyrococcus furiosus* ferritin (PfFtn) and *Archaeoglubus fulgidus* ferritin (AfFtn), are classical ferritins. The PfFtn molecule exhibits the characteristic 432-point group symmetry of classical ferritins and bacterioferritins.[60] Both AfFtn and PfFtn have roughly spherical shapes with a hollow internal cavity, but AfFtn has been reported to exhibit 23 symmetry with four large pores (approximately 4.5 nm) that penetrate the shell (Figure 1).[8] *Archaeoglobus fulgidus* ferritin subunits are monomeric in solution and self-assemble in the presence of salt. The spontaneous assembly is mediated by high salt concentrations (> 150 mM). A 432 symmetry was observed when R151 and K150 in the E-helix were mutated, abolishing the tetrahedral pores in the 23-point group symmetry.[69]
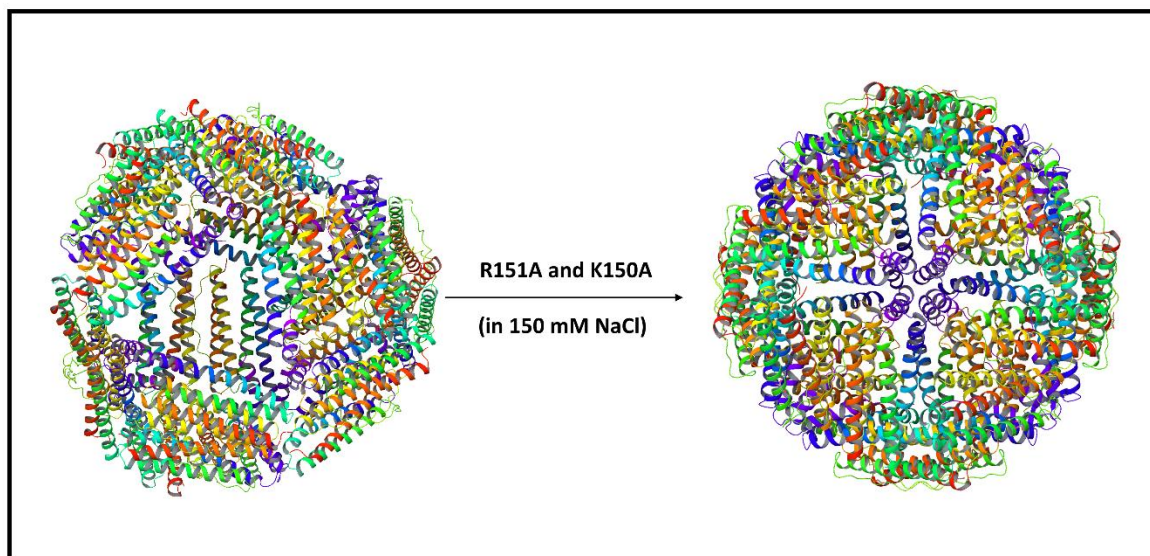
Figure 1.3.2. AfFtn pore closure. The conversion of tetrahedral AfFtn (1S3Q) to octahedral AfFtn (3KX9) upon mutation of R151 and KI50 in the E-helix. The Figure was made with Maestro (Schrödinger LLC) software.

1.3.5. Ferritin Iron Homeostasis

Iron is one of the most abundant elements in the earth's crust, and its ability to exist in different oxidation states makes it important in biological systems. Iron has a variety of functions which include its presence in metalloenzymes involved in metabolism,[70] yet it can be toxic.[71] Despite it being an essential biological factor, a few organisms have evolved that do not require iron, yet flourish in little or no iron environments. An example is *Lactobacilli*, one of the common bacteria in infant gut.[72] Also, the *Borrelia burgdorferi* pathogen can cause Lyme disease at cellular iron concentrations since it possesses iron-independent pathogenesis unlike the vast majority of pathogens which require iron depletion before infecting their host.[73] The spontaneous aerobic conversion of Fe(II) to the highly insoluble Fe(III) at neutral pH (10^{-18}, pH 7)[8] in biological systems leads to Fe(II) deficiency. Iron depletion/ deficiency has been implicated in early embryonic death in mice[74] as well as in Parkinson's disease where the unusual distribution of iron in patients' brain was attributed to iron-mediated oxidative stress.[75]

Even though iron is essential for life, it is necessary to regulate cellular iron concentrations due to its limited solubility and toxicity. [36b] Hence, many organisms have developed types of machinery such as iron transporters and iron storage compartments for effective iron regulation. Ferritins are forms of iron storage compartments in archaea, bacteria, and eukaryotes. They store iron in a bioavailable ferric oxyhydroxide form to prevent reactive oxygen formation (superoxide, hydrogen peroxide, hydroxyl radical). A Fenton reaction involving Fe (II) with hydrogen peroxide (H_2O_2) generates hydroxyl radicals while the superoxide and peroxide species can undergo a Heber-Weiss reaction to generate additional reactive hydroxyl radicals (equation 1). Excess amounts of free radicals could lead to a self-propagating chain reaction that destroys lipid membranes (lipid peroxidation) and hinders other processes such as cell viability, enzymatic activity, and antioxidant defense mechanisms.[76] Another enzymatic regulatory system is the DNA binding protein from starved cells (Dsp). Dsp has a ferritin-like fold of 12 identical subunits with a hollow interior for storing iron. Dsp has a DNA binding motif and an iron sequestration mechanism that prevents iron-induced DNA damage.[17]

$$Fe^{3+} + O_2^{\bullet-} \longrightarrow Fe^{2+} + O_2$$

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + {}^-OH + {}^\bullet OH \quad \text{(Fenton reaction)}$$

$$O_2^{\bullet-} + H_2O_2 \longrightarrow O_2 + {}^\bullet OH + {}^-OH \quad \text{(Haber-Weiss reaction)}$$
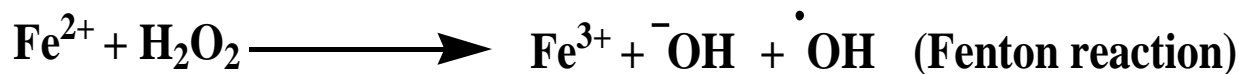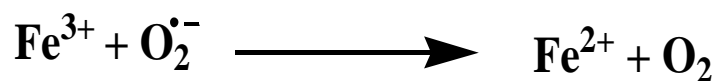
Equation 1: Fenton and Haber-Weiss reactions. Drawn in ChemDraw

1.3.6. The Acquisition and Release of Iron from a Ferritin Cage

1.3.6.1. The Ferroxidase Center

The conversion of ferrous iron to ferric iron in ferritin occurs in the catalytic ferroxidase center (FC) which is located in the middle of the four α-helical domains of each subunit (illustrated in Figure 1.3.3a). The FC of heteropolymeric ferritins (for example, horse spleen ferritin, HoSF) is located in the heavy chain (H subunit) while the light chain has been reported to show pH-dependent iron uptake with non-detectable Fe (II) oxidation. Also, the rate of Fe (II) oxidation is similar to background oxidation of Fe (II) by molecular oxygen, suggesting that the light chain has no ferroxidase center.[77] The catalytic center has two iron binding sites (A and B) that form the dinuclear iron-binding site in most ferritins,[78] but a third site C has been observed close to A and B towards the interior surface of some ferritins (Figure 1.3.3 b and c)

Due to the variations found in the structure of the ferroxidase centers, these centers are classified into the H-chain-type ferroxidase centers, bacterial ferroxidase centers, and the bacterioferritin ferroxidase centers. The FC of eukaryotic ferritins, for example, BfMF, shows site A and B iron atoms pentavalently coordinated with the site A iron atom coordinated by three oxygens (bidentate Glu 23 and monodentate Glu 58), one water molecule and one nitrogen from His61 whereas the B site iron is coordinated by four oxygens (Glu103, Asp140, Gln137, and Glu58) and one water molecule. The Glu58 residue bridges site A and B. Eukaryotic FC binds the

native Fe (III) and other metals such as Tb (III), Mg (II), Zn (II) and Cu (II). These metal ions affect the coordinating ligands at the C site as well as its conformation relative to the sites A and B. Lawson *et al.* observed that Tb(III) at site C of HuHF is coordinated by Glu61 and Glu64,[16] His57 replaces the Glu 64 residue in the presence of Zn(II) (human H-type apoferritin)[79] and Glu140 for Mg(II) bound in soybean ferritin.[80] Mutation of the residues that coordinate a Mg(II) ion at site C resulted in a decreased rate of Fe(II) oxidation indicating that site C is an essential component of the ferroxidase center.[80]

Bacterial ferritin FC closely resembles the H-chain ferroxidase center, for example, the site A in *Chlorobium tepidum* ferritin (CtFtn) is coordinated by Glu17, His53, and Glu50 that bridges to the B site. The B site iron atom is additionally coordinated by Glu94 and second bridging glutamine (Glu30) from site B to C which is further coordinated by Glu49, Glu126, and Glu129.[81] This type of coordination is highly conserved in the hyperthermophilic ferritins, such as that the *Archaeoglobus fulgidus* ferritin and the *Pyrococcus furiosus* ferritin.[8,60] The Fe-Fe distance of AfFtn is 3.18 Å for site A and B whereas site C is situated 6.3Å from this center(Figure 1.3.3 c).[82] The levels of metal bound at sites A, B and C could be affected by the crystallization conditions such as the pH of the crystallization solutions which would affect the protonation/deprotonation state of the coordinating functional groups.

Unlike the two FC sites described above, bacterioferritins have a symmetrical ferroxidase site similar to that observed in the methane monooxygenases and ribonucleotide reductases (Figure 1.3.3 b).[83]The symmetrical ferroxidase center of Bfr contains a dinuclear iron and a site C (not observed in the PDB:1Bfr crystals in Figure 1.3.3 b).[84] Unlike the other ferroxidase centers there are two bridging glutamates (Glu51 and Glu127) between sites A and B. The A site iron atom is coordinated by Glu18/His54 pair and the B site iron atom is coordinated by Glu94/His130.[78] The dinuclear sites are 3.6 Å apart,[85] more extended than the observed prokaryotic ferritins and *Archaeoglubus ferritin* above, but similar to other heme-containing ferritins (3.7 Å in *D. desulfuricans* and 3.5 Å in *A. vinelandii*).[86]
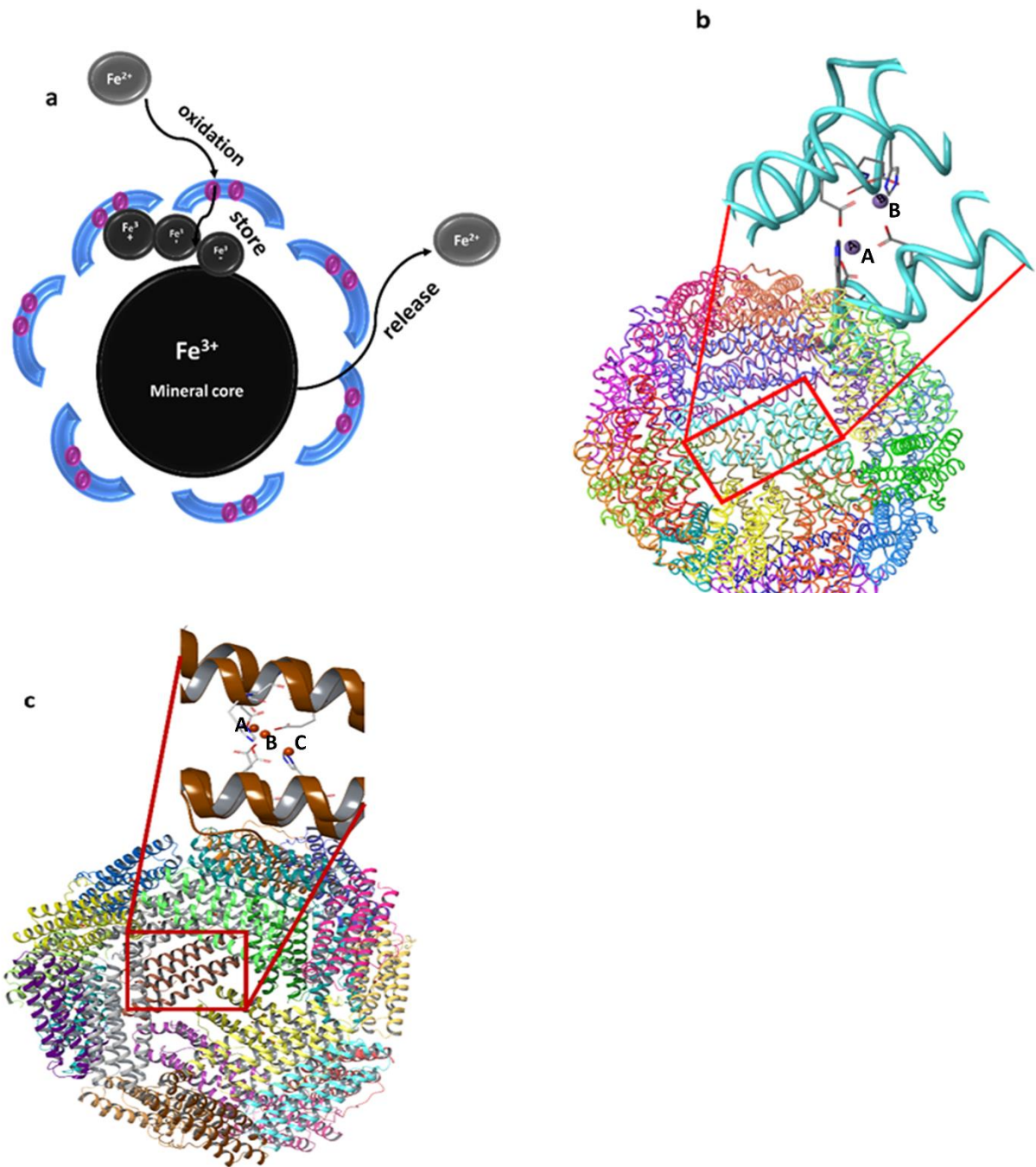
Figure 1.3.3. Ferritin Fe (II) oxidation. (a) A schematic representation of ferritin iron homeostasis. (b) The ferroxidases center, located within Bfr (PDB:1BFR) subunit (cyan,in red box) with 2 Mn^{2+} bound expanded (c) The ABC type ferroxidase site observed in AfFtn crystals (PDB:1SQ3), all 3 Fe binding sites are occupied by Fe^{2+} ions. The expanded subunit(in brown) illustrates the arragements of the 3 Fe atoms at the ferroxidase sites. Figure generated with Maestro (Schrödinger LLC).

1.3.6.2. The Pathway of Fe (II) to and from the Ferroxidase Center

Small molecules and charged ions are believed to enter the cage interior via a diffusion gradient. This gradient exists due to the sequence of amino acids lining the pores that are different from the inner cavity and are believed to aid in propelling these molecules and ions. For instance, Fe^{2+} can diffuse in and out of the cavity during Fe homeostasis through the 3- and the 4-fold axis of symmetry. However, these axes of symmetry are not aligned with the ferroxidase center, located in the 2-fold axis of symmetry of a subunit. So, it has become necessary to find the exact route of Fe (II)/Fe (III) to and from the cavity. This has led to the proposal of three possible mechanisms for Fe(II) delivery to the ferroxidase center: (1) The existence of a ferroxidase channel, a 1-fold axis of symmetry, that was proposed by molecular dynamics to show a direct connection of the catalytic center to the outside environment.[87] (2) A Fe(II) chaperone which induces a conformational change in the ferritin that results in local conformational changes in response to Fe(II). This conformational change may lead to ferritin channels that could be used for translocation. The conformational change is similar to the process of methane and molecular oxygen transfer in methane monooxygenase (MMO). (3) Human poly(rC)-binding protein (PCBP1) may act as a chaperone for Fe(II) delivery to ferritin based on the increased levels of Fe(III) mineral from the interaction between human PCBP1 and ferritin in yeast expression systems, but this interaction is absent in human cell lines.[88] Alternatively, channels within the cage architecture could be the main route for Fe(II) acquisition. Analysis of molecular dynamics simulations undertaken on human ferritin iron uptake indicate that the ferroxidase site directly connects to the exterior via the ferroxidase channel. Also, Fe(II) reaches the FC by interacting with two strong (Asp131 and Glu 134) and two weak binding sites (His111 and Cys130) found in the 3-fold channel.[87] Moreover, chemical modification of Cys126 located at the 3-fold channel in horse spleen ferritin resulted in decreased Fe(II) oxidation[66] and the crystal structure of iron-loaded bullfrog M-type ferritin (BfMF) revealed Fe(II)/Fe(III) in both 3 and 4-fold channels.[89] The B-pore channel at the junction of three bacterioferritin subunits is also involved in Fe(II) entry.[88] Even though the exact mechanism of Fe(II) entry remains unclear, Fe(II) may simultaneously reach the ferritin core through several channels and subsequently to the C site where it later diffuses to the remaining sites in the ferroxidase center. The involvement of this site in iron entry and exit is observed in a change in the conformation of its ligands upon Fe (III)

binding in *Archaeoglobus fulgidus* ferritin (AfFtn) and the diminishing rate of Fe (II) oxidation in *Pyrococcus furiosus* ferritin upon mutating the coordinating ligands.[8]

To answer the question as to how the Fe (III) is stored at the ferroxidase center, Ebrahimi *et al.* proposed the Fe(III)-Fe(II) displacement mechanism for ferritins.[90] This mechanism suggests that Fe(III) remains in a metastable state at the oxidation site until an incoming Fe(II) displaces it. This differs from the accepted substrate model which states that Fe (III) spontaneously leaves the oxidation site and proceeds towards the nucleation site after oxidation. This mechanism is evident in the observation of Fe (III) in the ferroxidase center of eukaryotic (BfMF)[89] and bacterial ferritin, *Chlorobium tepidum* ferritin (CtFtn).[81]

The exciting structure and function of the ferritin cage provide an extremely interesting structural model system for biotemplating. The subsequent sections are based on two ferritins from microorganisms, *E. coli* bacterioferritin (Bfr) and the thermophilic ferritin from the archaea *Archaeoglobus fulgidis* (AfFtn). Their structural properties and potential biotemplating strategies are explored.

1.4. Detail Description of the Ferritins AfFtn and Bfr

1.4.1. *E. coli* Bacterioferritin (Bfr)

E. coli bacterioferritin (Bfr) is the iron storage protein in *E. coli.* It is a maxi ferritin composed of 24 identical subunits. A heme cofactor is sandwiched between two neighboring subunits (Figure 1.4), making up a total of 12 heme cofactors per intact bacterioferritin. The association of identical subunits results in a relatively large protein (~ 450 kDa). Additionally, each subunit interacts with five adjacent subunits to form a 432-symmetry structure (Figure 1.4).[63a] Also, the interaction of subunits generates 3-fold and 4-fold axes of symmetry that are used for Fe (II) transport.

The Bfr subunit consists of 158 amino acids having a molecular mass of 18,945 Da with low proline content and a resultant 80 % alpha-helical content.[59] The Bfr structure closely resembles the ferritins from animals and plants (Horse spleen ferritin, HoSF, human ferritin, HuFtn). However, only 20 % of the amino acids of Bfr are similar to these ferritins.

The roughly spherical Bfr structure has an interior diameter of 8 nm and an exterior diameter of 12 nm. The association of homopolymeric Bfr subunits results in the N-termini of each subunit on the exterior surface and C-termini exposed on the interior surface. Similarly, the self-assembly of subunits generates 2, 3, and 4- fold axes of symmetry that connects the interior to the exterior surface of the Bfr cage. In addition to these channels, a B-pore is formed which is found only in Bfr. The B-pore in combination with the 3- and 4- fold channels are proposed for iron and small molecule entry/exit in and out of the cage.[88]

The 24-mer capsid in ferritin is a copolymer of two chains, heavy and light chains, but Bfr is a homomeric assembly of 24 heavy chains. However, The ferritin from *Pseudomonas aeruginosa* is the only bacterioferritin that contains two different chains (α and β).[61] The significant difference between Bfr and other ferritin members is the presence of the twelve protoporphyrin IX Fe (II) (heme) per intact capsule. These heme molecules are intercalated between every two subunits, and they are coordinated by two methionines from each of these two subunits. Moreover, the heme is not accessible to the exterior environment due to a cluster of four water molecules above the heme cofactor, and two extended loops at the dimer interface.[91] The heme chromophore is responsible for the characteristic absorbances at 417, 530 and 560 nm that is not observed for ferritins, but similar to the reduction spectrum of cytochrome b_1 hence it was initially thought of as a cytochrome with ferritin-like structure.[59] The same spectral pattern has

been observed for other bacterioferritin family members such as bacterioferritins from *Rhodobacter capsulatus* (Rc) Bfr[63b], *Azotobacter vinelandii,* and *Pseudomonas aeruginosa.*[59] Finally, bacterioferritin contains iron at three different locations: (1) the ferroxidase center, (2) the heme cofactor, and (3) at the interior cavity where iron is mineralized. In contrast, the classical ferritins and mini ferritins have iron in the ferroxidase center and interior cavity.

Bfr's oxidation/reduction chemistry occurs at the ferroxidase site[92] that is located in the four-α-helix bundle of the subunits (Figure 1.3.3b). The oxidation/ reduction processes protect the organism against Reactive Oxygen Species (ROS). This process occurs at the ferroxidase site as previously described. However, the exact mechanisms of iron acquisition, oxidation, and subsequent nucleation are unknown. The *in vivo* core demineralization in bacterioferritin has been described and a bacterioferritin-associated ferredoxin (Bfd) binds to the 2-fold channel on top of the heme molecule in ferritin. This brings the Bfd's iron-sulfur cluster, [2Fe-2S], close to the heme group, allowing for the transfer of electrons from [2Fe-2S] to the heme molecule. The electrons are then transferred to the inside of the Bfr cavity, close to the iron mineral and subsequent reduction of the mineral core.[93,94] This then reduces the Fe^{3+} to the soluble Fe^{2+} form, which is believed to diffuse out through the protein pores.
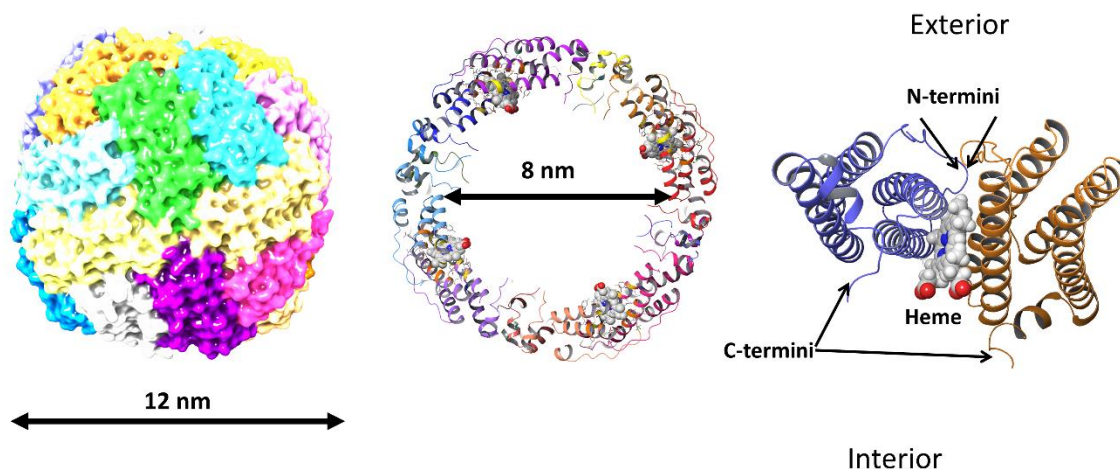
Figure 1.4. Structural features of Bfr. The surface diagram shows the exterior 12 nm dimensions, and the interior dimension is shown in the cartoon. The positions of the N- and C-termini are on the exterior and interior, respectively. All images were generated the 1Bfr crystal structure with the Maestro software (Schrödinger, LLC).

1.4.2. *Archaeoglobus Fulgidus* **Ferritin (AfFtn)**

The ferritin from the sulfate-reducing marine archaeon, *Archaeoglobus fulgidus,* is a nonheme-containing ferritin with 24 identical subunits. Monomeric *Archaeoglobus* ferritin is composed of 160 amino acids (approximately 20 kDa) and has a predominantly alpha helical secondary structure which folds into the typical four alpha-helix bundle structure, similar to other maxi ferritins. Despite the structural similarity, AfFtn shares only 31 % sequence identity with human H ferritin and 38 % with *E. coli* bacterioferritin. The AfFtn subunit and overall fold are structurally identical to bacterioferritins and eukaryotic ferritins except it forms a tetrahedral 23-point group symmetry arrangement instead of the 432 octahedral symmetry. Other interesting characteristics of AfFtn is its hyperthermostability, and its assembly/disassembly process which is mediated by variation of salt concentration.[8]

The roughly spherical AfFtn has four large triangular pores which are due to the presence of Lys150 in the E helix, compared to glutamate and leucine in human and *E. coli* ferritins (Figure 1.4.1). The Lys150 residue associates with neighboring residues at the four-fold interface. However, the presence of the positively charged residues Lys150 and Arg151 abolishes the four-fold interface and forms four large triangular pores in AfFtn.[8] Also, The AfFtn E-helix is longer than the Bfr E-helix as seen in Figure 1.6, while the loop sequences connecting the subunits are shorter in AfFtn than Bfr except for the 19-residue B-C loop. The shorter loop sequences further compact the AfFtn structure and enhance its thermostability. Another thermostabilizing feature of the thermophilic enzyme is reducing solvent accessibility.[95] This phenomenon occurs by the formation of higher oligomers, hexamers, in maxi ferritins. Hexamer formation in AfFtn brings the amine side chain of Lys 150 close to hydrogen bond with the carbonyl oxygen in Met111 and further stabilizes the tetrahedral symmetry.[8] The mutation of Lys150 has been found to collapse the triangular pores and results in the formation of an octahedral (432) symmetry cage identical in overall shape to the Bfr and HuHF ferritins as seen in figure.[69]

Iron oxidation/reduction in AfFtn is proposed to take place at its ferroxidase center. The AfFtn catalytic center is capable of binding three iron atoms similar to bacterioferritins with sites A and B situated 3.18 Å apart, and site C is 5.96 Å from the site B iron atom, towards the cage interior (Figure 1.3.3c). The A and B site iron atoms have Bfr-type coordination from glutamates and histidines. Lastly, the negative surface potential of the AfFtn exterior, similar to other

ferritins,[55] has been proposed to direct incoming Fe^{2+} through the triangular pore as well as other channels (B-pore and 3-fold) in the structure.[8]
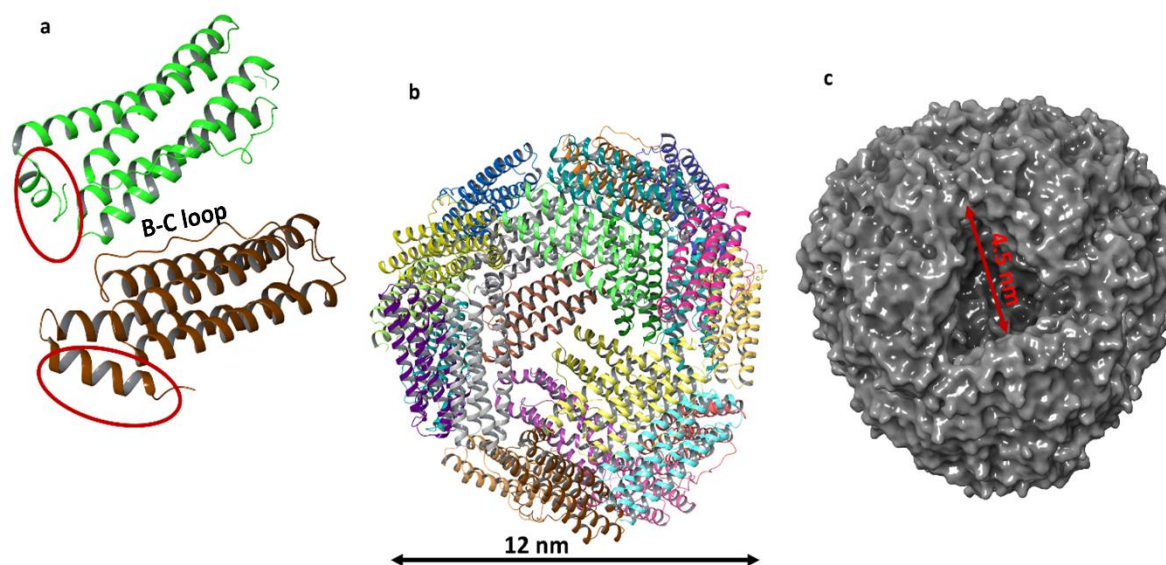
Figure 1.4.1. Structural features of Archaeoglobus ferritin (AfFtn, PDB ID 1SQ3). (a) AfFtn subunit (brown) is similar to a Bfr subunit (green) except for the longer E-helix of AfFtn (brown,in red circle) and the length of B-C loop. (b) Cartoon image of AfFtn showing the 12 nm exterior diameter. (c) The surface of AfFtn highlighting the 4.5 nm triangular pore. This Figure was generated with Maestro software(Schrödinger LLC.)

1.5. Biotemplating Strategies

1.5.1. Dissociation and Association of Subunits

Multisubunit assemblies may exist in different structural forms in response to different conditions such as pH, temperature, chaotropes, and ionic strength. For example, the Cowpea Chlorotic Mottle Virus (CCMV) undergoes pH and metal ion-dependent structural changes. The native capsid transitions to the swollen form when the pH is raised from 5.5 to 6.5 (in low ionic strength, I< 0.1 M) and further dissociates into dimers and releases its viral RNA at pH 7.5 and higher ionic strength (I > 1 M). The dimeric CCMV can be reassembled in the presence of novel guest molecules to repurpose the capsid.[96] Similarly, the thermophilic *Archeaeoglobus fulgidus* ferritin exists as intact 24-mers with 23 point symmetry, with a triangular pore, in high ionic strength buffer (I ≥ 150 mM) and dissociates into monomers in low ionic strength buffers (I ≤ 20 mM).[8] On the other hand, the icosahedral bacteriophage, P22, transforms from the empty 58 nm capsid to the mature 60 nm capsid upon incubation at 65 °C for 10 minutes and increasing the temperature to 75 °C results in the loss of several subunits and converts into a so-called "whiffle-ball" (WB) capsid form.[97,25b] Both P22 WB shape and triangular pore AfFtn can be used for molecular exchange in and out of the capsids.

Similarly, bacterioferritin subunits can dissociate from the capsule protein under declustering conditions such as decreased pH, higher temperatures, and with the addition of chaotropes (Figure 1.5); removal of the declustering agent causes re-association of the subunits into an intact 24-mer cage. This property makes Bfr an exciting and useful platform for bionanotechnology,[89] and advanced studies in biocatalysis, and materials science.[5] Small molecules can be entrapped into the ferritin cage through the iron channels, but in order to entrap larger, non-native guest molecules it becomes necessary to use these declustering techniques to separate the subunits and later assemble the subunits around the guest molecule as illustrated in Figure 1.5.
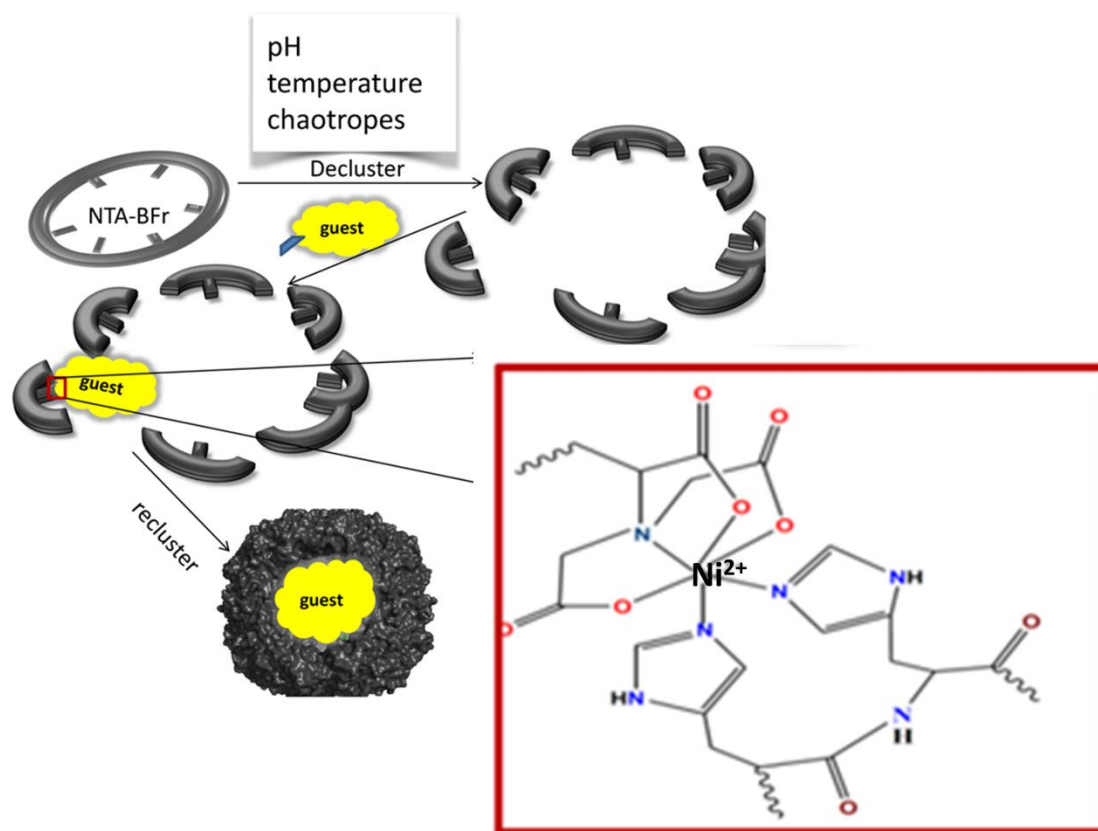
Figure 1.5. The Bfr declustering strategy. Large non-native molecules can be entrapped by dissociation of Bfr subunits under the three conditions to allow the guest to interact with a C-terminal affinity tag. The removal of declustering agents causes the subunits to re-associate controllably and entrap the guest molecule.

1.4.1.2 Guest Stabilization

The guest molecules could enter the cage interior and be stabilized inside the interior surface by a concentration-dependent approach or mediated by affinity interactions. Most metal ion mineralization employs a concentration-dependent approach wherein a high concentration of the guest metal outside the capsule protein is incubated with the intact ferritin and allowed to passively diffuse into the interior through the small pores in the shell after which they are precipitated inside the cavity.[98,30] The limitation of this method is that there could be nonspecific interactions between guest and host ferritin, which could result in a weakly trapped guest molecule and may require excess guest molecules to establish a diffusion gradient. Despite the limitations, this method has been successfully used to entrap intact enzyme, HRP, into the CCMV capsid.[51]

The second method stabilizes the guest by some interaction, either metal coordination or non-covalent interaction. Similar to IMAC protein purification, the insides of a cage can be engineeried with polyhistidine tags that interact with an NTA guest, in the presence of Ni^{2+} or Co^{2+} or vice versa as seen in Figure 1.7. This approach will be used to encapsulate commercial NTA-functionalized gold nanoparticles (5 nm) into both Bfr and AfFtn. Also, a metal-mediated non-covalent interaction was employed in the fabrication of a quantum dot fluorescent mCherry protein complex by the interaction between dihydrolipoic acid on the quantum dot (QD) surface, and a poly-histidine mCherry protein, containing a caspase 3 cleaving site.[99] The arrangement resulted in Förster resonance energy transfer being detected between the QD and the mCherry fluorescent protein, which was abolished upon caspase 3 cleavage to allow for protein sensing.

Alternatively, the host-guest interaction could be mediated by a covalent interaction. A cage protein containing a cysteine mutation was covalently modified with a fluorescent maleimide reagent which resulted in the encapsulation of a fluorescent dye within the cavity of a heat shock protein.[37] The abilities of both AfFtn and Bfr to decluster and recluster paves the way for the use of strategies to *controllably* encapsulate a diverse set of guest molecules such as inorganic nanoparticles, quantum dots, dyes, proteins, and magnetic resonance imaging (MRI) agents. As described above, efficient guest encapsulation can be aided by specific groups on the inner surfaces of the host or guest molecules.

1.5.1.3. Modification of the Heme Cofactor in Bacterioferritin

The heme cofactor in bacterioferritin is ligated to the neighboring subunits via bis-thioether axial coordination to a methionine from each of these subunits. This arrangement exposes the heme molecule to the interior of the cage, but it is occluded from the surface by the subunit protein coat. Hence the propionate groups from all the 12 heme cofactors point towards the interior of the cage where Fe (III) mineralization occurs. Despite the presence of multiple copies of heme cofactors within the Bfr molecule, its exact role is not completely understood. In 1995, Andrew and coworkers demonstrated that the heme cofactor is neither required for 24-mer assembly nor Fe(III) mineralization since the removal of the methionine 52 (Meth 52) that ligates the heme molecule does not significantly affect these processes.[100] However, recent studies, including work in the Honek laboratory, have shown that the heme molecule is involved in the stability of the intact 24-mer quaternary structure (Figure 1.6).[101] Moreover, the heme molecule has been implicated in Fe(III) core demineralization. Yasmin *et al*. demonstrated that the release of iron from the core is independent of the catalytic activity of the ferroxidase center wheras the process of Fe (II) uptake is the rate-determining step for Fe (III) release. The iron release is mediated by the heme cofactor which aids the transfer of electrons from a reductase protein to the bacterioferritin mineral core, hence the addition of small amounts of heme significantly enhanced the rate of Fe(III) reduction and Fe(II) release.[102] The heme cofactor in Bfr can also be used for other applications. The ability to modify and reinsert the heme cofactor into the dimer interface has been explored for H_2 production.[6] Alternatively, the Honek laboratory has demonstrated that other protoporphyrin analogs could be inserted in the Bfr dimer interface and used for modification of the Bfr interior cavity (Figure 1.6.1).[120] The heme can be labeled by first activating the propionate groups, followed by a nucleophilic attack on the activating group by a desired primary amine. This allows for modification of the heme with a range of functional groups. Finally, the heme macrocycle itself can be modified to extend its application.
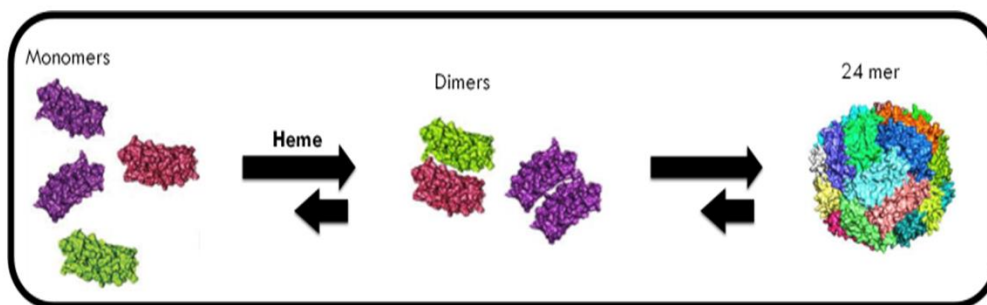
Figure 1.6. Heme-mediated Bfr 24-mer association. Monomeric Bfr subunits spontaneously form dimers upon heme ligation and ultimately produce an intact 24-mer bacterioferritin in solution.
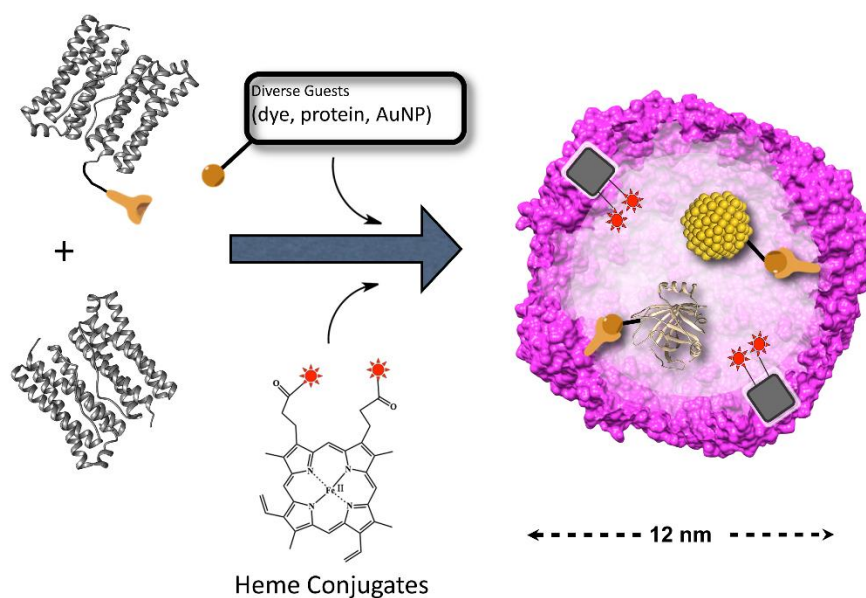
Figure 1.6.1. Bfr heme-mediated guest encapsulation. The heme cofactor has been used in the Honek laboratory to encapsulate guest molecules by synthesizing heme analogs attached to the guest molecule (through His-Ni^{2+}-NTA interaction) and reinserting these modified hemes into the dimer interface to eventually encapsulate the guest. As well as the subunits could be engineered with affinity groups for guest encapsulation.

1.5.2. Surface Modification

One of the most interesting characteristics of both AfFtn and Bfr is the physical separation between the N-termini and C-termini that are located on the exterior and interior surfaces, respectively. A potential application of ferritin is for targeted guest delivery, where the surface contains the targeting moiety that guides the internalized cargo to the desired cell. The targeting moiety could improve the therapeutic success of a drug and minimize the associated side effects of treatment. The exterior surfaces of these proteins could be modified via genetic methods to express cell-specific peptide sequences,[37] to bioconjugate antibodies to the surface via Cys conjugation[4b] and introduce peptide tags that serve as enzyme substrates for post-translational modification.[32,103] The *E. coli* bacterioferritin in this research will be genetically engineered to display surface tags that are amenable to further modifications employing the enzyme transglutaminase. These peptide fusions could be attached to the N-terminal methionine and be useful as long as they do not affect the overall quaternary structure of the ferritin cage. The technique of phage display has been used to identify peptide sequences that have affinity to a host of molecules, cells and surfaces, and their fusion to the external surface of a ferritin could provided additional targeting capabilities to the capsule protein.[104] Recombinant techniques can be used to engineer the ferritins into specific enzyme substrates. This will allow different enzyme manipulations on the ferritin surface provided the required sequences are sufficiently exposed. Another way to introduce surface tags is through chemoenzymatic post-translational modification. This approach labels purified protein with enzyme-substrate acceptor tags which are then irreversibly modified with desired guest molecules in the presence of the specific enzyme. Examples are the acceptor tag for biotin ligase, the LETPG peptide for sortase mediated modification, the 12-mer acceptor peptide for phosphopantetheine transferase (PPTase) and the glutamine tag sequence for transglutaminase-catalyzed reactions.[105,106]

Certain amino acids can be chemically modified by a variety of reagents, provided that the protein has the desired amino acid residue and in the appropriate position. A typical example of this is cysteine modification which has been used extensively for chemical modification, but many proteins including bacterioferritin and the *Archaeoglobus* ferritin lack this residue, and it is necessary to use other bioconjugate and chemoselective techniques that target other solvent-exposed amino acid residues.[107]

Thus, bioconjugation techniques can be used to modify proteins. The proteins can be modified with specialized reagents to introduce affinity labels, catalytic functionalities and modifications useful in targeting applications.[108] The AfFtn and Bfr capsids contain varying amounts of the 20 amino acids except for cysteine. Aspartic and glutamic acids, lysine, arginine, cysteine, histidine, and tyrosine, are frequently used for bioconjugation. The carboxyl groups of aspartate and glutamate can be activated by a carbodiimide reagent, such as 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), or carbonyl diimidazole (CDI). The activated compounds can then be reacted with a desired nucleophilic reagent. Lysines and cysteines can be modified by alkylation and acylation reactions.[109,105]

Secondly, the surfaces of ferritin capsids can be modified with organic and inorganic moieties through bioorthogonal conjugation. Bioorthogonal conjugation also uses genetic engineering to introduce non-native functional groups, unnatural amino acid, on different sections of the protein, unlike genetic fusion tags that are genetically restricted to the termini of the protein sequence.[110,111] Recently, the surface of human ferritin was modified by genetic engineering to introduce the unnatural amino acid 4-azidophenylalanine which subsequently was utilized in click chemistry to immobilize alkynes and small targeting molecules onto the protein surface.[112] Other reagents such as L-azidohomoalanine,[111] *p*-acetophenylalanine (pAcF), and *p*-azidophenylalanine (pAzF) have been successfully used for biorthogonal modification of many proteins.[113] Figure 1.7 below illustrates the chemical modification of ferritin to display diverse functional molecules.
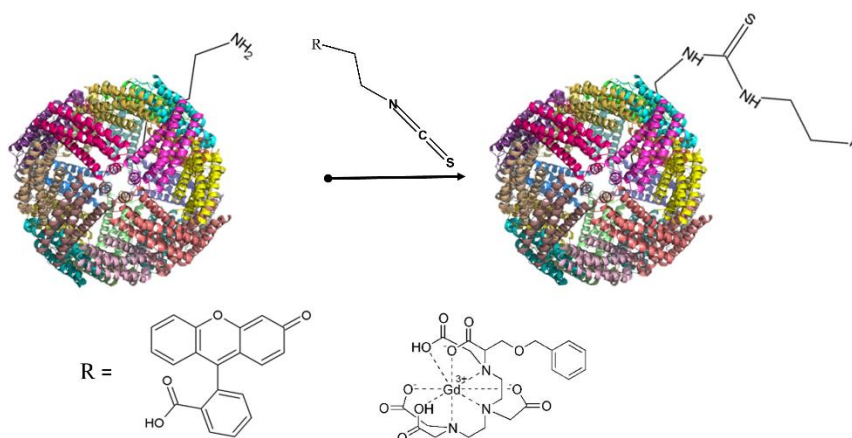
Figure 1.7. Ferritin chemical modification. The amino acids on the surface could be modified via genetic engineering, chemical modification or by biorthogonal approaches to display contrasting agents, antibodies, peptides, and other functional groups. A chemical modification approach is shown.

1.6. Project Goals

1.6.1.Surface Engineering of Bfr and Exploration of Chemoenzymatic Modification of the Eexternal Surface

The ability to overproduce mg to g quantities of several ferritins in different microbial expression systems has increased its applications in science.[114-116,59] Many of the published studies in this area have focused on the interior surface of the cage with less attention focused on developing flexible strategies to modify the exterior surface, even though it is equally amenable to strategic modification. Furthermore, one of the advantages of ferritin-based drug delivery vehicles is its ease of endocytosis by cell surface receptors, but this process suffers from non-specific internalization.[117,118] Hence, it is of interest to explore strategies to engineer the ferritin biomolecule for guided/targeted delivery and imaging applications. Additionally, the surface modification could generate materials with unique characteristics by a careful selection of reagents and ligating groups and expand the application of ferritins as building blocks for the fabrication of new protein-based polymers.[119]

In this project, the Bfr nanoplatform will be engineered into a large enzyme substrate. First, the surface/ external tags will be evaluated for their abilities to convert the Bfr subunits into enzyme substrates. For example, the Qtag-Bfr constructs will be incubated with the TGase enzyme in the presence of a primary amine substrate. The covalent attachment of the primary amine will be analyzed for adducts by employing mass spectrometry, SDS-PAGE, fluorescence spectrometry (for amines with fluorophores), and SEC (and UV-vis absorption) utilizing the associated wavelength of the ligands. After that, the constructs will be characterized for quaternary structures via SEC and their abilities to be reconstituted with heme to form 24-mers similar to HisBfr. Also, the effects of the linker sequence on the 24-mer reconstitution will be evaluated, employing both SEC and TEM. Lastly, the tendency for these constructs to serve as potential "on target" drug delivery vehicles will be investigated by encapsulating guest molecules in the 24-mer capsules and then modifying the surface of the encapsulated capsules or vice versa.

1.6.2. Studies on the Encapsulation of HisGFP within WTBfr and AfFtn

The use of Bfr and AfFtn ferritins in this project further extends the scope of ferritin bioengineering. For instance, the Bfr presents the additional heme cassette as a new strategy to mediate guest encapsulation. Work in the Honek laboratory has focused on bacterioferritin engineering via protein subunit (both exterior and interior) and cofactor modifications (both and *in silico*). The heme cassette has been previously used to encapsulate hydrophobic molecules, dyes, and an intact green fluorescent protein, while subunit modification has led to the encapsulation of 2 and 5 nm gold nanoparticles, streptavidin, and even an upconversion nanoparticle.[120-122]

The two ferritins will be chemically modified with NTA functionalities. For AfFtn, the AfFtnC54 mutant will be used for this application by using maleimide chemistry. The modification will be analyzed by mass spectrometry, and then the NTA-maleimide functionalized AfFtnC54 (AfFtnC54-NTA) will be used for encapsulating His-tagged enzymes. The effects of NTA modification on the 24-mer structure will be studied with SEC. The optimized conditions for forming the 24-mer AfFtnC54-NTA will then be used for encapsulating the enzyme. The main goal of enzyme encapsulation within AfFtn is to further probe the structural significance of the open triangular pore in AfFtn structure, how the pore shields substrates from the entrapped enzyme and the overall protection of the entrapped enzyme.

On the other hand, Bfr will be engineered to encapsulate HisGFP using the heme cassette. The ability to reinsert the heme cofactor into the dimer interface and favor 24-mer formation will be used as an approach to reinsert heme analogs. The heme will be initially modified via NHS ester activation followed by a lysine-NTA modification to form the bisNTA-heme analog. The produced heme analog will be characterized and then reconstituted into WTBfr to produce bisNTA-heme-Bfr. The WTBfr interacts with HisGFP in the presence of Ni^{2+} under declustering conditions. The encapsulated HisGFP will be evaluated for protection against fluorescence quenchers.

1.6.3.Hydrophobic Engineering of Bfr

Engineering of Bfr will be studied to extend the previous work on the hydrophobic engineering of the internal cavity of Bfr so that it can encapsulate hydrophobic guest molecules. The project will also evaluate the contribution of fatty acid length and the number of His-tagged Bfr subunits to wild type subunits to the hydrophobic engineering of bacterioferritin. The following experiments will be described: (i) The original C19-NTA will be encapsulated into Bfr capsules containing 50 % WTBfr and 50 % HisBfr, and a comparison of the sizes of the micelle-encapsulated Bfr to the previous work will be made. (ii) The contribution of fatty acid chain length will be investigated with C14-NTA and C16-NTA. These short chains will be encapsulated into 100 % HisBfr and analyzed with SEC, +ESI-MS, and TEM. (iii) The extent of hydrophobic engineering will be evaluated with pyrene fluorescence and analytical ultracentrifugation.

1.6.4. Bfr Pharmacophore Modeling

Computational docking studies employing virtual compound libraries (over 19 million) to determine if this type of approach can identify molecules to replace the heme group in Bfr and hence to find small synthetic molecules that can control a protein's quaternary structure. The ligands (compounds) identified from docking experiments were obtained from suppliers and reconstituted into the Bfr and analyzed with DLS, SEC, and TEM.

CHAPTER 2: EXPERIMENTAL DETAILS

2. Materials

All buffers were prepared with doubly distilled water (DDH$_2$O) and filtered through 0.2 or 0.45 µm disc filters (GTTP filters from Millipore) depending on the application. Other reagents were obtained as follows: Plasmids (Novogen, Mississauga, ON, Canada); Primers (Sigma-Genosys, Oakville, ON, Canada); Custom gene synthesis (Genescript Inc); Enterokinase, light chain (New England Biolabs, Whitby, ON, Canada); Recombinant Enterokinase (Addgene, Cambridge, MA, USA); Streptavidin, Fluorescein Isothiocyanate conjugate (Streptavidin FITC) (Sigma-Aldrich, Saint Louis, MO, USA); RNase Alert™ Lab kit (Thermo Fisher Scientific, Mississauga, ON, Canada); RNase S-tag peptide (APExBIO, Houston, TX, USA); Subtilisin A. (Sigma-Aldrich, Saint Louis, MO, USA); Hemin (Sigma-Aldrich, Saint Louis, MO, USA); N$_\varepsilon$-(N-(+)-Biotinyl-6-aminohexanoyl)-N$_\alpha$,N$_\alpha$-bis(carboxymethyl)-*L*-lysine tripotassium salt (Biotin-X-NTA) (Biotium, Fremon, CA, USA); Microbial Transglutaminase (MTGase) (ZEDIRA GmbH, Darmstadt, Germany); Maleimide NTA (AAT Bioquest, supplied by Cedarlane, Burlinton, ON, Canada); Ni-NTA-Nanogold™ (Nanoprobes, Yaphank, NY, USA.); TEM grids (Formvar-Carbon 400 mesh; Ted Pella, Redding, CA, USA); Molybdic acid ((NH$_4$)$_6$Mo$_7$O$_{24}$) (Sigma-Aldrich, Saint Louis, MO, USA.)

Chromatographic column and resins: Sephacryl™ S-300 HR and S-200 HR (slurry) (GE Healthcare Bio-sciences, Pittsburg, PA, USA); HiTrap™ 1 mL IMAC HP and HiTrap™ 1 mL benzamidine FF (HS); Superose™ 6 10/300 GL (GE Healthcare Bio-sciences, Pittsburg, PA, USA); Sephadex™ G-10, fine (GE Healthcare Bio-sciences, Pittsburg, PA, USA); UNO™ Q1 1.3 mL monolith ion exchange column and Macro-Prep® High S resin (Bio-Rad Laboratories, Mississauga, ON, Canada.)

2.1. Introduction

This chapter describes the instruments and methods employed in this project. The experimental section begins with protein production, followed by the various approaches employed to modify the proteins and the techniques used to analyze these modified proteins. The experiments in this section focus on engineering two microbial ferritins: *E. coli* bacterioferritin (Bfr), and *Archaeoglobus fulgidus* ferritin (AfFtn) platforms on their inner (Bfr and AfFtn) and outer (Bfr) surfaces. The interior modification is useful for the encapsulation of novel guest molecules such as gold nanoparticles, lipid/surfactant molecules, intact proteins (such as Green Fluorescent Protein (GFP), glucohydrolase, and Streptavidin). Outer surface modification was achieved via the recombinant DNA modification of Bfr N-termini with peptide tags (Glutamine-containing tags recognized as substrates for chemoenzymatic reaction with the enzyme microbial glutaminase or Sortase tags which are peptide sequences that can be utilized by Sortase enzymes for chemoenzymatic coupling with a co-substrate peptide or protein which has a sequence that the Sortase enzyme can react with).

Streptavidin and GFP were utilized as proof-of-concept "guests" to showcase the ability of the Bfr cage to encapsulate more complex biological molecules. Different constructs of ferritins were designed to enable encapsulations of these guest molecules; Streptavidin encapsulation utilized the interaction between a nitrilotriacetic acid (NTA) and His-tag to capture biotinylated Streptavidin-FITC (SF). The protein "host" platforms were genetically engineered to contain a C-terminal His-tag (HisBfr and HisAfFtn) fusion to interact with the biotin-NTA molecule in the presence of nickel (II) ions. The encapsulation of GFP utilized a reverse polarity approach to exemplify how His-tagged "guest" proteins could be encapsulated within the Bfr cavity. The propionate groups of the Bfr heme cofactor were functionalized with Nα, Nα-bis(carboxymethyl)-*L*-lysine and this modified heme reconstituted into wild-type Bfr (WTBfr) to interact with the His-tagged GFP during the encapsulation procedure. In contrast, AfFtn was designed to utilize an interior Cys-NTA moiety to interact with the HisGFP during the encapsulation procedure. The Cys-NTA was the result of the chemical modification of a cysteine residue (site-directed mutation; the sole cysteine residue in AfFtn) by the reagent N-(5-(3-maleimidopropylamido)-1-carboxy-pentyl) iminodiacetic acid (abbreviated as maleimido-C_3-NTA).

Encapsulation of hydrophobic molecules within the Bfr cavity was explored by utilizing NTA-functionalized lipids: nonadecanoyl-NTA (C19-NTA), palmitoyl-NTA (C16-NTA), and myristoyl-NTA (C14-NTA) which also interacted with His-tagged Bfr via nickel (II) ion-mediated affinity interactions. The lipids were encapsulated via cage-mediated micelle formation, wherein the surfactant molecules (or complex micelles) were made to interact with declustered His-tagged Bfr subunits followed by heme-mediated reclustering to the 24-mer capsule which constrained the hydrophobic surfactants to the Bfr interior to form a micelle or a hydrophobic aggregate. Conversely, a preformed micelle with known dimensions could also be encapsulated via the same NTA-Ni^{2+}-His interaction.

The encapsulation of inorganic complexes for applications in imaging and diagnostics was studied using 5 nm gold nanoparticles (AuNP). The AuNP particle has NTA functionalities exposed on the surface through a proprietary polymer (~ 1.8 nm long) which is attached to the outer surface of the AuNP. The HisBfr and HisAfFtn constructs interacted with the AuNP similarly to the micelle. All the encapsulated guest molecules within the ferritins were studied employing transmission electron microscopy (TEM), and size exclusion chromatography (SEC) was employed to monitor specific wavelengths associated with the guest and ferritin molecules. Fluorescent guest molecules were further analyzed utilizing fluorescence spectroscopy, and fluorescence quenching experiments were used to confirm the protection of the guest molecules within HisBfr from collision quenching agents such as iodide. Mass spectrometry was employed to obtain the relative masses of the guest and host materials in solution and to also confirm the presence of the guest molecule within the SEC-purified ferritins.

The ability to remove and reinsert the native heme cofactor back into the Bfr dimer interface and eventually, due to reclustering, within the intact ferritin, allows for the heme moiety to act as a "cassette or module" to add additional complexity and capabilities to the bacterioferritin scaffold. The heme cofactor has been observed to play a significant role in the formation of an intact Bfr 24-mer.[120] It has been reported to play a role in Fe^{3+} reduction. [102] The heme-mediated 24-mer formation was studied further by replacing the natural heme with commercial analogs and analyzing how they affect the formation of the Bfr quaternary structure. A protoporphyrin IX molecule and two others: Protoporphyrin IX dimethyl ester, Protoporphyrin IX Zn (II) were utilized in these experimental studies. Also, computational docking calculations were undertaken, using Molecular Mechanics (MM) simulations, to identify additional molecules

that might be capable of binding within the heme pocket; the Maestro computational software [123] was used for this purpose. The applications used to provide detailed binding information between the Bfr receptor and molecules in two commercial databases from Schrodinger. These were the CAP2011 and the Aldrich Market Direct® Phase Database. Molecules predicted to have very good binding affinities (and a few weak binders) were purchased and reconstituted into WTBfr to study if the compounds could also act in a similar fashion to the natural cofactor heme to stabilize the 24-mer quaternary structure (Chapter 4).

Exterior surface modification of Bfr was achieved by extending the N-terminus with sequence tags which allowed Bfr to undergo enzymatic modifications on its surface to attach various compounds to its surface. One tag was also employed that would decorate the surface with S-peptide, a fragment of the RNase S protein, which upon reconstition with the RNase S protein, should reconstitute the protein and result in reactivation of the RNase enzymatic activity on the surface of the ferritin. The Qtags were used for transglutaminase-catalyzed reactions on the ferritin exterior surface. The Qtag1 and Qtag2 peptide sequences are double duty constructs since they are both reported to be MTGase substrates as well as capable of reconstituting RNase S activity when bound to the RNase S protein.[106,124] Qtag3 and Qtag4 Bfr both have the MTGase substrate sequence reported by Lee and coworkers[103] but with slightly different $(Gly)_n$ linker lengths (Figure 2). A sortTag-Bfr construct was also designed to enable Bfr to undergo sortase-mediated modifications. Additionally, the construct containing an N-terminal deca-Histidine tag followed by an enterokinase proteolytic site and followed by the N-terminus of the Bfr (termed the HTEKBfr construct, figure 2c) was intended to aid in the preparation of a form of the wild-type Bfr (termed HWTBfr) after subsequent protein expression followed by metal affinity chromatography and then enterokinase incubation to cleave the His-tag sequence from the N-terminus of Bfr. This approach was taken to address initial challenges with the expression and isolation of wild-type Bfr from a plasmid that coded for this form of the protein.

Tag	Sequence
QTag1	MKETAAAKFERQHMDSPDLH
QTag2	MKETAAAKFERQHMDSGGGG
QTag3	MERLQQPTGGG
QTag4	MERLQQPTGGGG
SortTag1	MLVPRGGGGG
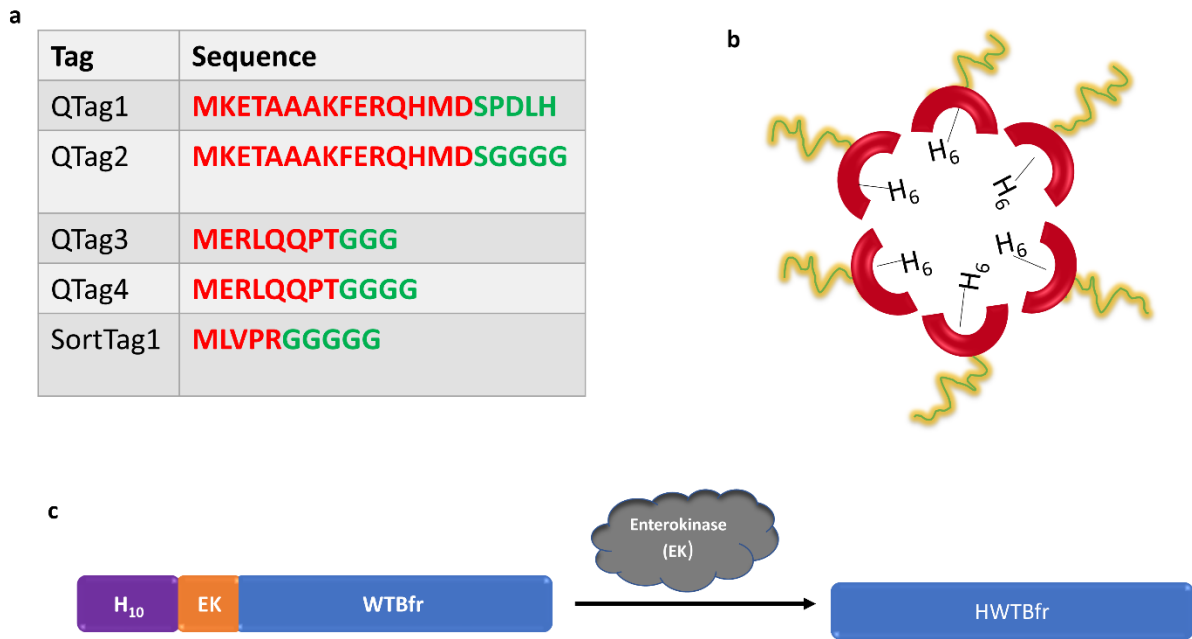
Figure 2. The Bfr constructs designed for surface modification and Wild-type production. (a) The original HisBfr construct with C-terminal H_6 tags were recombinantly modified with five different N-terminal surface tags into five new Bfr constructs, shown in the table. The red segment is the tag sequence and green is the linker sequence. (b) A representation of the constructs with both exterior and interior tags. The red shape represents the HisBfr subunit with H_6 tags to the interior and yellow N-termini surface tags. (c) Shows the production of HWTBfr from the HTEKBfr construct having an enterokinase recognition sequence.

Additionally, the thermostable ferritin from the archaea *Archaeoglobus fulgidus,* abbreviated as AfFtn, was explored in this work. The primary objective was to encapsulate a glycohydrolytic enzyme within the interior cavity of this capsule protein (~ 12 nm in diameter) and verify it it is protected against heat and explore its ability to hydrolyze a small substrate which could diffuse through the host's large triangular openings but block larger substrates from being hydrolyzed. This work would be a proof-of-concept that the substrate profile of an enzyme could be controlled by the porosity of the nanoparticle cage surrounding it. The protein was characterized similarly to the steps used to characterize HisBfr due to the similarity in structure (except that the triangular opening in AfFtn is larger than the narrow pores found in

43

bacterioferritin). Streptavidin and AuNP were encapsulated within the interior of the AfFtn cage to understand how it might accommodate guest molecules. Previous work in the Honek laboratory has shown the protection of Bfr-encapsulated Streptavidin-Fluorescein Isothiocyanate (FITC) conjugate, abbreviated herein as Bfr-SF, from iodide fluorescence quenching. However, it is envisaged that a collisional quencher such as iodide should be able to quench the fluorescence of the guest entrapped within AfFtn due to the large triangular openings in the AfFtn host and the ability of iodide to freely diffuse through it.
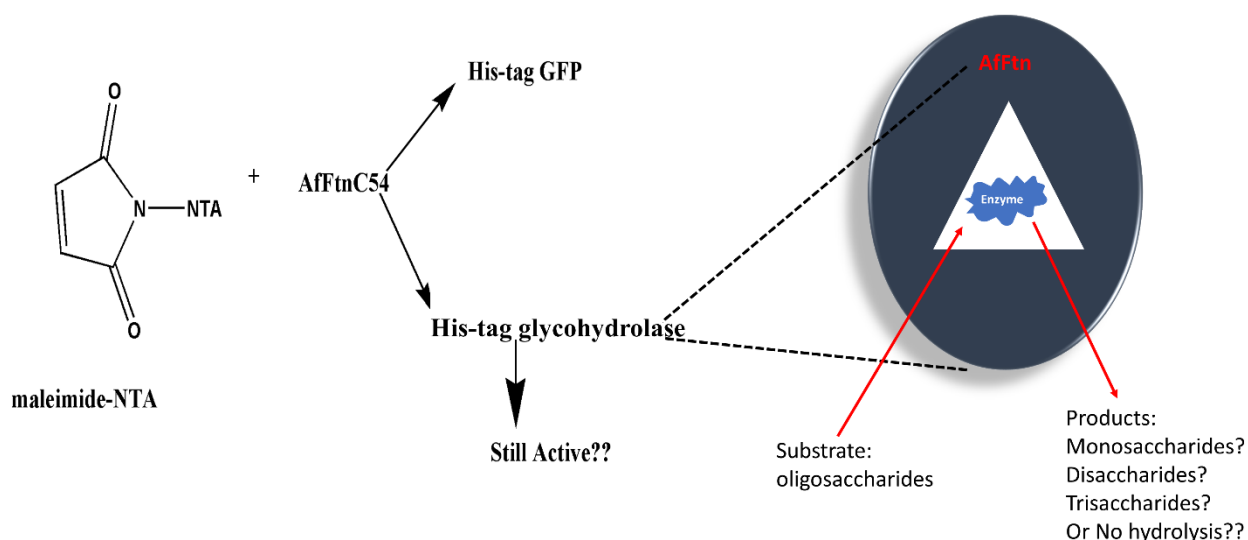
Figure 2.1. A representation of *Archaeoglobus fulgidus* ferritin (AfFtn) catalytic container.
    The guest molecules are encapsulated by employing an AfFtn cysteine modification(AfFtnC54) which is chemically modified to attach an NTA affinity tag. Alternatively, and not represented in this figure, streptavidin-FITC can be encapsulated by metal affinity interactions using the C-terminal His-tagged AfFtn. The encapsulated guest molecules will then be analyzed for their protection (GFP and Streptavidin-FITC) from quenching agents or their catalytic effects (encapsulated enzymes such as glucohydrolase).

44

2.1.1. Design and Expression of BFr Constructs

The original DNA/protein sequences for wild-type bacterioferritin (WTBfr) and the C-terminal His-tagged fusion version of bacterioferritin (HisBfr) were designed by Dr. Uthaiwan Suttisansanee, and Dr. Elizabeth Daub and the DNA and protein sequences for both of these constructs are shown in Table 1 and Appendix 8, respectively. Additional engineered versions of the Bfr protein were custom gene synthesized by Genscript (Piscataway, USA) and designed by Professor Honek. The constructs that were designed were: Qtag1-Bfr, Qtag2-Bfr, Qtag3-Bfr, Qtag4-Bfr, and Sorttag1-Bfr the tag sequences are in Figure 2, and the entire protein sequence is presented in Table 1. WTBfr and HisBfr proteins were overproduced from ampicillin resistance - containing pET-22b (+) plasmids, which contain a *lac* operon promoter, whereas the kanamycin resistance-containing pET-29b (+) plasmid was used for the Qtag1-Bfr construct. The pET-29b (+) plasmids with S-tag sequence (a subtilisin digest of RNase S protein) at the N-terminus and a C-terminal hexahistidine tag for purification was combined with the Bfr-pET-22b plasmid. The plasmids were then cut with both BglII and Xhol restriction enzymes, which resulted in ligation of the *Bfr* gene to the new vector that provided an S-tag at the N-terminus for surface modification(Appendix 8). The *lac* operon promoter allows for isopropylthio-β-galactoside (IPTG) induction during the production of *E. coli* QTag1-Bfr. QTag2-Bfr was designed similarly to QTag1-Bfr above; the only difference is in the sequence of the linker that ligates the S-tag to the surface of HisBfr (transformed into DH5α cells by Taylor Urquhart). These constructs are transglutaminase (TGase) substrates and contain the S-peptides for RNase S reassembly.[125] The QTag3-Bfr and QTag4-Bfr constructs were designed based on a known TGase substrate.[103] The latter construct used for surface modification was SortTag1, which allows for sortase-mediated enzymatic modification of the outer surface of Bfr.[126] The genes were further isolated, employing a QIAprep Spin Miniprep Kit (QIAGEN, Toronto, ON, Canada) and transformed into competent *E. coli* BL21 cells via heat shock for protein expression. Similarly, three mutants of Qtag2-Bfr (Qtag2E86A, Qtag2K22A, Qtag2K22A-E86A) were designed to study the effects of changing two of the N-terminal residues on the ability of the various N-terminal tags to form the intact 24-mer.

45

Construct	Sequence
WTBfr	M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G
HisBfr	M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
Qtag1-Bfr	**MKETAAAKFERQHMD** S P D L H M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
Qtag2-Bfr	**MKETAAAKFERQHMD** S G G G G M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
Qtag3-Bfr	**MERLQQPT** G G G M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
Qtag4-Bfr	**MERLQQPT** G G G G M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
SortTag1-Bfr	**MLVPR** G G G G G M K G D T K V I N Y L N K L L G N E L V A I N Q Y F L H A R M F K N W G L K R L N D V E Y H E S I D E M K H A D R Y I E R I L F L E G L P N L Q D L G K L N I G E D V E E M L R S D L A L E L D G A K N L R E A I G Y A D S V H D Y V S R D M M I E I L R D E E G H I D W L E T E L D L I Q K M G L Q N Y L Q A Q I R E E G **L E H H H H H H**
WTAfFtn	M A S I S E K M V E A L N R Q I N A E I Y S A Y L Y L S M A S Y F D S I G L K G F S N W M R V Q W Q E E L M H A M K M F D F V S E R G G R V K L Y A V E E P P S E W D S P L A A F E H V Y E H E V N V T K R I H E L V E M A M Q E K D F A T Y N F L Q W Y V A E Q V E E E A S A L D I V E K L R L I G E D K R A L L F L D K E L S L R Q F T P P A E E E K

HisAfFtn	MASISEKMVEALNRQINAEIYSAYLYLSMASYFDSIGLKGFSNWMRVQ WQEELMHAMKMFDFVSERGGRVKLYAVEEPPSEWDSPLAAFEHVYE HEVNVTKRIHELVEMAMQEKDFATYNFLQWYVAEQVEEEASALDIVE KLRLIGEDKRALLFLDKELSLRQFTPPAEEEK**LEHHHHHH**
AfFtnC54	MASISEKMVEALNRQINAEIYSAYLYLSMASYFDSIGLKGFSNWMRV QWQEEL**C**HAMKMFDFVSERGGRVKLYAVEEPPSEWDSPLAAFEHVY EHEVNVTKRIHELVEMAMQEKDFATYNFLQWYVAEQVEEEASALDIV EKLRLIGEDKRALLFLDKELSLRQFTPPAEEEK
HisGFP	**MASHHHHHH**MVSKGEELFTGVVPILVELDGDVNGHK FSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTF TYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG HKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGS VQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITHGMDELYK
HTEKBfr	**GHHHHHHHHHH**SSGHIDDDDKHMKGDTKVINYLNKLLGNELVAIN QYFLHARMFKNWGLKRLNDVEYHESIDEMKHADRYIERILFLEGLPN LQDLGKLNIGEDVEEMLRSDLALELDGAKNLREAIGYADSVHDYVSR DMMIEILRDEEGHIDWLETELDLIQKMGLQNYLQAQIREEG
HisPfFtn	MLSERMLKALNDQLNRELYSAYLYFAMAAYFEDLGLEGFANWMKAQ AEEEIGHALRFYNYIYDRNGRVELDEIPKPPKEWESPLKAFEAAYEHE KFISKSIYELAALAEEEKDYSTRAFLEWFINEQVEEEASVKKILDKLKF AKDSPQILFMLDKELSARAPKLPGLLMQGGE**LEHHHHHH**

Table 1. Protein sequences for the major constructs used in this project. The highlighted residues represent the Poly histidine fusion tags as well as mutations (AfFtnC54). The N-termini colored segments (red is the tag and green are the linkers) are for surface modification (Qtag and Sort tag) and enterokinase (EK) recognition sequence (HTEKBfr). The decahistidine (purple) segment in the HTEKBfr sequence was used for IMAC purification and the EK recognition sequence (orange) was used for His-tag removal. The corresponding DNA sequences for these constructs are in Appendix 8.

Protein expression was achieved by inoculating a single bacterial colony from an Amp-agar plate (or a Kan-agar plate- 100 ug/mL Amp or 50 ug/mL Kan) into a 1.5 L Luria-Bertani media (LB- 15 g Bio-tryptone, 15 g NaCl, 7.5 g Yeast). Cultures were grown at 37 °C to 0.6 OD_{600}, and protein expression was induced with 1 mM IPTG and incubated overnight with shakng at room temperature. The cells were harvested by centrifugation at 6371 g for 30 min in a Beckman JA10 rotor. The harvested cells were resuspended in 50 mL lysis buffer (100 mM NaCl, 50 mM Tris-pH 8.0, 1 mM PMSF, 1 mM EDTA) and then homogenized at 17000 psi on an Avestin Emulsiflex-C5 (Mannheim, Germany) with cooling. Insoluble proteins and DNA were pelleted by centrifuging at 20,000 g for 40 min in a Beckman JA 25.5 rotor.

2.1.2. Purification of His-tagged Proteins

All H_6-tagged variants (Table 1) were purified by heat shock at 70 °C for 10 minutes and IMAC chromatotraphy. The clarified protein solution to an equilibrated 1 mL Nickel HisTrap™ HP column (Binding buffer: 50 mM Tris-pH 8, 100 mM NaCl, 20 mM imidazole) attached to a BioLogic DuoFlow Chromatograph (BIO-RAD) instrument. The protein was loaded onto the column with binding buffer and washed with about 10X column volumes (CV) until the UV at 260, 280 and 418 nm readings stabilized (and then the OD was zeroed). The bound protein was then eluted with buffer containing a higher concentration of imidazole (100 % of elution buffer: 50 mM Tris-pH 8, 100 mM NaCl, 300 mM imidazole). The molecular weight of the subunits of proteins in the isolated chromatographic fractions was estimated employing a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) protocol. Positive electrospray mass spectrometry (+ESI-MS) was further used to verify the protein monomeric mass. The purified proteins were then quantified using UV-vis absorbance ($\varepsilon_{280} = 21430$ cm^{-1}M^{-1} for Bfr, 33920 cm^{-1}M^{-1} for AfFtn), aliquoted into small storage tubes and then stored at -80 °C for further experiments. The purification steps are summarized in Figure 2.2.
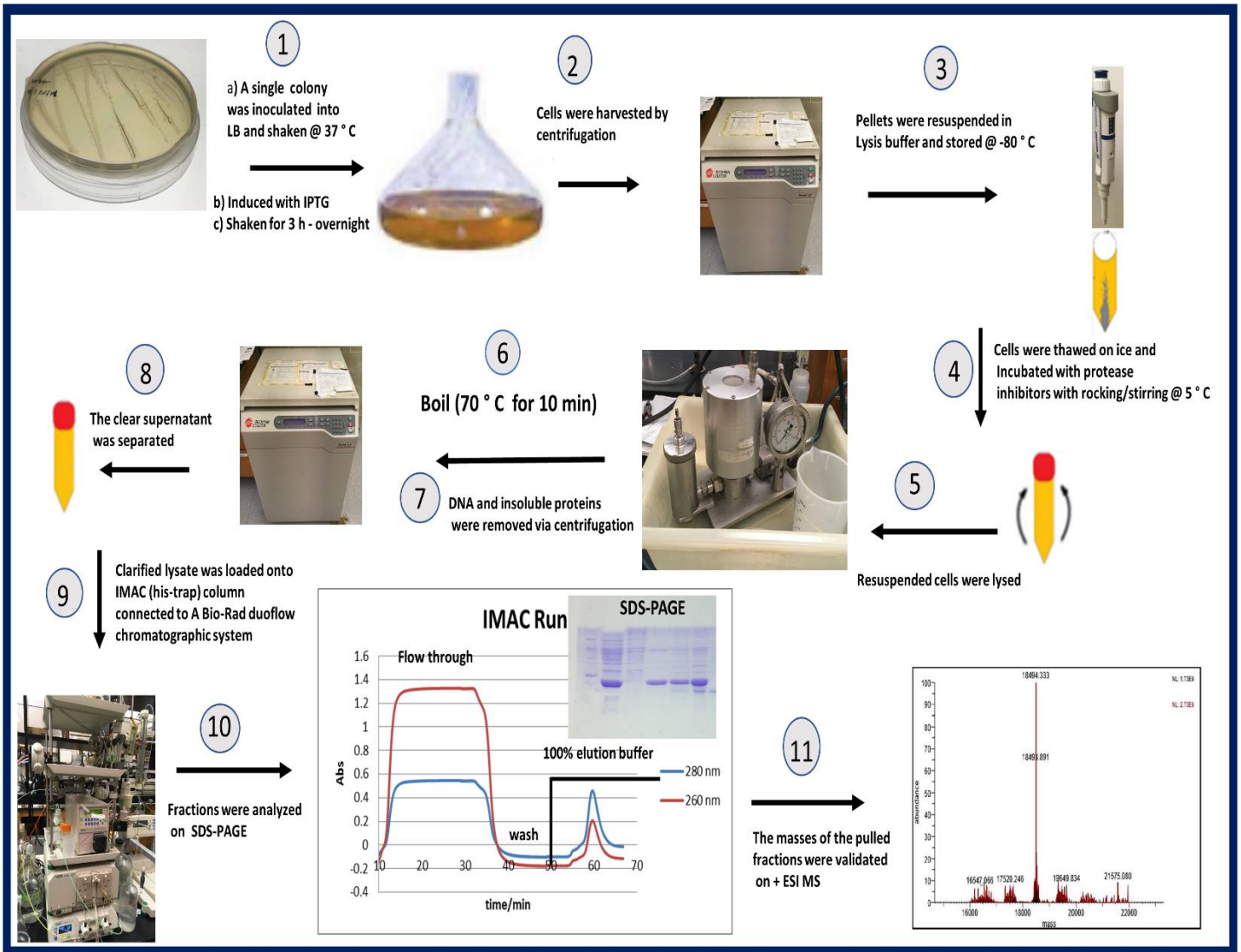
Figure 2.2. Schematic representation of the His-tagged protein overproduction and purification. The WTBfr production only differs at the chromatographic step, where ion exchange chromatography or ammonium sulfate precipitation is employed followed by size exclusion chromatography.

2.1.3. Purification of WTBfr

The WTBfr protein was expressed similarly to the other constructs. However, the purification process was different due to the lack of the C-terminal hexahistidine tag. The harvested cells, from 1.5 L of culture, were resuspended in 100 mL of lysis buffer (50 mM MOPS- pH 7.4, 5 mM $MgSO_4$, 1 mM PMSF) and then frozen at -80 ˚C. The 50 mL thawed cells were incubated with DNase (50 µg) and RNase (50 µg) at room temperature for 30 min and lysed at 17000 psi. The lysate was centrifuged for 1 h at 20000 g, and the lysate was applied to a pre-equilibrated BioRad UnoQ1 column (50 mM MOPS- pH 7.4, 5 mM $MgSO_4$). The column was washed with the same buffer until the absorbance at 269, 280 and 418 nm decreased to 0 OD, bound proteins were then eluted with 100 % of 50 mM MOPS- pH 7.4, 1 M NaCl. The fractions were analyzed on a 15 % SDS-PAGE gel, and WTBfr-containing fractions were pooled, concentrated in 10 kDa cutoff Amicon centrifugation tubes and buffer exchanged into 50 mM MOPS- pH 7.4, 0.5 M NaCl. The concentrate was loaded onto a Sephacryl® S-200 10/300 HR column to obtain pure WTBfr protein, verified by SDS-PAGE and MS. Figure 2.2.1 below shows the generalized purification protocol.

In order to optimize the protein yield, the ion exchange step was substituted with ammonium sulfate precipitation summarized in Figure 2.2.1 below:

Cells harvested by centrifugation, 6 371 g for 30 min

1

Resuspended in buffer A and lysed with emulsiflex, 17000 psi

2

Lysate boiled @70 °C for 10 min (80 °C for WTAfFtn) and centrifuged @ 20 000 g for 20 min

3

Incubate clarified lysate @ 30 °C (30 min) with DNase 1 (500 units/50 mL)

4   5

40 % ammonium sulfate precipitation @ 10 °C_ 15 min   80% ammonium sulfate precipitation @ 10 °C_15 min

4 a   6

Discard precipitate (yellow)   Keep precipitate (Bfr, orange red) (AfFtn, pale yellow)

7

Buffer A = 5 mM MgSO4, 1 mM PMSF, 50 mM Mops-pH 7.5
Buffer B = 50 mM Mops-pH 7.5 , 150 mM NaCl   Resuspend in Buffer B and run SEC on AfFtn

Figure 2.2.1. Protein purification via ammonium sulfate precipitation.

2.1.4. Overproduction and Purification of Enterokinase (EKC122S)

The ampicillin-resistant pET-15b-EK-C122S-His5 bacterial stab (obtained from Addgene, deposited by Skala and Coworkers[127]) was spread onto an LB-Amp agar plate and incubated at 37 ˚C overnight. A single colony was restreaked on to a new LB-Amp plate for another 37 ˚C overnight incubation. Another colony from the second plate was inoculated into a 5 mL LB-Amp broth for DNA extraction. EKC122S DNA was extracted with the Genelute^{TM} plasmid miniprep kit (Sigma-Aldrich) and transformed into competent BL21 cells via heat shock. EK protein was expressed and purified by Evan Shepherdson (Chem494 undergraduate student under my supervision) following the procedure reported by Skala *et al.*[127] with few modifications.

A 1 L solution of LB containing 100 µg/ml ampicillin was inoculated with an overnight culture to a starting OD of 0.05 and shaken (37 ˚C, 220 rpm/min) until the OD_{600} reached 1.0 where it was induced with the addition of solid IPTG (0.5 mM final concentration). The cells were harvested after 3 h of induction in a Beckman JA10 rotor (15 min, 4000 g, 4 ˚C). The harvested pellet was suspended in 50 mM Tris- pH 7 based on its wet mass (3 mL buffer per 1 g of the pellet) and lysed at 17000 psi in an Avestin Emulsiflex-C5 (3 cycles, interrupted by 5 min of chilling on ice). The lysate was washed 3x with 15 mL of suspension buffer I (50 mM Tris-pH 7, 1.5 M NaCl, 60 mM EDTA, 6 % Triton x 100), shaken for 30 min at room temperature and centrifuged at 18000 g. The pellet was washed with 30 mL of suspension buffer II (50 mM Tris-pH 7, 20 mM EDTA) and the inclusion bodies (IB) were removed by centrifugation (18000 g, 20 min, 23 ºC). The residue was washed 2x, and 1 g of IB was dissolved in 20 mL of Solubilization buffer (7.5 M guanidine-HCl pH 9, 50 mM Tris, 100 mM BME), and stirred overnight at 23 ˚C. Undissolved particles were removed by centrifugation at 43000 g for 20 min at room temperature, and the pH of the supernatant was adjusted to pH 3.5 with 5 mM citrate buffer and dialyzed against 5 mM citrate buffer (12.4 kDa cut off) for 4 h. The white fluffy pellet was harvested by centrifugation (18000 g, 20 min, 23 ºC) and resuspended in solubilization buffer (1:10 (w/v)). The EK in the clear solution was refolded in 2 L of refolding buffer (0.5 M arginine, 50 mM Tris-pH 8.3, 20 mM $CaCl_2$, 1 mM EDTA, 5 mM cysteine-HCl, 0.5 mM cystine) for 72 h at 23 ºC. The resulting solution was concentrated by employing 200 mL Amicon concentrators (10 kDa Millipore disc filters), and the concentrate was dialyzed overnight (at 23 ºC) against the storage buffer supplemented with 3 mM NaN_3. The EK protein was analyzed using SDS-PAGE and further purified on a Uno® Q1 column (binding buffer: 50 mM Tris-pH 7.5, 20 mM NaCl,

Elution buffer: 50 mM Tris-pH 7.5, 1 M NaCl) followed by an IMAC column (binding buffer: 20 mM imidazole, 0.5 M NaCl, 50 mM Tris-pH 75, Elution buffer: 200 mM imidazole, 0.5 M NaCl, 50 mM Tris-pH 7.5). The pooled fractions exhibited protein bands from the SDS-PAGE gel corresponding to EK of molecular weight of approximately 26 kDa.

2.1.4.1. Recombinant EK His-Tag Removal

A reaction mixture containing 1 mg /mL HTEKBfr (in 50 mM NaCl, 2 mM $CaCl_2$, 20 mM Tris- pH 8) and 300 nM pure EK was incubated at 37 ˚C for 15-24 h. The reaction was terminated by adding PMSF (to 1 mM final concentration) or exchanging the buffer for distilled water. Samples were then analyzed on both SDS-PAGE gels and mass spectrometry (+ ESI) for the cleaved substrate (no His-tagged protein). Removal of EK and the unreacted substrate was achieved by loading the mixture onto a 1 mL HiTrap™ Benzamidine FF column (in 500 mM NaCl, 50 mM Tris- pH 7.5) and eluting with 500 mM NaCl, 10 mM HCl- pH 2. The eluted fractions were neutralized by the addition of 1 M Tris-HCl- pH 9 (1 mL fraction to 200 µL buffer). The resulting fractions were concentrated and quantified for encapsulation studies. Alternatively, the cleaved protein could be separated by running the mixture through an IMAC column where the recombinant EK binds, and the cleaved protein flows through.

2.1.5. Purification of Other Ferritins

Other members of the ferritin family were also explored in this project: *Archaeoglobus fulgidis* ferritin (AfFtn), the AfFtnC54 ferritin variant, and the *Pyrococcus furiosus* ferritin (PfFtn). These constructs were designed by Dr. Honek and commercially synthesized by Genscript, which included codon optimization. The plasmids AfFtn-pET-22b and PfFTn-pET-22b that encode for the C-terminal His-tagged AfFtn and PfFtn respectively were transformed via heat shock into competent BL21 cells similarly to the other constructs described above (DNA sequence, Appendix 8).

An aliquot (5-10 mL) from a 100 mL (LB-Amp) overnight culture that was grown from a single colony was used to inoculate 500 mL of LB-Amp (40 µg/mL amp) to obtain a starting OD_{600} of about 0.05. The cultures were shaken at 37 ˚C to 0.4 OD_{600} where they were spiked with 40 µg/mL Amp and shaken until 0.6 OD_{600}, and induced with 1 mM IPTG. The cells were then shaken at 25 ˚C overnight and harvested via centrifugation at 6000 g for 40 min. The harvested

cells were resuspended in buffer A [lysis buffer: 25 mM HEPES pH 7.5, 10 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF)]. The resuspended cells were later lysed and purified similarly to the other His-tag constructs above. Buffers for the AfFtnC54 variant were supplemented with 2.5 mM DTT or 2.5 mM BME to prevent Cys oxidation to the disulfide.

2.2. Bfr Surface Modification

2.2.1. Transglutaminase (TGase)-Catalyzed Reactions

The N-terminal modification (S-peptide) allows microbial TGase (MTGase) to enzymatically modify the surface of the protein with primary amines.[128] Previous experiments on His-tagged bacterioferritin demonstrated that even though there are glutamines (Q) and lysines (K) on the surface of the protein, the glutamines are not accessible to MTGase enzyme hence there is a need for a peptide to be present that would provide the needed glutamine residue in the proper flanking amino acid sequences for MTGase catalyzed surface modifications.
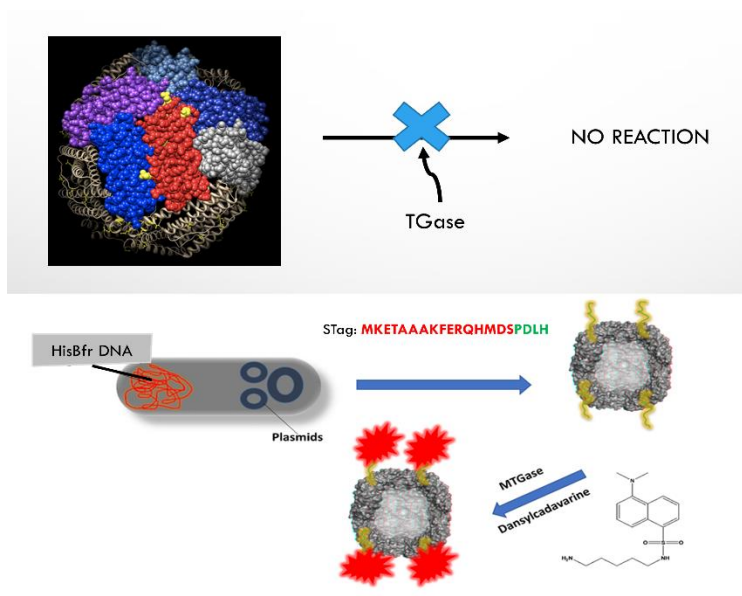
Figure 2.3. Schematic representation of Bfr (1Bfr)-MTGase Reaction. The top panel shows that surface Gln on HisBfr are not modified by TGase. The Bfr is recombinantly overproduced with surface tags (bottom panel) and becomes an MTGase substrate which is covalently modified with dansylcadaverine. The dansylcadaverine are represented as red shapes attached to the yellow Stag shapes.

A reaction mixture containing MTGase (1 equiv. 2.66×10^{-11} moles), QTag-Bfr (226 equiv., 6×10^{-9} moles), dansylcadaverine (DC) (451 equiv., 1.2×10^{-8} moles), and NaCl (38 equiv., 1×10^{-9} moles) in 100 mM Tris-HCl-pH 8, was incubated at different temperatures (from 4 to 50 ˚C) for different times (ranging from 1 h to 24 h). The reaction mixture was then loaded onto a desalting column Sephadex^{TM} G10 (GE Healthcare) to exchange the buffer and also remove the excess enzyme and crosslinked products. The optimum reaction condition was chosen based on the fluorescence intensity of the protein band on an unstained SDS-PAGE gel. The gel was further stained with Coomassie brilliant blue dye to visualize for the molecular weight of covalently modified subunit and cross-linked products.

The extent of crosslinking and band intensities were used to set optimum reaction conditions for the enzyme to be 10 ˚C for at least 2 h in a water bath. +ESI-MS was utilized to ascertain the overall mass of the DC-Qtag-Bfr protein adduct. The samples for mass spectrometry were run through an IMAC column to separate the modified His-tagged proteins from the excess enzyme. The fractions in the imidazole gradient were pooled, and buffer exchanged into DDH$_2$O and then injected into the MS instrument (by adding 10 uL of protein to 90 uL of 1:1 MeOH: H$_2$O with 0.1% formic acid mixture). Alternatively, reaction products were dried using a speed-vac or lyophilized and resuspended in DDH$_2$O for MS analysis.

2.2.2. Cyanogen Bromide Treatment

The dansylated proteins were further treated with cyanogen bromide (CNBr) to assess which glutamine was labeled with DC. A protein treated with CNBr will be cleaved into fragments at the C-terminal side of methionine residues.[143] The choice of this reagent was based on an initial prediction from The ExPASy peptide cutter tool,[129] which predicted 9 possible fragments from Qtag-Bfr. The CNBr was suspended in a 70 % formic acid (FA) solution to a final concentration of 50 mg/mL FA and was used to unfold and cleave the protein. The reaction mixture was incubated at room temperature overnight, and the excess reagents were evaporated in an SC10 speed-vac. The resulting protein powder was resuspended in DDH$_2$O and buffer-exchanged employing a Sephadex G10^{TM} column (equilibrated with filtered DDH$_2$O). The fractions containing protein were pooled and analyzed using (+ESI-MS) mass spectrometry.

2.2.3. QTag-Bfr RNase Reconstitution

The abilities of Qtag 1 and Qtag 2 to reconstitute RNase activities were explored in collaboration with Taylor Urquhart (Honek laboratory). First, the production the S-protein by subtilisin digest was undertaken, followed by chromatographic purification as reported by Genz *et al.* with slight modification.[106] illustrated in Figure 2.4 below.

Cleave buffer: 25 mM Tris–HCl pH 8.0, 50 mM KCl
Binding buffer: 20 mM Na-phosphate pH 5.0
Storage buffer : 20 mM Bis-Tris–HCl pH 6.5
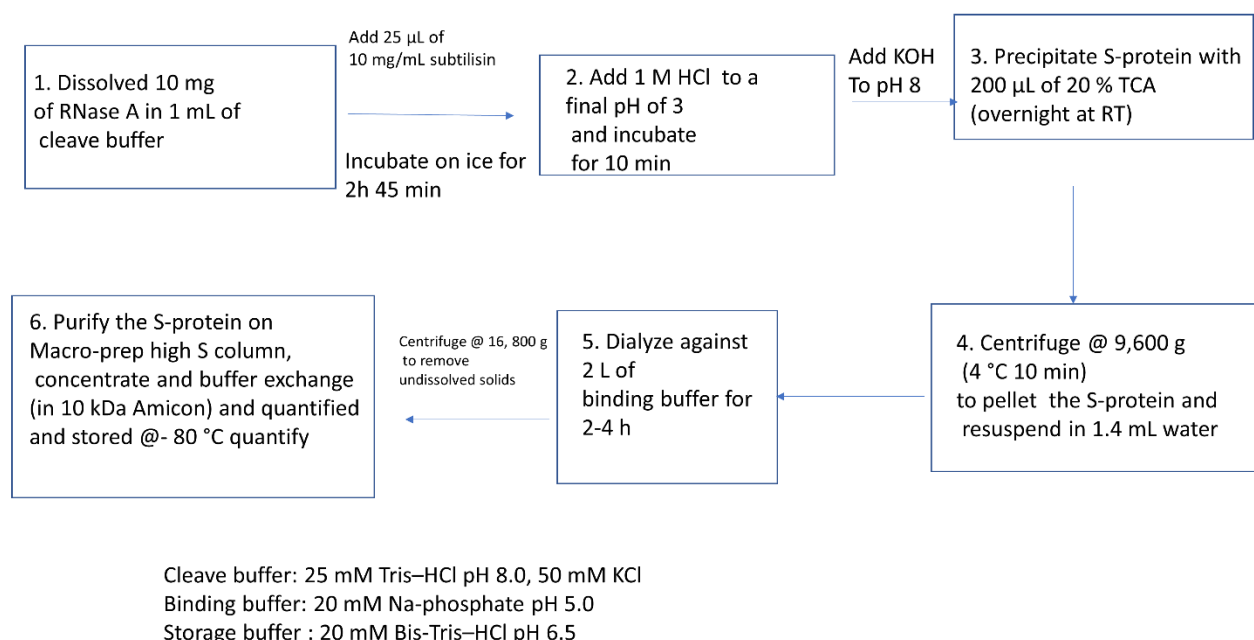
Figure 2.4. Production of RNase S-protein as per the method reported by Genz *et al.*[106]

2.2.3.1. RNase Reconstitution

RNase S protein, and therefore RNase activity, was fully reconstituted by incubating commercial S-peptide with the S-protein, prepared in the above protocol. Firstly, a commercial RNase S-tag peptide (KETAAAKFERQHMDS, from APExBIO, Houston, TX, USA) was incubated with the purified S-protein to test for reconstitution of RNase S activity on a commercial substrate. The reconstitution mixture contained 6 µM each of the S-protein and peptide in an RNase-free vial, from the kit, and the mixture was incubated on ice and then diluted with RNase-free water (kit) to a final concentration of 60 nM. The RNaseAlert™ substrate (Thermo Fisher) is a propriety RNA substrate which is tagged with a fluorescent reporter

55

molecule on one end of the oligonucleotide and a quencher on the other end. The proximity of the quencher dampens fluorescence from the fluorophore until cleavage by an RNase. The RNA substrate is then cleaved, and the fluorophore and quencher are spatially separated in solution, which results in bright green fluorescence emission of the freed fluorophore (Figure 2.4.1). The RNaseAlert® substrate was prepared by resuspending it in 5 µL of reaction buffer (from the kit), and then 45 µL of the 60 nM reconstituted RNase S was added and quickly transferred into a black, opaque, half-area 96-well plate and placed in the plate reader. Top read fluorescence was recorded by setting the excitation and emission wavelengths at 490 nm and 520 nm respectively, with 515 nm emission cut-off. Also, the PMT was set to 6, and all measurements were recorded at 37 °C. These conditions were used to prepare and record RNase samples from the Qtag-Bfr constructs (Qtag 1 and 2), with the Qtag-Bfr acting as the S-tag peptide.

Secondly, the rate of hydrolysis was estimated by fitting the data to a modified first order reaction:

$$F = Fmax(e^{-k(t+T)})$$

where F is the maximum fluorescence (RFU), proportional to the hydrolyzed substrate, Fmax is the total amount of substrate, k is the rate constant, t is the elapsed time and T is the elapsed time between the first reading and the time it was displayed the screen.
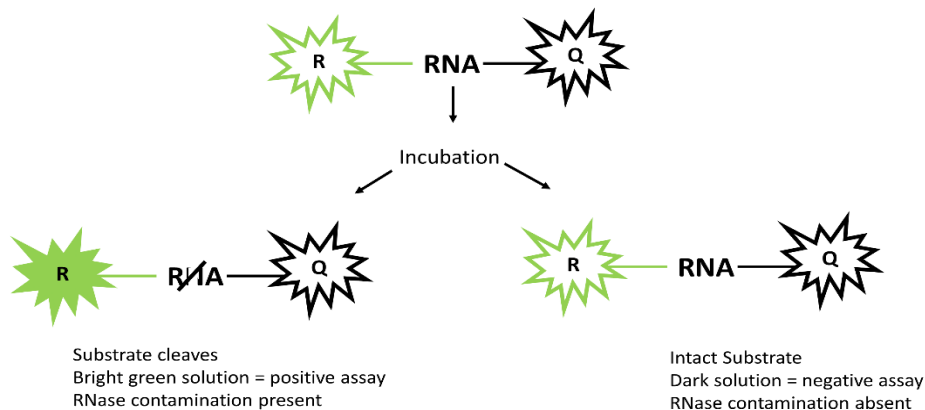
Figure 2.4.1. Schematic representation of the RNase assay. The commercial RNase substrate with a fluorescent probe and a quencher. Subsequent hydrolysis of the substrate increases the distance between the two moieties and prevents quenching. Image was generated in PowerPoint.

2.3. Heme Reconstitution

Heme was reconstituted into bacterioferritin following the procedure reported by Wong *et al.*[101] A 0.01 M solution of hemin in a 0.1 M NaOH aqueous solution was diluted with 0.2 M MES buffer-pH 6.5 to a final concentration of 1.5 mM of hemin. The resulting mixture was then centrifuged at 13250 *g* and the supernatant collected. The protein to be reconstituted was buffer-exchanged into a high salt buffer (1 M NaCl, 0.2 M MES-pH 6.5), mixed with 2 molar equivalent of hemin solution and heated at 80 ˚C for 15 min and then allowed to cool back to room temperature. The sample was then centrifuged at 13250 *g* to remove excess hemin, and further buffer-exchanged employing a 10 kDa spin column to remove excess and nonspecifically bound hemin. The nonspecifically bound hemin could also be removed via dialysis or Sephadex G25[TM] column chromatography (Figure 2.5). The product was analyzed employing UV-visible spectroscopy for the emergence of a Soret band at approximately 418 nm, which is an indication of oligomerization. Size exclusion was used to separate the 24-mers from dimers and monomers. The constructs that did not form 24-mers were mixed (50:50 WT: His) to decrease the steric interactions between C-terminal His-tags. The QTag2-Bfr variants (Qtag2E86A, Qtag2K22A, Qtag2K22A-E86A) were used to study (by Daniel Feng (Chem494), under my supervision) the structural effects of the Bfr-N-terminal sequence. These variants were purified on IMAC columns and analyzed on SEC for their size distributions and further reconstituted with heme to observe the formation of a possible 24-mer.
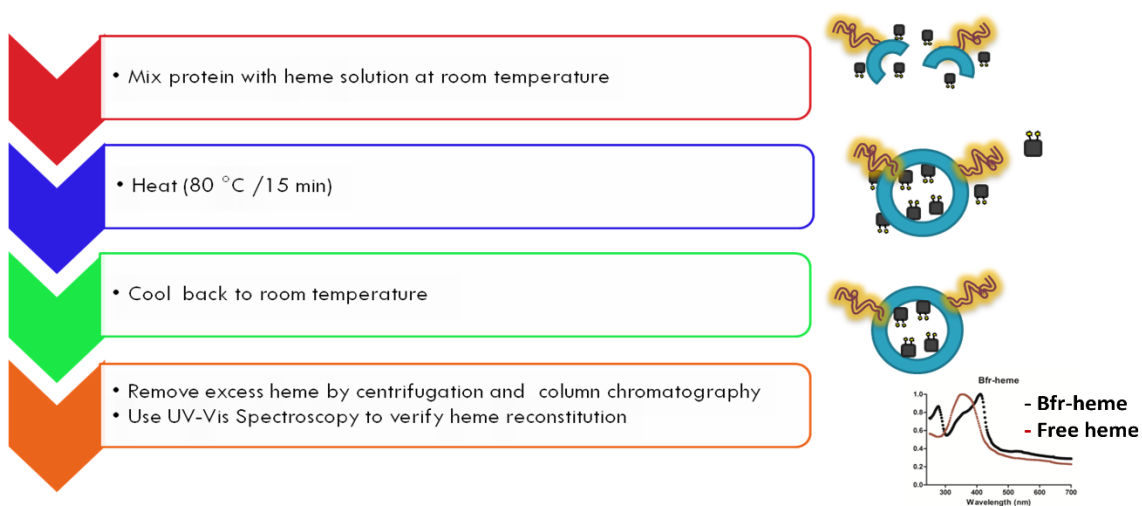
Figure 2.5. Schematic representation of heme reconstitution.

2.3.1. Heme Analog Synthesis

Bis-N-hydroxysuccinimide-heme (Bis-NHS-heme) was synthesized (by Anton van der Ven; M.Sc. thesis) by the addition of hemin (200 mg, 0.307 mmol) and PS-carbodiimide resin (954 mg, 1.3 mmol/g, 4.0 equiv.) to anhydrous dimethylformamide (3.2 ml) with stirring at 23 °C for 10 minutes (Step1). N-Hydroxysuccinimide (NHS) (88 mg, 0.76 mmol, 2.5 equiv.) was added to this solution and stirred at 100 °C for an hour in a Biotage Initiator 8 microwave synthesizer (Biotage Inc., NC, USA) to produce mono- and bis-NHS-modified hemes. The dried NHS-Heme was further reacted with bis(carboxymethyl)-L-lysine (Lys-NTA) (10.3 mg; 0.127 mmol NHS-Heme in 343 uL DMF, 10 mg; 0.381 mmol Lys-NTA in 1 mL of DMF) in a microwave reactor for 1 h at 70 ˚C to produce bis- and mono-NTA-hemes (step 2). The product was analyzed by TLC and mass spectrometry. The product was dried in vacuo to remove DMF and resuspended in 0.1 M NaOH supplemented with 50 % methanol and loaded into a C-18 samplet (3 grams of resin) and left to dry overnight. The sample holder was washed with 50-100 mL of 10 mM phosphate- pH 8 and eluted with a 0-100 % methanol gradient which separated mono- from bis-products. The bis-Lys-NTA-heme product was collected, dried and further resuspended in distilled water for a final cleanup using a P2 size exclusion column using distilled water as eluent. The purified bisNTA-heme was aliquoted and dried in a speed vac and stored at -20 ˚C for later use.
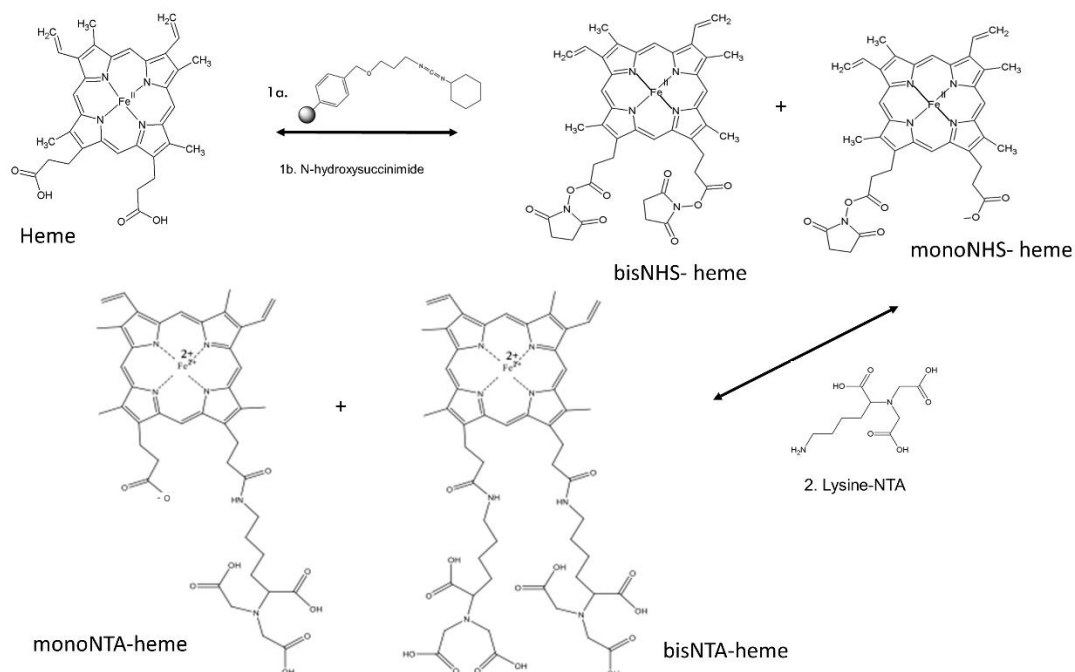
Figure 2.5.1 Synthesis of Heme-NTA. Step1 activation into the NHS ester (bis and mono NHS-heme) followed by labeling with the Lys-NTA into bis and mono-NTA heme.

2.3.2. His-tagged GFP Encapsulation

The HisGFP BL21 stock was gifted to us by Dr. Jeanne Hardy (University of Massachusetts, Amherst) for encapsulation studies. The encapsulation of N-terminally His-tagged GFP into WTBr via a reverse polarity strategy was mediated by the interaction between the bis-NTA-heme cofactor of WTBfr and the polyhistidine tag of the His-tagged GFP guest. The overproduction of His-tagged GFP was similar to other His-tagged proteins with little modification; the plasmid was expressed in an *E. coli* BL21 cell line and induced with 1 mM IPTG at 37 °C. The induced media was shaken at 25 °C and harvested after 3 h to limit inclusion bodies. The purification procedure was identical to the His-tagged protein purification outlined above. Firstly, harvested cells were lysed at 17000 psi in Anvestin emulsiflex C25 and centrifuged at 20000 *g* to remove insoluble DNA. The clarified lysate was loaded onto a pre-equilibrated 1 mL HiTrap™ IMAC column and washed with equilibration buffer (150 mM NaCl, 20 mM imidazole,50 mM Tris-HCl, pH 7.5) until the OD_{490} was zero. The bound proteins were

then eluted with 100 % of elution buffer (150 mM NaCl, 300 mM imidazole, 50 mM Tris-HCl, pH 7.5). Finally, the pulled fractions were buffer exchanged into storage buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5). A few microliters (100 µL) of the protein was washed (in a spin column) with DDH$_2$O for +ESI-MS to verify the mass of the pure protein. The verified protein was then aliquoted into 500 µL portions and stored at -80 °C for future experiments.

Bis- NTA-heme was incorporated into wild type bacterioferritin by dissolving it in 0.2 M MES buffer and mixing with WTBfr in a ratio of 1:2 Bfr: Bis- NTA-heme (in 0.2 M MES-pH 6.5, 1 M NaCl) ratio and heated at 80 °C for 10 min. The mixture was allowed to cool to room temperature and centrifuged at 13250 g and dialyzed (0.2 M MES-pH 6.5, 1 M NaCl; up to 24 hours) to remove excess and non-specifically bound heme. The WTBfr-heme complex was then mixed with His-tagged GFP and Ni^{2+} in a 1:10:100 molar ratio respectively and incubated at 60 °C for 90 min followed by a 30 min cooling period at room temperature and then a size exclusion chromatographic step (Superose® 6 10/300 HR; buffer 25 mM HEPES-pH 7.4, 150 mM , NaCl, 300 mM KCl) was employed. The absorbance of a protein fraction from this column having a molecular weight corresponding to an intact 24-mer bacterioferritin with 495 nm absorption was an indication of GFP encapsulation. The fractions were analyzed on a native-PAGE gel, and mass spectrometry and TEM experiments were conducted on the purified nanomaterial.

2.3.3. Ellman's Reagent (DTNB) Modification of HisGFP

Guest protection was studied by comparing the accessibility of Cys residues in GFP to a bulky reagent such as DTNB. Free HisGFP was incubated with up to 250-fold molar excess DTNB for up to 3 h at room temperature with stirring. The reaction was centrifuged at 13250 g to remove precipitates and washed (with 150 mM NaCl,200 mM MES-pH 6.5) using a 100 kDa Amicon device to remove unreacted reagents. The HisGFP-TNB complex was analyzed using + ESI mass spectrometry and buffer-exchanged into 25 mM HEPES, 20 mM HEPES- pH 7.5 for UV absorbance measurement at 430 nm ($\varepsilon_{430\ nm}$ =12205 mM^{-1}cm^{-1}).[130] Similarly, encapsulated HisGFP was incubated with DTNB and the results were compared. The results illustrate how the Bfr shell protects the SH groups in GFP from reacting with DTNB.

2.4. Gold (Au) Nanoparticle Encapsulation

The capabilities of the cage proteins to encapsulate larger guest molecules was studied with a commercially available 5 nm gold nanoparticle (AuNP) (Ni-NTA-Nanogold[TM], Nanoprobes, Yaphank, NY, USA). This guest molecule has an NTA functionality attached to the surface through a proprietary polymeric linker (Figure 2.6). The nanoparticle solution was initially characterized by UV-visible spectroscopy and was observed to absorb at 518 nm. Also, size distribution analysis employing DLS indicated that the Au particles had an 8 nm diameter on average, which could be due to the presence of the polymeric linker. A 24-molar equivalence of the guest proteins (1 cage) was mixed in a ratio of 60:40 His: WT protein mixture and buffer-exchanged into the declustering buffer (8 M GdnHCl, 25 mM HEPES-, pH 7.6, 100 mM NaCl, for Bfr and 10 mM NaCl, 25 mM HEPES, pH 7.6 for AfFtn). The protein was incubated with 100 molar equivalence of Ni^{2+} ions and 2 molar equivalence of heme (only for the Bfr ferritin host) for approximately 30 min, followed by the addition of 1 molar equivalence of the AuNP in solution. The mixture was allowed to incubate at 23 °C for at least 1 h before the removal of the declustering agent (by increasing NaCl concentration to 150 mM for AfFtn and dialyzing against 25 mM HEPES, 150 mM NaCl, pH 7.6 for Bfr). The encapsulated proteins were applied to a Sephacryl® S-300 10/300 HR column, pre-equilibrated with 25 mM HEPES, 150 mM NaCl, pH 7.6 and the fractions were monitored at four wavelengths: 280 nm, 260 nm, 418 nm (heme in Bfr), and 518 nm. The elution times were compared to prior runs of free AuNP and free protein. Fractions containing encapsulated AuNP (Au coeluted (518 nm) with 24-mer) were collected and concentrated for further analysis.
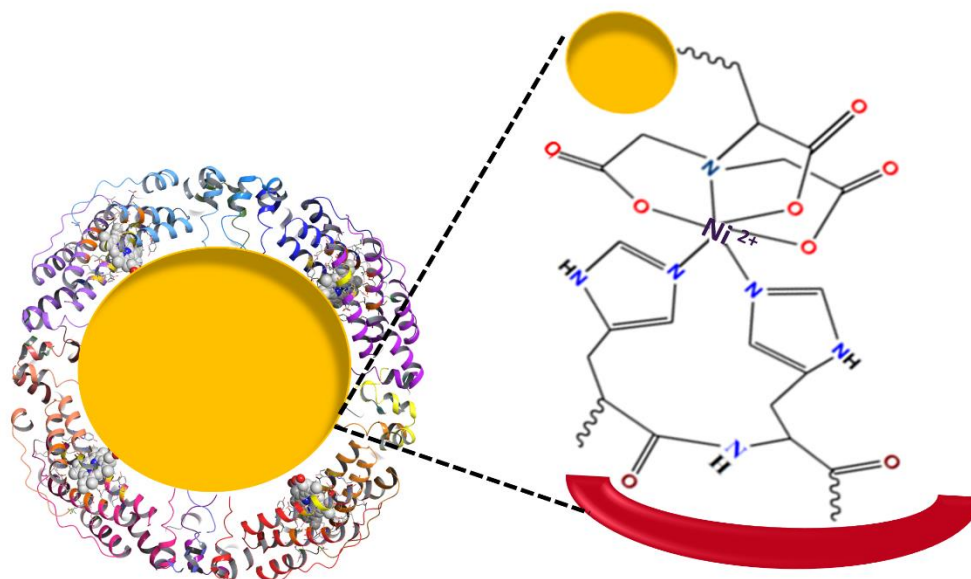
Figure 2.6. The AuNP- Ni^{2+}-HisAfFtn interaction. The AuNP, yellow sphere on the left, is stabilized by the interaction between the NTA functionalities on AuNP and two histidines from the polyhistidine tags of the subunits that coordinates Ni^{2+} (both HisBfr and HisAfFtn). The ferritin subunits are represented in red shape on the right. The Bfr cluster was generate with Maestro (Schrödinger LLC)

2.4.1. Bfr-Targeted Guest Delivery

The host-guest AuNP (5 nm)-protein complex (Qtag-Bfr-Au) was dansylated (DC-Qtag-Bfr-Au) to illustrate that Bfr can be repurposed on both surfaces while a guest is within the protein shell. TGase-catalyzed reactions (Section 2.2.1) were used to modify the Qtags with DC using the procedure above (Section 2.5). The declustering buffer was then removed by dialyzing the mixture overnight against a reclustering buffer (150 mM NaCl and 25 mM MES-PH 7.4). Dansylation was carried out similarly to that described in Section 2.2.1. Moreover, the mixture was run on a Sephacryl® S-300 (10/300) size exclusion column using the reclustering buffer to separate the DC-Qtag2-Bfr-Au from the other oligomers in solution. The desired fractions were analyzed on a native-PAGE gel to observe a dansylated 24-mer protein band, while TEM was used to visualize the encapsulated AuNP.

2.5. Lipid-NTA Encapsulation

2.5.1. Synthesis of C19-NTA

To investigate the hydrophobic engineering of the internal cavity of bacterioferritin, a strategy to load the interior cavity of Bfr with lipids was undertaken. The lipids were designed to have NTA functionalities attached to them so that any potential micelles or hydrophobic aggregates formed from them might likely interact with the hexahistidine tag of His-tagged Bfr, resulting in encapsulation of a large hydrophobic molecular ensemble. Nonadecanoyl-NTA was synthesized by Omaima Ben Krayem (M.Sc. University of Waterloo, 2017) using the procedure outlined by Mènard and coworkers which begins with carboxylic acid activation of the fatty acid, followed by modification with bis(carboxymethyl)lysine (Lys-NTA).[118] The carboxy group of nanodecanoic acid was activated with 1-ethyl-3-(3-dimethyl aminopropyl) carbodiimide (EDC) to enable N-hydroxysuccinimide (NHS) ester formation. Nanodecanoic acid (1 equivalent, 29.3 g, 98.2 mmol) and 1.5 equivalents of N-hydroxysuccinimide (16.9 g, 147.3 mmol) were dissolved in 700 mL methylene chloride (CH_2Cl_2), followed by stirring at room temperature for 24 h. The resulting solution (containing Nonadecanoyl-NHS) was diluted with 700 mL of water. The organic phase was separated by washing the mixture with 3 x 250 mL of CH_2Cl_2, 250 mL of brine and dried over anhydrous Na_2SO_4. The methylene chloride was evaporated and the resulting solid was recrystallized with 700 mL EtOAc to obtain powdered nonadecanoyl-NHS (white powder). The powdered product (3.5 g, 9.1 mmol, one equivalent) was combined with one equivalent (2.4 g, 9.17 mmol) of bis(carboxymethyl)lysine (Lys-NTA) and triethylamine (7 equivalents; 9 mL, 64.2 mmol) and the mixture was dissolved in DMF (50 mL) at room temperature for 48 h with stirring. The mixture was evaporated under reduced pressure, and the residue was resuspended in 150 mL H_2O, which was then acidified with the addition of 1N HCl to precipitate the product. The product was filtered and dried under vacuum to obtain white powdered $C_{19}H_{54}N_2O_7$. A few mg of the final product was dissolved in DMSO and analyzed on a Bruker Avance 500 NMR spectrometer. Mass spectrometry analysis was performed using a linear ion trap quadrupole (LTQ) to identify the molecular mass of the product as well as the impurities that could be present. The MS sample was prepared by making a 1 mg/mL solution in water. The lipid was solubilized by adjusting its pH to 12 with 1 N NaOH and then adding a few drops of 6 M HCl to a

pH of 8. The dissolved lipid (10 µL) was added to 90 µL of MS solvent (1:1:0.1, H_2O: MeOH: FA) and injected into the ion source at a flow rate of 5 µL/min.

2.5.2. Synthesis of C14-NTA and C16-NTA

A simplified version of the NTA functionalization strategy, similar to the heme-NTA synthesis, was employed by April Marple (Chem494 undergraduate student in the Honek laboratory). Commercially available NHS-ester activated fatty acids (myristic (C14) and palmitic (C16)) were directly reacted with bis(carboxymethyl)lysine (Lys-NTA). One equivalent (50 mg each, 0.141 mmole C14-NHS, 0.154 mmole C16-NHS) of the fatty acid-NHS ester was dissolved in 2 mL of anhydrous DMF and incubated with 1 equivalent of bis(carboxymethyl)lysine (Lys-NTA) (39.6 mg) and 7 equivalents of Et_3N (0.138 mL, 0.99 mmoles and 0.15 mL, 1.08 mmoles for C-14NHS and C16 NHS, respectively). The resulting mixture was then let to stir for 48 h at 23 $^{\circ}$C. Then the solution was evaporated in vacuo and the solid was redissolved in 4 mL of doubly distilled water and acidified with 1 N HCl until a white solid appeared. The solid was filtered under suction and further dried in a desiccator. The modified lipids/fatty acids (74 % and 79 % yield for C14-NTA and C16-NTA, respectively) were further analyzed employing thin layer chromatography (TLC) (stationary phase: silica, mobile phase: 100 % dichloromethane), mass spectrometry (Appendix 6), and NMR (^{1}H-NMR and ^{13}C-NMR) (Appendix 7).
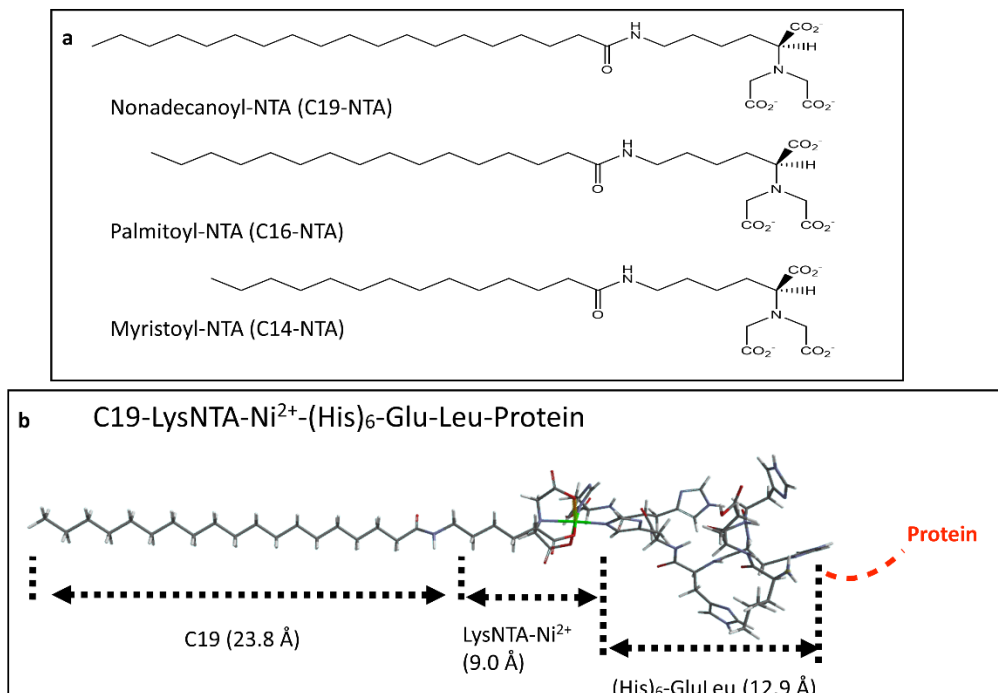
Figure 2.7. Functionalized lipids for encapsulation. (a) Chemical structures of the three lipid-NTA that were synthesized and (b) models of the extended length of the C19-NTA lipid when it interacts with the C-terminal H_6-tag from the protein. Distance measurements were made using structures drawn in Spartan (Wavefunction, Inc).

2.5.3. Lipid-NTA Encapsulation

The lipid-NTA was encapsulated utilizing two different approaches: Bfr-mediated encapsulation and lipid-mediated encapsulation. The Bfr-mediated encapsulation uses lipids at surfactant concentrations below the critical micelle concentration (CMC), whereas the fatty acid-mediated encapsulation encapsulates a likely preformed micelle into the Bfr cage. DLS analysis was used to determine the aggregation concentration (2 mg/mL) of the synthesized surfactant before the encapsulation studies. Additional encapsulation studies with mixtures of 50:50 HisBfr: WTBfr were employed to study the effects of the multiple C-terminal tags on the Bfr-lipid complex and the resulting dimensions of the reclustered capsule protein. The Bfr-mediated encapsulation reaction contained, 1 equivalent His-tagged Bfr (2.04×10^{-8} moles, in declustering buffer), 45 equivalents of C19-NTA (9.25×10^{-7} moles, ½ CMC), 2 equivalents of heme ($4.08 \times$

65

10^{-8} moles) and 2 equivalents of Ni^{2+} (4.08 x 10^{-8} moles). A ¼ CMC reaction was also set up to investigate lipid effects on Bfr cage dimensions. The lipid-mediated approach used a similar reaction setup with ND-NTA lipid at micelle concentration (1.85 x 10^{-6} moles). The mixtures were incubated at 5 °C for 90 min followed by overnight dialysis in a 12.4 kDa cutoff tubing (buffer: 50 mM HEPES-pH 7.4, 150 mM NaCl) to remove the excess GdnHCl, centrifuged to remove precipitate and loaded onto a Sephacryl® S-300 10/ 300 -HR column. The protein fractions eluting at a time corresponding to an intact capsule protein (24 subunits) were collected and analyzed using mass spectrometry, dynamic light scattering, and transmission electron microscopy.

Alternatively, a mixture of 50 %: 50 % / HisBfr: WTBfr, or using shorter chain lipids can be used to overcome the internal space limitation (see Results and Discussion later in this thesis) since the lipids may not require all 24-mer H6-tags for encapsulation. Hence a reaction mixture containing equal parts of WTBfr and HisBfr was used as host while keeping all other reagents the same as the encapsulation at the 2 mg/mL lipid above for C19-NTA. In the case of C14 and C16, the amount of lipid that formed micelles within Bfr interior dimensions were encapsulated into 100 % HisBfr cage due to their decreased chain length.

2.6. Protein Encapsulation into AfFtn
2.6.1. Streptavidin-FITC (SF) Encapsulation into AfFtn

The ability of AfFtn to encapsulate intact proteins was demonstrated by attempts to encapsulate fluorescein-labeled streptavidin (SF). The streptavidin molecule has an isothiocyanate-functionalized fluorescein attached to ε and N-terminal amines in the streptavidin molecule as seen in Figure 2.8 below. The derivatization results in the formation of a thiourea bond between streptavidin and the FITC labeling reagent. The conjugation results in bright orange color with excitation and emission bands at 495 nm and 520 nm respectively. The encapsulation procedure begins with incubating the SF protein (1.85 x 10^{-9} moles) with Biotin-X nitrilotriacetic acid, tripotassium salt (Biotin-X NTA) (3.33 x 10^{-8} moles) for 30 min at 23 °C with 2.04 x 10^{-8} moles of a 50:50 WTAfFtn: HisAfFtn mixture and a 4.0 x 10^{-8} moles Ni^{2+} were added followed by incubation at 50 °C for 90 min. The resultant mixture was allowed to cool to room temperature and centrifuged at 13250 g to remove precipitate before injecting onto a Sephacryl® S-300

10/300HR column. The chromatographic peak corresponding to a molecular weight of the 24-mer structure with 495 nm absorbance (for SF) was collected and analyzed on a Native-PAGE gel, and UV-visible spectroscopy and TEM were employed to further explore the overall structure of the host-guest protein complex.
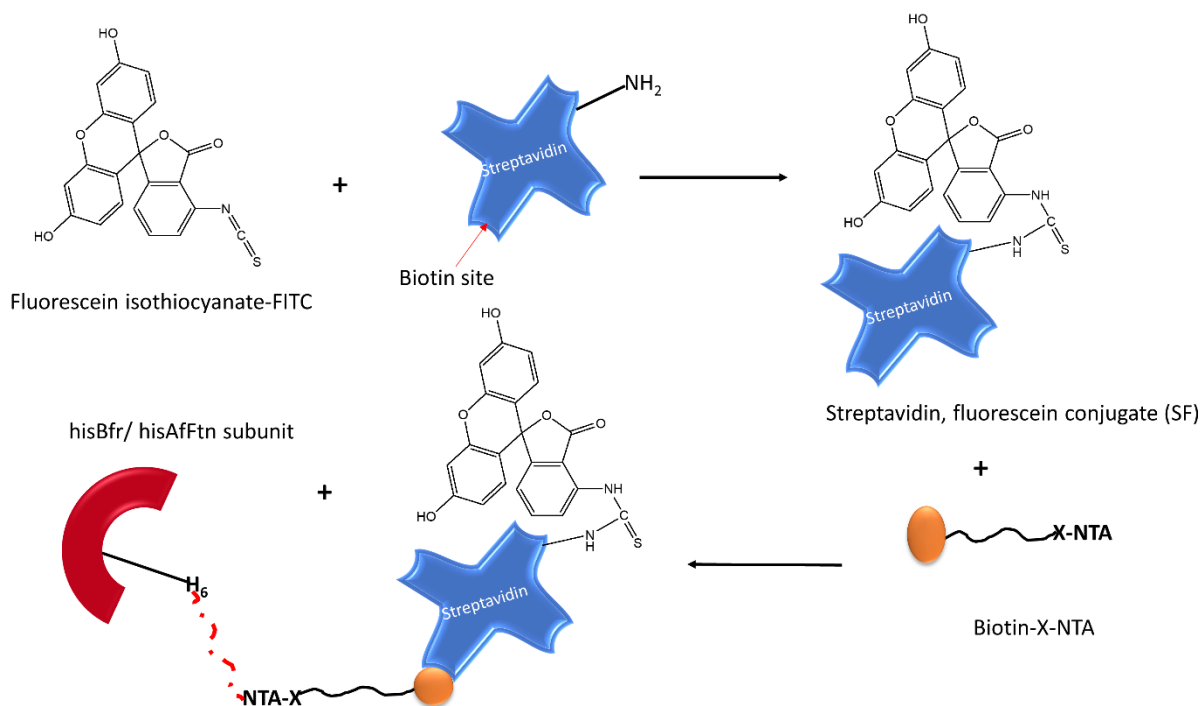
Figure 2.8. Schematic representation of streptavidin derivatization and encapsulation: The streptavidin-fluorescein conjugate (SF) was obtained from Sigma Aldrich and modified with biotin-X-NTA, which interacted with the streptavidin and also the hexahistidine from the capsule protein subunits (red dots).

2.6.2. Glucohydrolase Encapsulation into AfFtn

A catalytically functional nanoreactor was engineered from an AfFtnC54 variant through affinity mediated encapsulation (Figure 2.1). The C54 group was modified with Maleimido-C3-NTA, which formed a protein adduct with an NTA affinity tag attached to it. The NTA would

then be expected to interact with a His-tagged guest enzyme (in the presence of Ni^{2+}). A solution containing AfFtnC54 (1.2×10^{-8} moles in 20 mM NaCl, 25 mM HEPES-pH 7.4) was sparged with Argon gas and reduced with 50-100 molar excess TCEP ($0.6-1.2 \times 10^{-6}$ moles) at room temperature for 30 min in order to reduce any inter-subunit disulfide bonds forming in solution. Maleimide-X-NTA (20x molar excess, 2.4×10^{-7} moles) was dissolved in the protein buffer and flushed with Argon gas and was added to the protein solution. The solution was then capped and allowed to react with the Cys groups of AfFtnC54 for about 60 min. The AfFtnC54-Maleimido-X-NTA complex was analyzed utilizing mass spectrometry. The prepared protein was then investigated with (i) His-tagged GFP as a model protein of a His-tagged enzyme and (ii) a commercial recombinant alpha-glucosidase from *Bacillus stearothermophilus* (Megazyme, Inc.; Cedarlane, Inc as a local supplier). The above protocol for maleimide labeling was further optimized by removing the TCEP reduction step (detailed explanation in Chapter 3).

2.7. Instrumentation

2.7.1. Mass Spectrometry Analysis

Accurate masses of purified proteins were verified employing a Thermo Scientific Q Exactive[TM] Orbitrap instrument with an Electrospray Ionization (ESI) source. About 100 µL of the pooled fractions, identified by SDS-PAGE electrophoresis, was washed with DDH_2O in a Nanosep® microcentrifuge spin column until all salts and buffers were removed and then suspended in DDH_2O. The mass was acquired in positive ion mode by using a 1:1 MeOH: H_2O with 0.1% formic acid mixture as a mobile solvent. The protein to be analyzed was then diluted with this solvent to a final concentration of about 1-10 pmol/uL and injected at a flow rate of 5 uL/ min into the ESI ion source. The injected sample was heated and bombarded with ions of collisional energies between 40-60 eV, causing them to fragment into various daughter ions. The daughter ions (in mass/charge ratios, isotopically unresolved) were deconvoluted into their corresponding intact protein masses using either the Thermo Protein Deconvolution software: https://assets.thermofisher.com/TFS-Assets/CMD/manuals/Man-XCALI-97576-Protein-Deconvolution-30-User-ManXCALI97576-A-EN.pdf or the Thermo BioPharma software: http://tools.thermofisher.com/content/sfs/manuals/Man-XCALI-97813-BioPharma-Finder-User-ManXCALI97813-EN.pdf.

2.7.2. Size Exclusion Chromatography (SEC)

The Size Exclusion Chromatographic technique was used to determine and separate tertiary and quaternary structures in solutions of purified, heme reconstituted, and encapsulated proteins. Aliquots (0.25-1 mL; filtered and centrifuged) were loaded on to three different kinds of columns, Sephacryl™ S-300 10/300 HR, Sephacryl™ S-400 26/60 HR, Superose® 6 10/300 GL, depending on the time and the amount of sample available. The flow rates used were between 0.5-1.3 mL, and 1 mL fractions were collected. The BioLogic DuoFlow™ system used was equipped with a QuadTec UV-Vis detector which allowed for the simultaneous visualization of four wavelengths, making it very useful for detection of guest molecules with different absorption wavelengths. Calibration profiles were prepared by running BioRad Gel filtration standards (#1511109 containing: 5 mg thyroglobulin- 670,000 Da, 5 mg g-globulin-158,000, 5 mg ovalbumin -44,000, 2.5 mg myoglobin-17,000 Da, and 0.5 vitamin B12-1,350 Da) (Appendix 2) and used to estimate the elution times/ volumes of the samples injected, which then were used to estimate an approximation of the molecular weights of the various protein complexes.

2.7.3. Dynamic Light Scattering (DLS)

The size distribution of the particles in solution was analyzed by employing a Zetasizer Nano ZS model ZEN3500 (Malvern Instruments Ltd, Worcestershire, England) equipped with a He-Ne laser with 4.0 mW power at a 532 nm wavelength. The size measurement (DLS method) protocol was used. The fluctuations in scattered light at a specific angle were detected with an avalanche photodiode array (APDs) which was visualized in the form of intensity over a period of time. Water was used as dispersant with a refractive index of 1.33, and the temperature was set to 20 °C. All samples were centrifuged and filtered before the measurement and were placed in a 45 uL Hellma cuvette, with 0.3 mm path length and run in triplicates. Sizes were estimated from the Stokes-Einstein relationship ($D = kBT/6\pi\eta RH$), assuming that they all existed as spheres in solution and the dispersant viscosity (η) was set as the sample viscosity.

2.7.4. Transmission Electron Microscopy (TEM)

TEM analysis was performed on a CM10 Philips transmission electron microscope modified with an Advanced Microscopy Techniques image capturing CCD camera, housed in the

Department of Biology. The images were collected in bright field mode with 100 keV accelerating potential. 0.1 mg/mL samples were buffer-exchanged into DDH_2O water to remove excess buffer and salt (when necessary). The samples were applied onto copper grids (Formvar-Carbon 400 mesh; Ted Pella, Redding, CA, USA), placing the grid onto a 20 uL droplet of the sample. The grids were lifted up and dabbed with a piece of filter paper to remove the excess sample. The thin layer of the sample was then washed twice similarly to how the samples were applied and finally placed onto a droplet (20 uL) of 0.5-1 % Molybdic acid solution the excess dye solutions were dabbed with pieces of filter paper and put into Petri dishes and incubated at room temperature overnight before the TEM analysis.

2.7.5. Fluorescence Measurements

Fluorescence measurements were carried out on a Photon Technology International Fluorometer, A1010B Steady-State fluorescence system equipped with an LPS-220B lamp power supply, an SC-500 shutter control, MD-5020 motor driver, and 814 photomultiplier detector. Solutions of 500 µL were placed in a 1 mL micro fluorescence quartz cuvette and inserted into the sample holder and excited with the appropriate excitation wavelengths. The emission scans were collected at +15 nm from the excitation wavelength until it tailed off. A 1 nm slit width was used in most cases except for dilute samples where it was set to 2 nm.

2.7.6. Ultraviolet (UV) Measurements

All UV measurements were carried out on a SpectraMax 5 microplate reader. The sample solution and buffers were filtered with a 4.5 µm syringe filter to removed unwanted particles before placing in a 45 uL Hellma cuvette (0.3 mm pathlength). On the other hand, optical density (OD 600) measurements were performed on the BioRad SmartSpec™ plus using 1.5 mL disposable polymethylmethacrylate cuvettes.

2.7.7. Analytical Ultracentrifugation (AUC)

The AUC samples were prepared in our laboratory, frozen, and transported, on ice, to the Biomolecular Interactions & Conformations Facility of Western University. Sedimentation velocity studies were carried out (by Lee-Ann Briere, facility manager) using a Beckman Optima XL-A Analytical Ultracentrifuge. An An60Ti rotor and double-sector cells with Epon charcoal

centerpieces were used. Centrifugation was carried out at 20 °C, at 20 000 rpm or 30 000 rpm centrigugation speed depending on the sample. Forty-five absorbance measurements at either 280 or 300 nm were collected at 10 minute intervals, in 0.002 cm radial steps. Data were analyzed using the c(s) distribution model in Sedfit and normalized in GUSSI. The partial specific volumes (Vbar) were also calculated from the amino acid compositions, using the program SEDNTERP. SEDNTERP was also used to calculate buffer, densities, and viscosities seen in Appendix 9. These parameters were used to choose the speed for the samples.

CHAPTER 3. RESULTS AND DISCUSSION FOR EXPERIMENTAL SECTION

3. Brief Introduction

This chapter combines both the results and the discussions of the experiments performed in this thesis(results from experiments in chapter 2). The chapter begins with the proteins used, discussing how the constructs were designed and transformed and the protein productions and purifications as well as the characterizations (sections 3.1.-3.1.5.). Subsequent sections (section3.2.-3.9.)focus on the biotemplating strategies and, where specific enzymes and reagents were utilized, will contain a brief description of the enzyme or the conjugation technique.

3.1. Protein Production

The recombinant proteins employed in this work were all designed and overproduced in the Honek laboratory as described in the Methods section (chapter 2). Several of the proteins were designed in our laboratories for particular functions (His-tagged; Cysteine mutation, etc.) and then commercially obtained through custom gene syntheses (Genscript, Inc.). The presence of C-terminal hexahistidine tags was useful for immobilized metal affinity interaction (IMAC) purifications of all the HisBfr, HTEKBfr, HisAfFtn and HisPfFtn constructs. A BioRad DuoFlow instrument with a BioLogic QuadTec™ Detector was used for all chromatographic separations, allowing the detection of up to four wavelengths associated with the protein modification. These wavelengths were 280 nm, 418 nm, 490 nm, and 518 nm, which correspond to the spectroscopic properties of the proteins, heme, GFP, and 5 nm Au nanoparticle, respectively.

3.1.1. Purification of His-tagged Protein

The His-tag on the interior surface of Bfr and AfFtn constructs require buffer conditions that will decluster the subunits and cause the various quaternary structures to exist in equilibrium. The declustering process is necessary for the IMAC purification which is dependent on the exposure of the His-tags. For instance, AfFtn subunits associate into a 24-mer at salt concentrations of 50 mM (see DLS analysis, Figure 3.8.1) and greater so, 20 mM NaCl was used in both binding and elution buffers. Before the IMAC column purification, the lysate is heated to 80 °C or 70 °C for 10 minutes to remove thermally labile host proteins and up to 100 °C for PfFtn, followed by cooling and centrifugation before applying to the nickel His-trap column.

The general elution profile for the isolation of these His-tagged proteins employing an IMAC purification strategy includes a large flowthrough peak and a subsequent washing period, where the monitored UV absorbance is reduced back to the baseline, followed by the elution of the His-tagged protein as a sharp peak when an imidazole gradient is subsequently employed (Figure 3). The presence of a 418 nm peak in the Bfr fractions was indicative of the presence of the heme-group in Bfr (Figure 3a), unlike AfFtn and PfFtn which do not contain the heme cofactor (Figure 3b). A few microliters (20 µL) from several fractions such as the flowthrough, the wash, and the imidazole eluted regions were analyzed utilizing 15 % SDS PAGE for the detection of specific protein bands and the desired proteins were observed to be predominantly in the elution peak (imidazole gradient peak) with sufficient purity for many subsequent studies. The presence of pure proteins in the imidazole gradient indicated that most of the proteins existed as monomers and dimers under the purification conditions, that exposed the C-terminal to Ni^{2+} His trap column. In instances where most or all of the proteins were in the flowthrough, such as the case of PfFtn, the binding and elution buffers were supplemented with 6 M GdnHCl to descluster the subunits. Unlike other proteins which completely unfold in 6 M GdnHCl and require subsequent refolding steps, Bfr dissociates in subunits which reassociate upon removal of the declustering agent. The fractions containing the pure protein of interest were pooled (usually all peak fractions in the imidazole gradient) and buffer exchanged into a storage buffer, and the dilute samples were concentrated, quantified and stored (in small aliquots) for further analysis.

The average yield of all the ferritins was approximately 30 mg/L of cell growth, estimated from the Beer's equation with an extinction coefficient of 21430 $M^{-1}cm^{-1}$ and 33920 $M^{-1}cm^{-1}$ for Bfr and AfFtn, respectively (extinction coefficients obtained from the amino acid sequence using the ProtParam tool; https://web.expasy.org/protparam/). The generalized elution profiles for heme and non-heme constructs are shown in Figure 3a, b, and c respectively. Electrospray ionization mass spectrometry was used to validate the estimated subunit mass obtained from SDS-PAGE electrophoresis; a 100 µL aliquot of each of the pure proteins was exchanged into DDH2O. The washed sample was mixed with $1H_2O$: 1MeOH/0.1 % FA   in a 10 :90 ratio (10 µL protein: 90 µL solvent) solution and ionized into daughter ions (source spectrum; mass to charge ratios, Figure 3.1a) and deconvoluted to obtain the exact masses of the samples (Figure 3.1 b-e). The 1 amu mass deviation (M-1) in all samples could result from poor isotopic resolution of the quadrupole at the higher mass range, unlike HisBfr where a mass deviation of 152.2 amu in HisGFP (Figure

3.1d) depicts the loss of the N-terminal methionine and water from the HisGFP chromophore formation. The mass loss could result from the activity of the bacterial methionyl aminopeptidase (MetAP) in the expression system. The loss of an N-terminal methionine was observed for HisAfFtn as well. (Figure 3.1e)
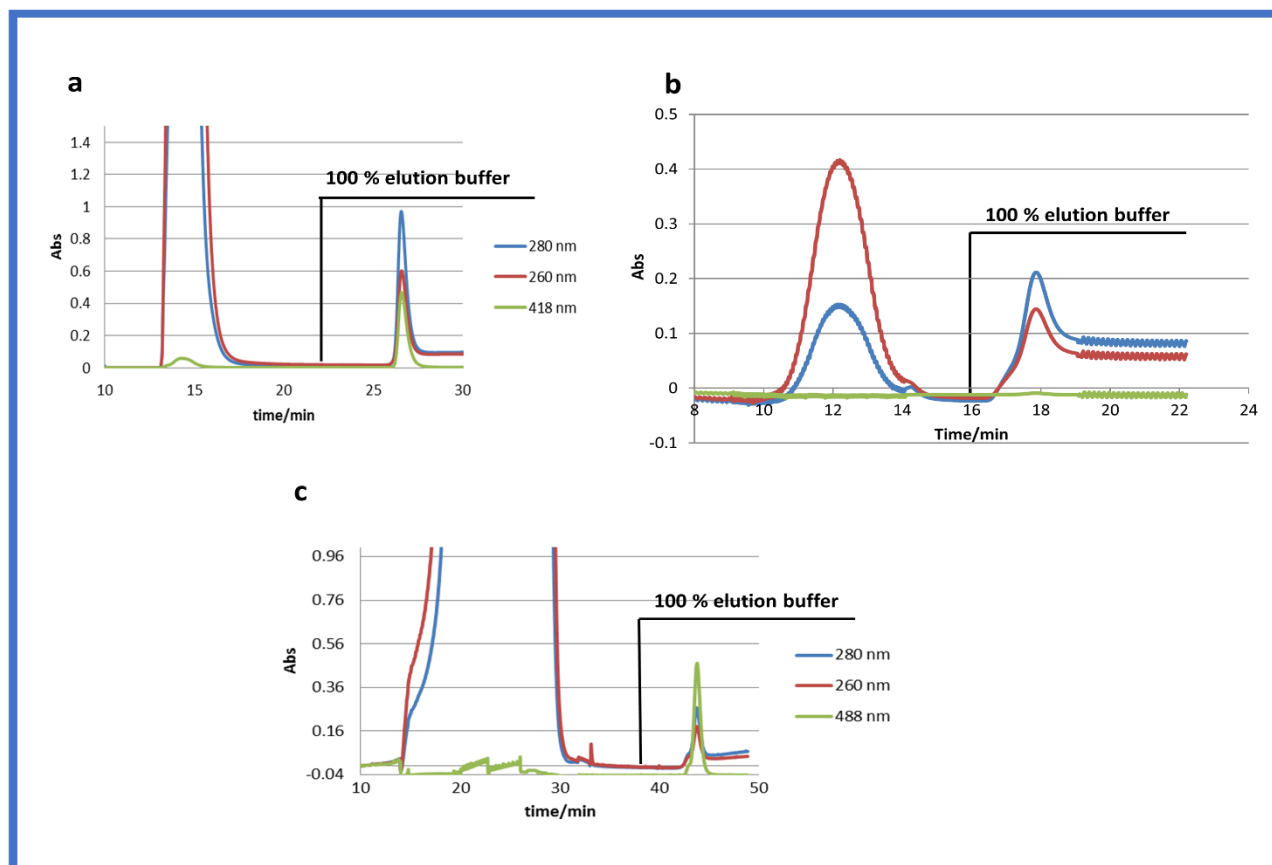
Figure 3. Generalized chromatograms of nickel his-trap affinity purifications. (a) Elution profile for heme-containing His-tagged proteins. (b) Elution profile for non-heme-containing His-tagged proteins. (c) His-tagged GFP elution profile. The blue, red, and green curves represent 280 nm, 260 nm, 418/488 nm absorbances, respectively.
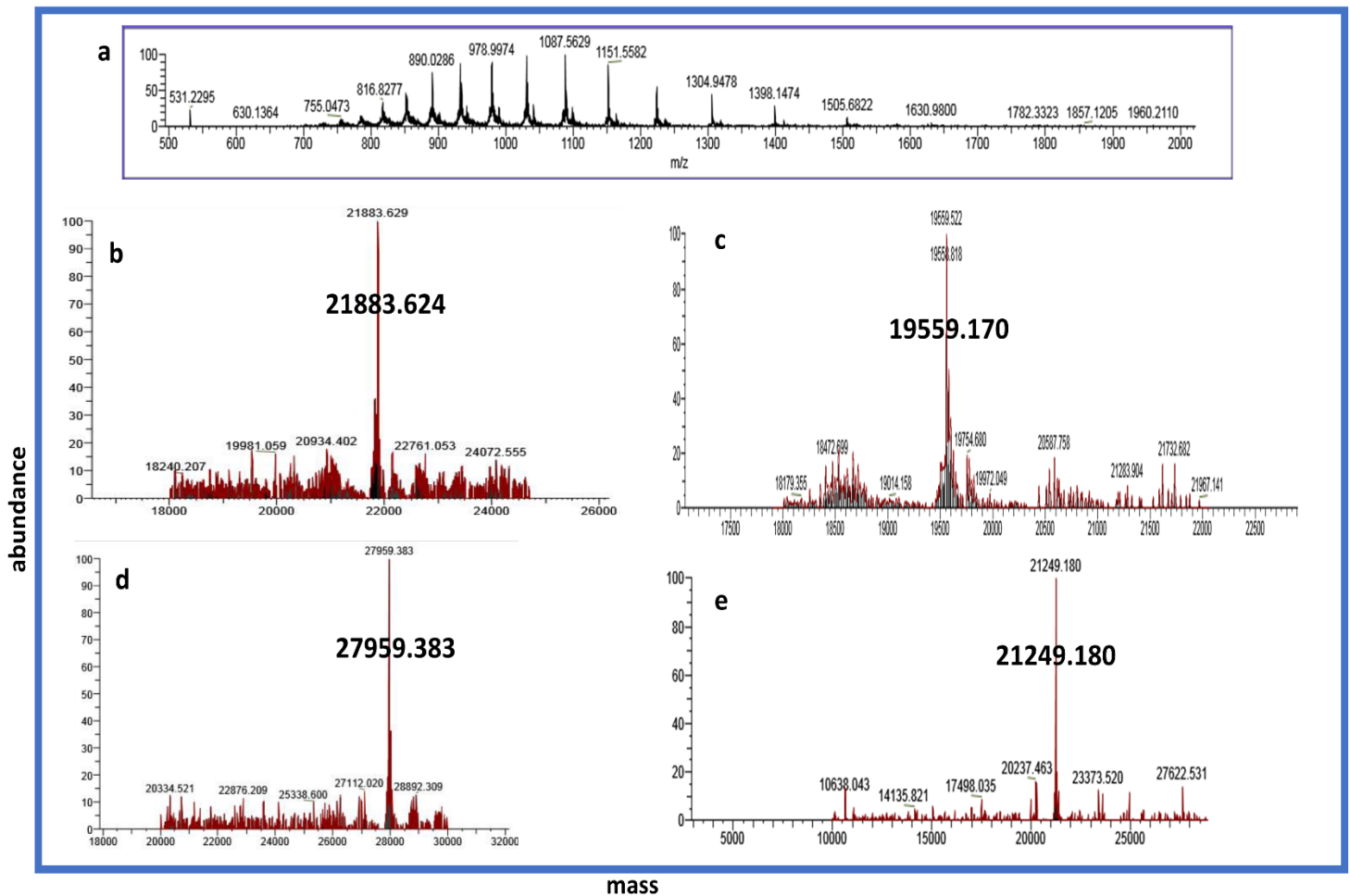
Figure 3.1. Positive ion electrospray ionization mass spectrometry data. (a) the source spectrum of HisBfr; (b) Qtag1-Bfr deconvoluted spectrum (theoretical mass 21884.76 Da; observed mass 21883.62 Da); (c) HisBfr deconvoluted spectra (theoretical mass 19560.15 Da; observed mass 19559.17 Da); (d) HisGFP deconvoluted spectra (theoretical mass 28111.68 Da; observed mass 27959.38 Da); (e) HisAfFtn deconvoluted spectra (theoretical mass 21381.24 Da; observed mass 21247.18 Da).

3.1.2. Purification and characterization of *Pyrococcus Furiosus* Ferritin (PfFtn)

The third thermophilic ferritin studied in this project was the hyperthermophilic archaeon *Pyrococcus furiosus* ferritin (PfFtn). PfFtn has been reported to show increased Fe(II) oxidation at 85 °C with a half-life of 48 h at 100 °C and 85 min at 120 °C.[131] The PfFtn is similar in structure,

dimensions and mass to Bfr (but does not contain the heme cofactor) and AfFtn, but has the octahedral symmetry in Bfr unlike the open triangular pore of AfFtn (tetrahedral symmetry). [131] The PfFtn construct used here has a C-terminal hexahistidine tag which is on the interior of the capsule which was used for IMAC purifications. Due to the similarity to the other His-tagged proteins used, the same purification protocol was initially used to produce pure HisPfFtn. However, no protein was observed in the 100 % imidazole gradient despite heating the samples at 80 °C for 10 min before loading unto the column (Figure 3.1.2a). The flowthrough and the protein peaks from the imidazole gradient were collected and analyzed using 15 % SDS-PAGE, but HisPfFtn protein bands were not observed in the gradient peak; the protein was found in the flowthrough. Analysis of the SDS PAGE gel image (data not shown) indicated that the gradient peak, in Figure 3.1.2a, was due to the high concentration of imidazole (300 mM) in the elution buffer and did not contain protein.

The elution of HisPfFtn in the flowthrough indicated that none of the HisPfFtn was bound to the IMAC column and perhaps the protein existed in its 24-mer form, with unexposed hexahistidine tags on the inside of the capsule. Hence the sample was reheated at 100 °C for 30 minutes to attempt to purify the protein in a one-step heat treatment as reported previously for this protein by Tatur *et al.*[131] However, the protein produced from this step showed conterminant bands on SDS-PAGE gel, necessitating further purification. Hence, it was further purified by IMAC chromatography. The equilibration buffer was supplemented with 6 M GdnHCl (25 mM HEPES, 150 mM NaCl, 20 mM imidazole, 6 M GdnHCl-pH 7.5) and the same for the elution buffer (25 mM HEPES, 150 mM NaCl, 300 mM imidazole, 6 M GdnHCl-pH 7.5) to decluster the subunits and expose the C-terminal H_6-tags. The elution profile presented in Figure 3.1.2 b shows that HisPfFtn eluted in the 100 % elution buffer and was verified by both 20% SDS-PAGE (not shown) and mass spectrometry (Figure 3.1.2c). The mass spectrometry ionization produced a single mass peak at 21373.232, which corresponds to the M-1 peak. Pure HisPfFtn was then dialyzed against 150 mM NaCl, 25 mM HEPES overnight and quantified using its UV-vis absorbance ($\varepsilon_{280\ nm}$= 31400 cm^{-1} M^{-1}). The amount of purified HisPfFtn was estimated to be 20 mg/L of cell growth which is lower than the average 30 mg/mL yield obtained from the other 2 ferritins. The decrease is due to the several steps involved in the isolation. Finally, the quaternary structure of the isolated protein was analyzed in solution by SEC and DLS analysis. Analysis of the SEC profile, Figure 3.1.2d, indicated that the PfFtn existed entirely in its 24-mer form, and

analysis of the DLS also yielded an average size distribution of 15±2 nm (Figure 3.1.2e), consistent with its 24-mer quaternary structure.
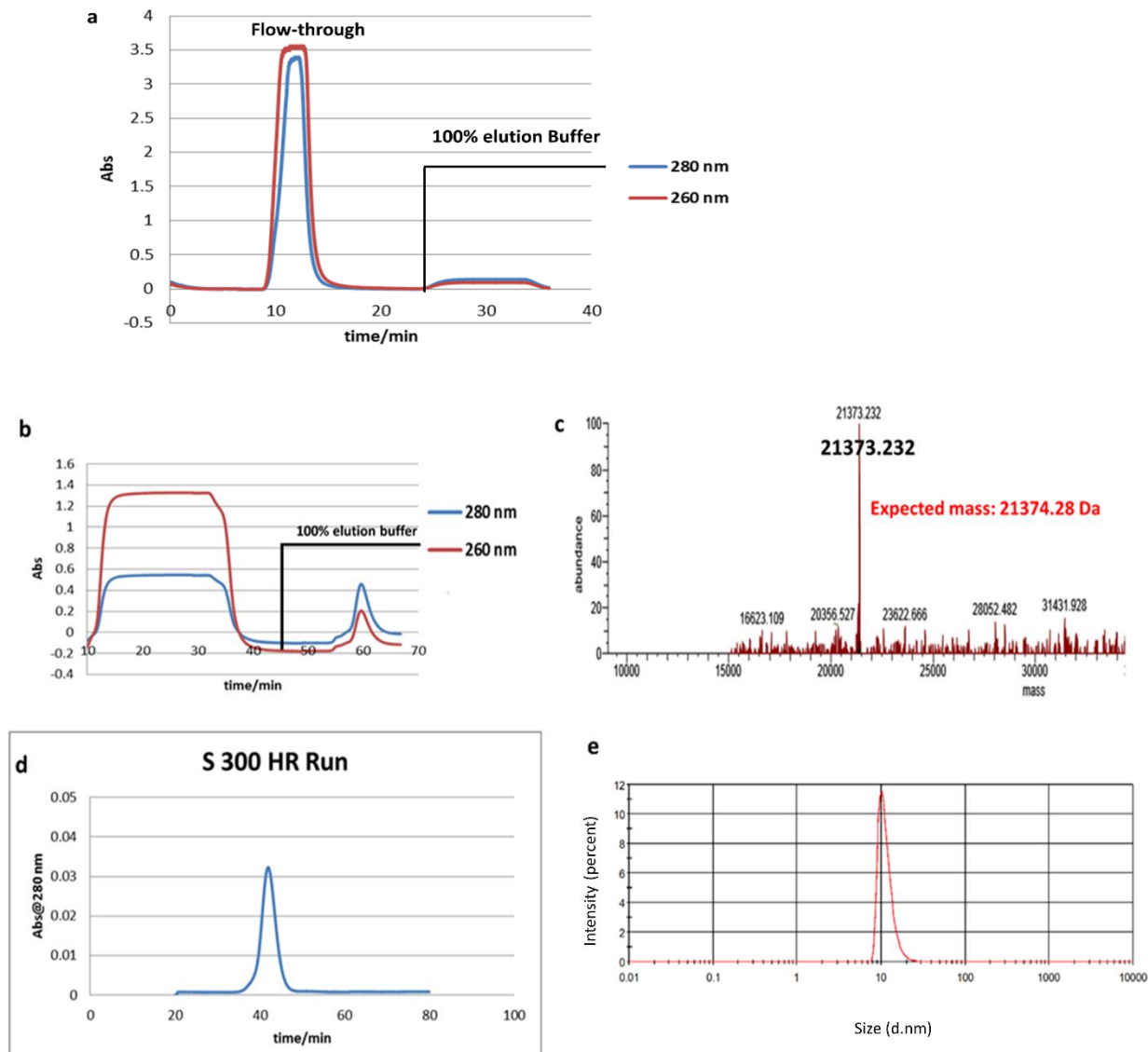
Figure 3.1.2. Purification and characterization of HisPfFtn. (a) The elution profile of HisPfFtn without 6 M GdnHCl (b) Elution in the presence of 6 GdnHCl (c) +ESI-MS data of the pure HisPfFtn (d) Size exclusion chromatography profile of pure HisPfFtn (d) DLS analysis of HisPfFtn (in150 mM NaCl, 25 mM HEPES-pH 7.5).

3.1.3. Production of Recombinant H5 Enterokinase and HWTBfr

Previously WTBfr protein from *E. coli* was purified by heat treatment followed by three chromatographic steps: ion exchange chromatography followed by hydroxyapatite chromatography and lastly size exclusion chromatography (Anton van der Ven, MSc). Each of these steps resulted in protein loss and consequently, low yields of isolated protein. Hence there was the need to find an alternative approach for the WTBfr protein production. A new Bfr construct was recombinantly engineered with an N-terminal decahistidine (H_{10}) tag and an enterokinase (EK) cleavage site. The H_{10} tag was used for IMAC purification which was subsequently removed via an EK cleavage reaction. The tag removal resulted in the isolation of WTBfr with an N-terminal histidine amino acid (HWTBfr).

Enterokinase (Enteropeptidase) is a serine protease (in the duodenum) that converts trypsinogen to active trypsin. The trypsin catalyzes the hydrolysis of food proteins as well as activates digestive zymogens into their active enzymes. EK is composed of two chains, light and heavy chains, which are connected via a disulfide bond. The heavy chain with its transmembrane domain is responsible for anchoring the enzyme to the intestinal brush border while the light chain serves as its catalytic domain.[132,133] The catalytic domain is capable of maintaining the enzyme's catalytic activity with or without the covalently linked heavy chain. The catalytic domain recognizes the DDDDK-X or DDDDR-X sequence and cleaves C-terminal to K or R [134] [135]. Figure 3.1.3 shows the catalytic domain of the EK light chain in bovine enteropeptidase complexed with a trypsinogen inhibitor analog, VDDDK.[136]

Figure 3.1.3. A closeup view of the active site of bovine enteropeptidase light chain bound to a VDDDK ligand (PDB ID: 1EKB). The C-terminal lysine (Lys 306, ball and stick representation) from the peptide is covalently bound to His 57 and Ser 195 from the enzyme as reported by Lu *et al*. [136] (image prepared with Maestro)

The HTEKBfr constructs were initially cleaved with a commercial EK enzyme (New England BioLabs, Inc.) which produced a yield of approximately 70 % of HWTBfr from HTEKBfr, but the EK enzyme was not efficient as it required a high concentration to obtain a few milligrams of cleaved HWTBfr. The typical cost of producing HWTBfr from HTEKBfr, with commercial EK, was high; the typical 25-30 mg of HTEKBfr (per L of growth) was estimated to require about 12800 units (27 vials; calculated total cost of ~ $ 4500) of EK.[137] Hence, the protocol from Skala and co-workers was adapted. [127] This research group engineered a truncated (light chain) bovine EK (EK_C122S_His5) which similar to Figure 3.1.3, but contained a pentahistidine tag and required an *in vitro* folding procedure. The C-terminal polyhistidine tag facilitated enzyme purification as well as the removal of the enzyme after the His-tag cleavage reaction. Another advantage of this engineered EK (EK_C122S_His5) over the commercial

enzyme is its ease of removal by IMAC chromatography, unlike the commercial source that requires a benzamidine column purification. The elution buffer for the benzamidine column is very acidic and, in some cases, resulted in protein precipitation. Also, the C122S mutation present in this engineered EK eliminates the cysteine residues that form a disulfide bridge with the heavy chain and prevents inclusion bodies. Moreover, two mutations, P92R and E152D, in the sequence, do not impact the native EK structure. Finally, there is an N-terminal propeptide, MGPIDDDDK, which autoactivates the enzyme.[127]

The amount of active enzyme produced, 1.5 mg/L, was similar to the reported yield of 2 mg/L despite the slight variations in the protocols. (Appendix 1 for SDS page gels and chromatograms). EK cleavage reactions were carried out on HTEKBfr constructs using the pure EK_C122S_His5 enzyme under different buffer conditions to obtain an optimal set of conditions. Three buffer systems were employed; the buffer suggested for the commercial EK (50 mM NaCl, 2 mM $CaCl_2$, 20 mM Tris-HCl-pH 8.0), the buffer used by Skala *et al*. (50 mM NaCl, 50 mM Tris-pH 7.5) and a Bfr substrate buffer supplemented with $CaCl_2$ (100 mM NaCl, 2 mM $CaCl_2$, 50 mM Tris-pH 8). The EK_C122S _His5 construct was expected to cleave similarly to the commercial light chain, but there were additional peaks detected in the +ESI data (Figure 3.1.2) which would appear to correspond to nonspecific cleavage at X+1 (18494 Da) and two others in the flexible regions of the tag (19702 Da and 19528 Da). The cleavage pattern observed confirms the reported promiscuous behavior of the EK enzyme,[127,133,135,138] and necessitated a new protocol for WTBfr production as described in the next section.

What was learned from these experiments despite the outcome are not limited to: (1) acquiring the technique of isolating an active ezyme, (2) mastered the technique of unfolding and carefully refolding an enzyme to remove inclusion bodies and activating the ezyme, (3) how to adjust the chrotatographic condition,benzamidine chromatography, to preserve the intergrity of the purified proteins. Moreover, documenting the outcome contributes to the knowledge about fussion tags and how they can be used for complex systems.
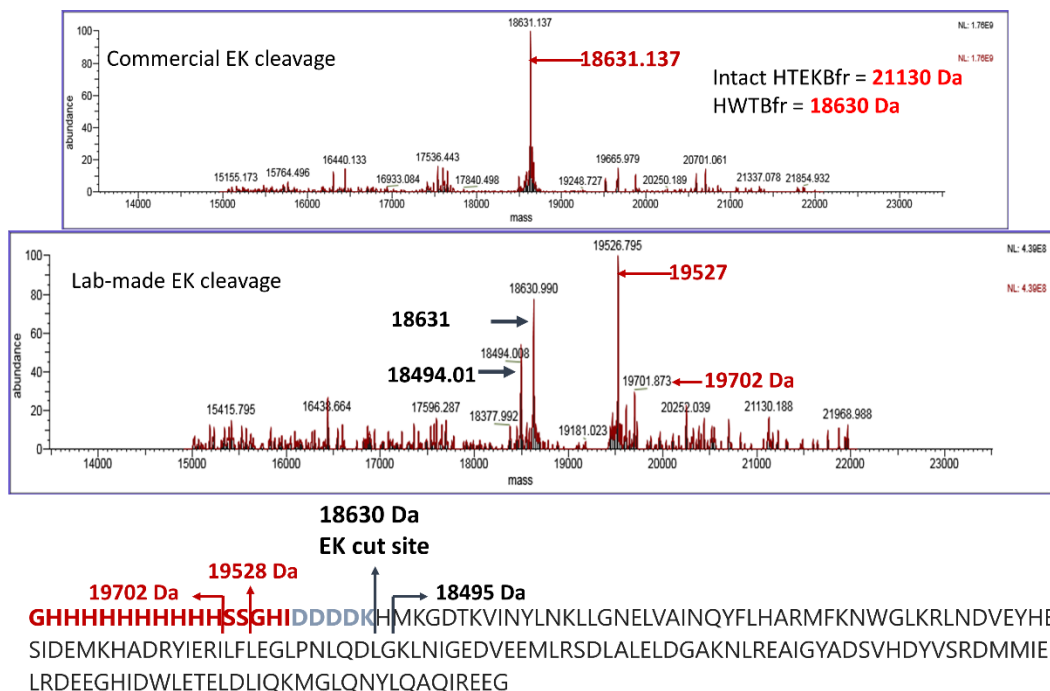
Figure 3.1.2. +ESI MS results for enterokinase His-tag removal. The top panel shows the total conversion of HTEKBfr to HWTBfr using the commercial EK while the bottom shows a mixture of fragments from the expressed EK enzyme. The sequence indicates the possible cleaving sites.

3.1.4. Removal of the EK Enzyme after Cleavage

Since the expressed EK enzyme did not produce reasonable amounts of HWTBfr, the remaining stock of the HTEKBfr was cleaved with the commercial EK enzyme. After the cleavage reaction, the enzyme was removed by running a benzamidine column which removes serine proteases such as EK. The buffer contained 0.5 M NaCl and 0.05 M Tris-HCl- pH 7.5 while the elution buffer contained 0.5 M NaCl and 10 mM HCl- pH 2. Due to the acidic nature of the elution buffer, the fraction tubes were supplemented with 200 µL (each) of 1M Tris-HCl -pH 9 to neutralize and stabilize the eluent. Figure 3.1.3 below shows the elution profile from a 1 mL benzamidine column running at 1 mL/min. Most of the cleaved proteins were observed in the 100 % elution buffer (Figure 3.1.3b, lane 5) with little unreacted HTEKBfr (Figure 3.1.3b, lane 4) which was not ionized on the +ESI since it showed the cleaved product in 100 % abundance (Figure 3.1.2, top panel).

The final step was to run the isolated HWTBfr fraction (from lane 5) on size exclusion chromatography and to obtain the quaternary structures in solution. The sample was exchanged into SEC buffer (150 mM NaCl, 50 mM HEPES-pH 7.5) and loaded onto a Sephacryl™ S-300 10/300 HR column. The elution profile showed HWTBfr as dimeric. However, the presence of heme did not shift the protein to higher oligomers, unlike the other Bfr constructs. A significant shift toward the 24-mer form required extended heating at elevated temperature, as seen in Figure 3.13c. The extended heating led to significant protein loss and may not serve as a useful host for the intended purposes. Therefore, no attempts were made to separate the reaction from the lab-made EK cleavage. Even though the lab-made EK cleavage showed fragments that corresponded to WTBfr, its abundance was low relative to the HWTBfr and others (Figure 3.1.2. bottom panel) requiring more purification steps to produce very little WTBfr.

Despite the unexpected results obtained here, it highlights the fact that the association of Bfr subunits beyond dimers is greatly affected by the N-terminal sequence and even the presence of a single bulky amino acid (like histidine) can altogether abolish 24-mer formation. The previous WTBfr production from WTBfr plasmids (without any tags) was further optimized, in the next section, to produce WTBfr for protein engineering.
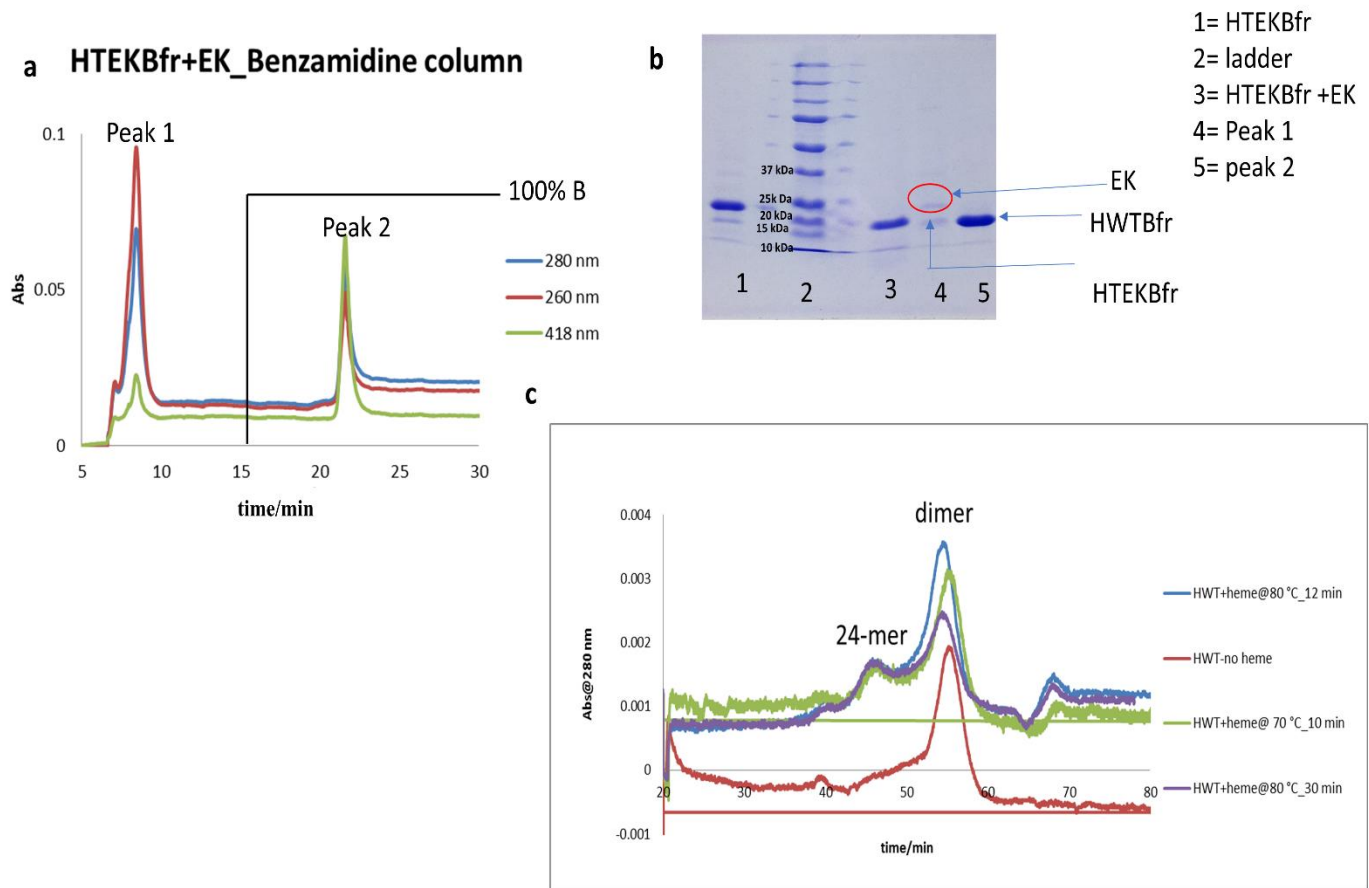
Figure 3.1.3. Enterokinase removal and characterization of HWTBfr. (a) The serine protease removal with 1 mL benzamidine FF column, flow rate 1 mL/min. (b) The SDS-PAGE (15 %) analysis of the fractions from (a). (c) Sephacryl™ S-300 10/300 HR column profile run on HWTBfr with and without heme at different temperatures. HWTBfr heme (red), HWTBfr + heme at 80 °C for 12 min (blue), HWTBfr + heme at 70 °C for 10 min (green), and HWTBfr + heme at 80 °C for 30 min (purple).

3.1.5. Wild-type (WT) Bacterioferritin Purification

The WTBfr from constructs were expressed in *E. coli*, and the cells were harvested and lysed similarly to the His-tagged constructs. The crude supernatant from the cell lysate was heated for 10 min at 70 °C and allowed to cool to room temperature before centrifugation to remove insoluble and denatured protein from the lysate. The protein solution was applied to an ion exchange column, BioRad UnoQ1, in binding buffer and eluted with a 1 M NaCl gradient (Figure 3.1.4). SDS analysis on the fractions showed about 80 % of the protein in the salt gradient (Figure 3.4) with a few higher molecular band proteins, requiring further purification steps. The Sephacryl™ S-200 HR, 10/300 column was used for this purpose. The pooled fractions were concentrated down to 10 mL in the eluent buffer (50 mM MOPS, 1.5 M NaCl, pH7.5), the higher salt concentration aided DNA removal. Analysis of the gel image (Figure 3.1.4 b) showed significant loss of protein in the flowthrough, but exhibited higher 260 nm absorbance than 280 nm absorbance, indicating high DNA contamination and hence these fractions were not collected. The 100 % gradient fractions showed higher molecular weight protein bands, which necessitated an additional purification step. The Sephacryl™ S-200 10/300 HR column (running at 0.5 mL per minute, in 500 mM NaCl, 50 mM MOPS-pH 7.5) was used to remove contaminants; the elution profile generated two well-resolved peaks that were analyzed on 15 % SDS-PAGE for WTBfr bands. The first peak, at retention time ~20-30 min, was WTBfr (Figure 3.1.4 c, gel), but the second peak at time 35-43 min did not show any WTBfr protein bands. The 260 nm absorbance in the fractions of peak 1 significantly decreased. However, the higher molecular weight bands persisted in lanes associated with the WTBfr peak, which were initially assigned to oligomeric forms of WTBfr (monomer, 18.5 kDa; dimer, 37 kDa; tetramer 74 kDa). There were no visible protein bands in peak 2.

Positive ESI mass spectrometry was used to validate the mass on the WTBfr fractions, and the deconvoluted spectrum showed a WTBfr protein mass (Figure 3.1.5.) as the most ionizable species in solution. The abundance of monomeric WTBfr could imply that the protein solution is purely WTBfr, and the higher molecular weight bands resulted from subunit association. The association could have occurred upon cooling of the SDS samples before electrophoresis. The quaternary structures were analyzed by employing size exclusion chromatography. The WTBfr protein existed as an ensemble of 24-mers, dimers, and monomers (Figure 3.1.6). The shoulder

peak (RT, 40 min) corresponds to aggregates, and the broad peak is a distribution of WTBfr oligomers.

Despite the reasonable amount of WTBfr protein (22 mg/L) obtained from the above protocol, there was a significant amount of protein loss in the flowthrough that resulted in low protein yield. This problem persisted after column regeneration. Hence an ammonium sulfate precipitation protocol (Figure 2.2.1) was utilized after the heat treatment and DNase I incubation steps. This protocol replaces the ion exchange chromatographic step, yet produced pure proteins with reasonable yield (approximately 35-50 mg of protein per liter of cell growth).
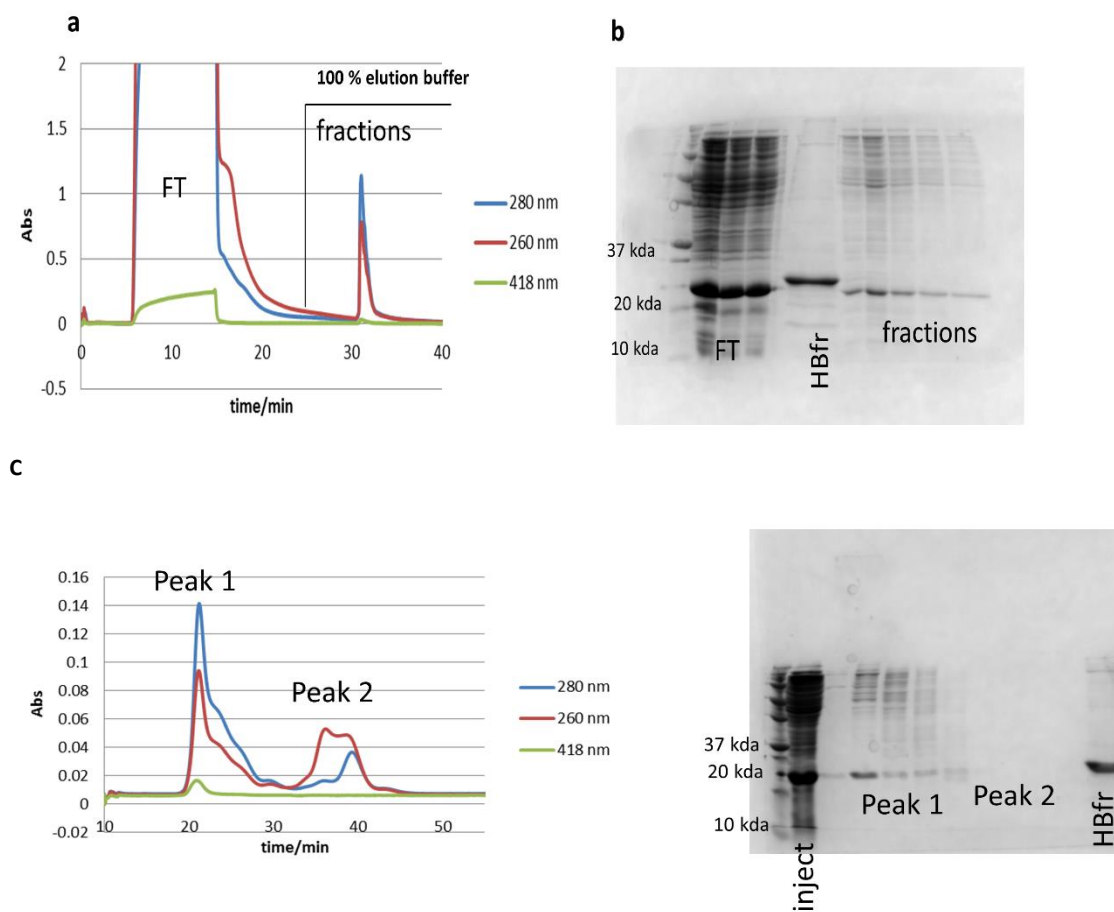
Figure 3.1.4. Ion exchange and size exclusion chromatography. (a)The elution profile of loaded WT lysate shows two peaks, flowthrough (FT) and fraction peaks in the high salt gradient. (b) SDS-PAGE gel with Pure HisBfr (HBFr, 19.5 kDa) as a reference. (c) A Sephacryl™ S-200 10/300 elution profile for unoQ1 fractions and its corresponding SDS-gel image.
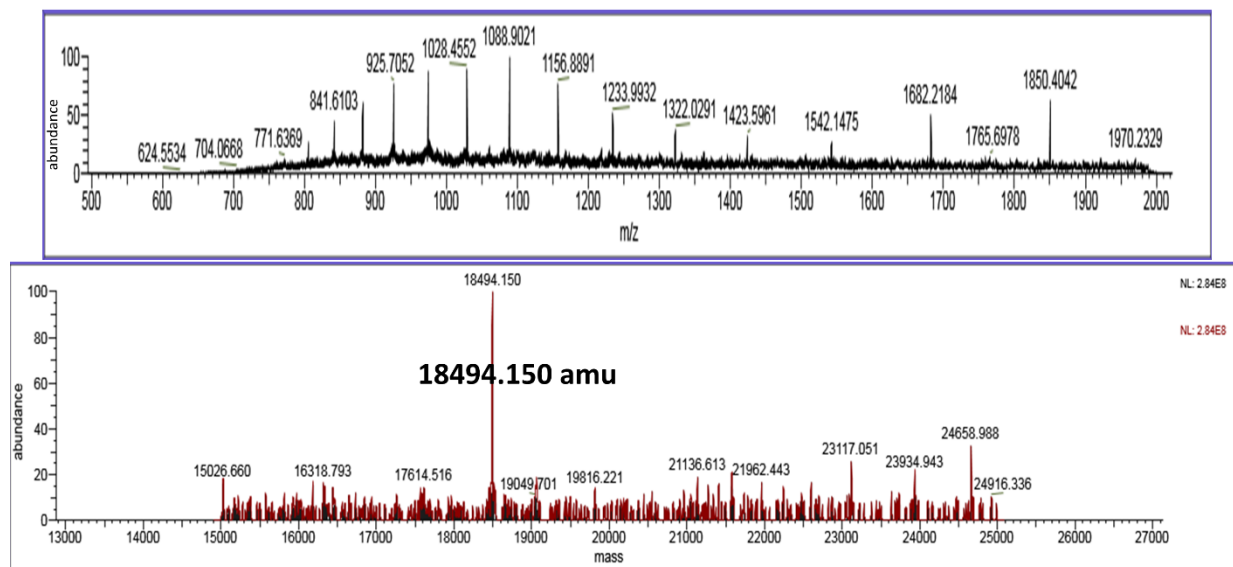
Figure 3.1.5. Mass spectrum of pure WTBfr. The observed mass was 18494.150 amu was 1 amu less than the expected mass, 18495.03. The +17 ion (1088.9021 @ 1100 m/z) is the most abundant peak source spectrum (top panel).
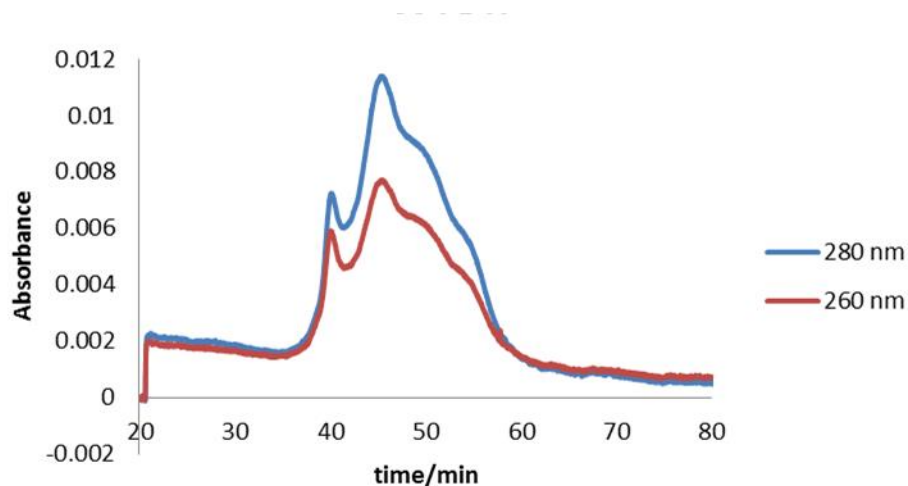
Figure 3.1.6. Size exclusion chromatography of WTBfr after Uno Q1 and S-200 HR runs. The Sephacryl™ S-300 10/100 HR column was run at 0.5 mL/min with 150 mM NaCl, 50 mM HEPES -pH 7.5 buffer.

3.2. Heme Encapsulation

The Bfr heme cofactor has been reported to drive the formation of the 24-mer quaternary structure of this protein[88] as well as to facilitate iron release from the ferroxidase site.[102,100] The ability to remove and reinsert heme into the dimer interface makes the Bfr platform amenable to heme-mediated modifications without employing changes in the amino acid residues in the Bfr protein. The heme encapsulation is monitored via UV-Vis spectroscopy due to the heme group's spectral changes upon binding to Bfr, and the resultant oligomerization can be visualized utilizing size exclusion chromatography (SEC) as well as transmission electron microscopy (TEM). Size exclusion chromatography was used to visualize the oligomeric forms (quaternary structures) of the protein in solution by loading purified protein samples onto high-resolution Sephacryl™ columns. The relative molecular masses of the proteins were estimated from calibrations curves (from the separate columns. Appendix 3). SEC was also used to isolate encapsulated Bfr from free Bfr monomers and dimers after host-guest interaction, based on their hydrodynamic radii or distinct host absorbance peaks.

Unlike other heme-binding proteins which require the heme cofactor for proper folding (example, mitochondrial cytochrome c),[139] Bfr properly folds in the absence of its heme cofactor, into apoBfr, leaving an empty heme-binding site. The recombinant Bfr proteins used in this work initially contained deficient levels of the native heme cofactor which could be ascribed to the overproduction of high levels of proteins within a short period with the inability of heme biosynthesis to provide the high levels of heme required for complete holoprotein formation. Hence most of the Bfr constructs were initially monomeric or dimeric in solution and required the addition of exogenous heme to form the intact 24-mers. Heme binding begins with a subunit dissociation step (declustering), followed by heme addition and a final reassociation (reclustering) step. The declustering agent used must sufficiently open up the heme-binding interface, hence temperatures ranging from 70 °C to 80 °C were used. At the encapsulation temperature, the Bfr molecules were dissociated, but not denatured.[101] Another route that was investigated was incubating at lower temperatures (35-60 °C) for extended times (1-2 h). However, incubating the Bfr and the heme at these lower temperatures for extended periods generated only dimers in solution (data not shown). The products of the binding interactions were visualized by employing UV-Vis spectroscopy or the appearance of a Soret band at approximately 415 nm in the protein/heme absorbance spectrum (Figure 3.2).

Additionally, the reaction mixture was assessed using SEC for the nature of quaternary structures in the solution since dimers, trimers, and other intermediate forms of Bfr also produce Soret bands. The Soret band results from the axial ligation of methionine 52 to the heme molecule.[100] Figure 3.2 illustrates the heme encapsulation process for WTBfr with the UV-vis scan (left panel) showing the emergence of a Soret band upon adding heme to apo WTBfr. A size exclusion profile of the encapsulated WTBfr generated a major 24-mer band (green and purple curves) from the heme reconstituted WTBfr, unlike the ensemble of oligomers in the apo WTBfr (blue and red curves).
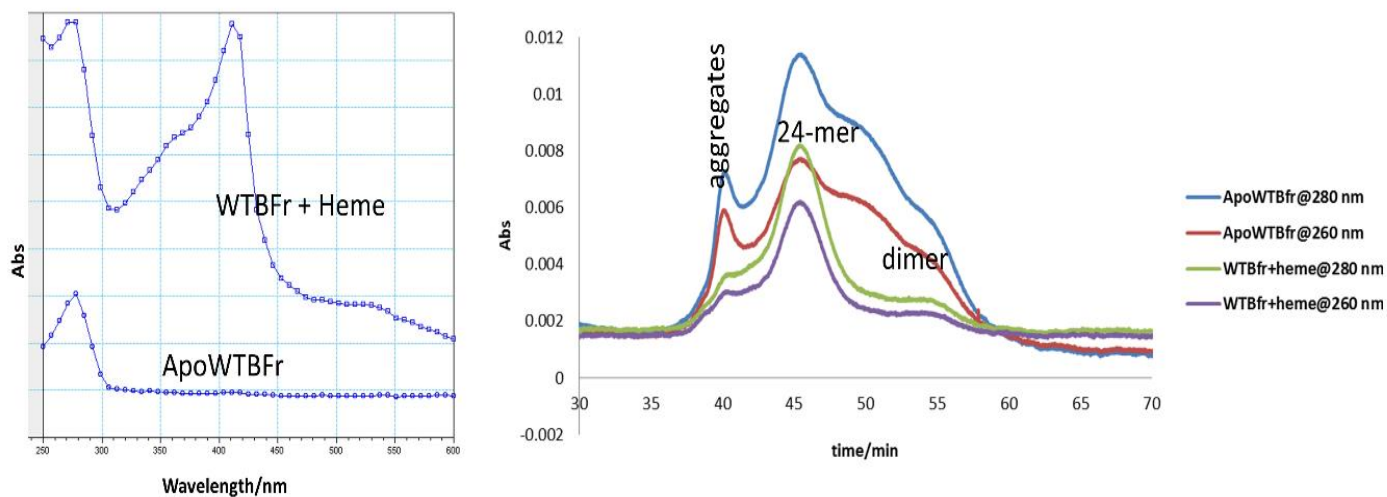
Figure 3.2. WTBfr heme encapsulation. The SEC column, Sephacryl™ S-300 10/300 HR, was run at 0.5 mL/min in a 150 mM NaCl, 50 mM HEPES, pH-7.5 buffer.

3.2.5. Sedimentation Velocity Studies

Even though WTBfr may not be required for most of the guest encapsulation studies (except His-tagged guest molecules), much effort was made to obtain significant quantities of WTBfr since it has been demonstrated to control the size of the Bfr capsule.[120] The process of mixing WTBfr and HisBfr constructs is useful for reducing the internal steric crowding associated with the multiple his-tags in the 100 % HisBfr capsule. Previous research by Anton van der Ven[120] showed that larger and dynamic guest molecules such as streptavidin could be

88

encapsulated by using a mixture of 40 % WTBfr and 60 % HisBfr. Also, C19-NTA lipid encapsulation studies, described in section 3.6, requires a 50 %: 50 % mix of HisBfr and WTBfr to regulate the Bfr capsule dimensions. Hence at least 40 % of HisBfr in a capsule is sufficient to stabilize the guest via His-Ni^{2+}-NTA interactions.

Analytical Ultracentrifugation analysis was used to study the overall sedimentation patterns of Bfr samples: (a) 100 % HisBfr + heme and (b) 50 % HisBfr + 50 % WTBfr + heme. The heme was used to lock the subunits into the 24-mer conformation. The primary aim of the AUC experiments was to characterize the two Bfr mixtures. The expected molecular weight of the 100 % HisBfr based on the amino acid sequence is 469 kDa, whereas the 50 % HisBfr + 50 % WTBfr is 456.6 kDa assuming they exist as $\alpha_{12}\beta_{12}$. Also, other oligomeric forms are expected since there is an equilibrium between subunits and the 24-mer in solution (even in the presence of heme).

The sedimentation velocity plots (Figure 3.2.1) were obtained by fitting the velocity and the residuals scans Appendix 10) with the Sedfit and normalizing in Gussi. The sedimentation velocity analysis on samples 100 % HisBfr + heme, a) and 50 % HisBfr + 50 % WTBfr, b) produced slightly different sedimentation patterns shown in 3.2.1. However, the degree of symmetry (f/fo) similar for both samples with sample a) having a value of 1.2 and b) having a value of 1.17; hence, the mixing did not affect the roughly spherical arrangement of subunits. Overall, sample b (Figure 3.2.1. b) which is a mixture of WT and His Bfr showed fewer species than the 100% HisBfr sample, the predominant species having a molecular weight of 472.3 kDa whereas the mixed sample, b exhibited a molecular weight of 498.8 kDa, consistent with the 100 % Histagged Bfr having the greater molecular weight. The data indicates that there could be mixing of subunits in the 50 % WTBfr + 50 % HisBfr subunits which led to lower molecular weight species in solution.
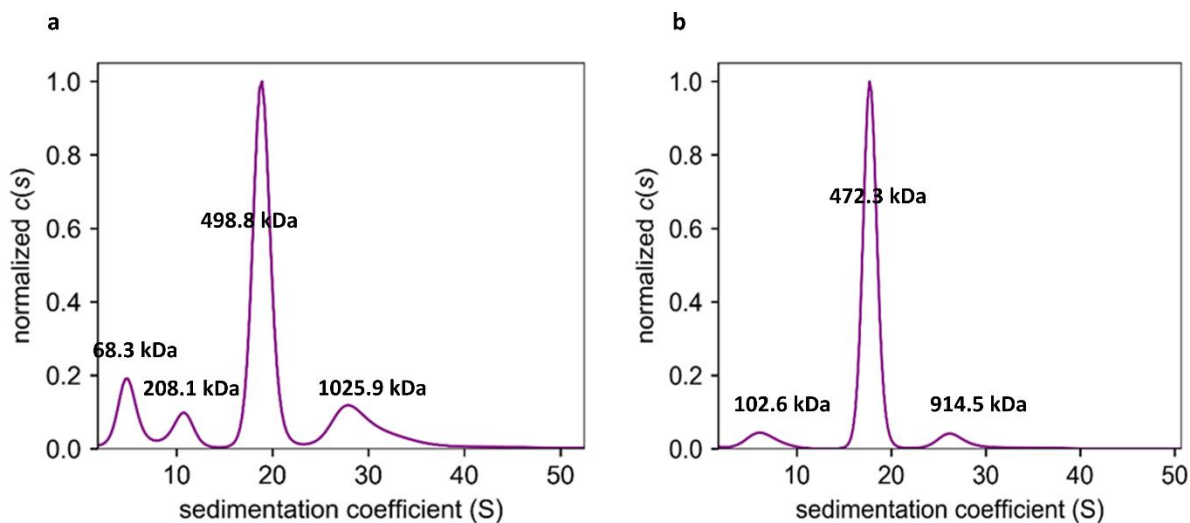
Figure 3.2.1. The distribution of species in solution. For: (a) 100 % HisBfr + heme showing 4 populations of molecular weights: 68.3 kDa with 5.0 S (sedimentation coefficient), 208.1 kDa having 10.6 S , 498.8 kDa with 18.9 S, and 1025.9 kDa with 30 .6 S. (b) 50 % HisBfr + 50 % WTBfr + heme has 3 populations: 102.6 kDa, 472.3 kDa , and 914.5 kDa MWs having sedimentation coefficients of 6.4 S, 17.7 S, and 27.6 S, respectively. Sedimentations were performed at 20 000 rpm.

3.3. Transglutaminase (TGase)-catalyzed Surface modification

Transglutaminase is an enzyme that catalyzes the posttranslational formation of a covalent isopeptide bond between a γ-carboxyl group of glutamines in a protein of interest (POI) and an incoming primary amine, ε-amino lysine of another protein, and deamination of glutamine (when the donor group is replaced by water).[140] The reaction is illustrated in Figure 3.3. The covalent nature of the new bonds that are formed makes them resistant to proteolytic degradation, and they have exceptional mechanical properties. Hence TGases have become useful in food processing, and material sciences as well as health.[141,142] These TGases have become popular in the food industry due to their abilities to form cross-linked products that improve the texture of foods such as wheat proteins, beef myosin, soy proteins, and whey proteins.[141] For example, TGase has been used to produce cuts of meats (Steak) by joining small pieces of beef.
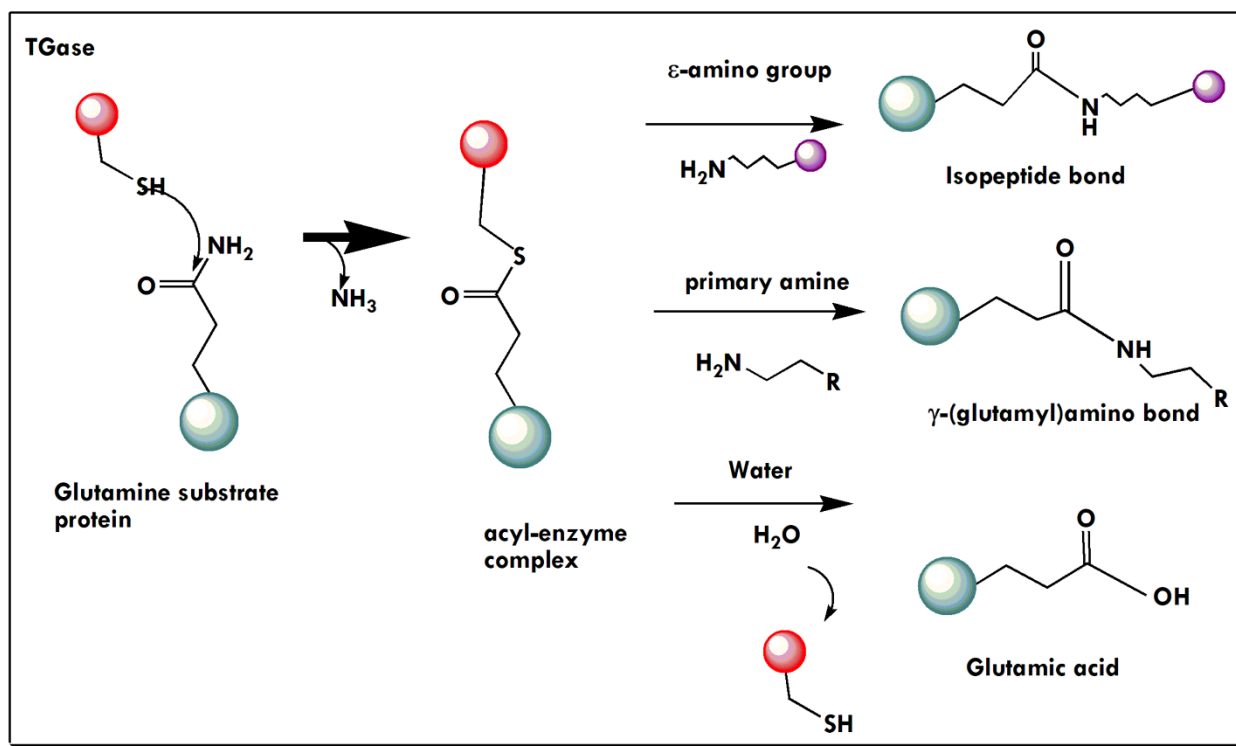
Figure 3.3. Transglutaminase-catalyzed modification: The TGase enzyme forms an acyl complex which is modified with primary amines, water, and ε-amino groups from proteins to produce glutamic acid, γ-amino bonds, and isopeptide bonds, respectively. Image was generated in ChemDraw.

The Bfr surface was modified with five different N-terminal tags to produce five new constructs for Bfr exterior modification. These constructs also contain C-terminal hexahistidine tags. The first four are Qtags in that they are transglutaminase substrates. Qtag1 and Qtag2 can also reconstitute RNase S activity if the RNase S-protein is added to these tags. The Qtag1 tag is composed of the first 15 amino acids from the RNase S-peptide. The 15 amino acids were genetically linked to Bfr through a PDLH spacer from the pET-29b (+) vector used to express this construct, described in the experimental section (section 2.1.1). The Qtag-Bfr constructs become amenable to several forms of TGase mediated modification that could be introduced: intact proteins, antibodies, fluorophores, and many useful molecules (that present accessible primary

amines). The data illustrates how a primary amine, dansylcadaverine could be modified on the Bfr exterior surface via a TGase mediated reaction (Figure 3.3.1). Also, the dansylated Qtag2-Bfr is capable of encapsulating a guest molecule (5 nm AuNP). Both exterior and interior modification demonstrates that Bfr could be used as a targeted guest delivery machinery.

The Qtag-Bfr constructs were covalently modified using TGase with dansylcadaverine (DC), molecular weight (MW) 335.465 gmol^{-1}, to generate fluorescent Qtag-Bfr products. The products were visualized on 15 % SDS-PAGE gels (unstained, Figure 3.3.1) and their intensities were measured using a PTI fluorimeter, using the excitation and emission wavelengths of dansylcadaverine which are 325 nm and 525 nm. Finally, electrospray ionization mass spectrometry (+ ESI-MS) was used to obtain the masses of the products (Figure 3.3.2).

The SDS-PAGE gel image (Figure 3.3.1) showed higher molecular weight bands, which could result from cross-linked subunits. TGases have been reported to show strict specificity towards glutamines; the desired Gln has to be in the preferred sequence. However, TGase modification of ε-amino Lysines does not require the lysines to be in a specific sequence; in general, any accessible lysines could be crosslinked.[128] The subunits of Bfr display six exposed lysines per subunit (Fig. 3.3.4a) that can potentially form crosslinked products with adjacent subunits. Moreover, the higher molecular weight labeled bands persisted after boiling the SDS sample, confirming that they are crosslinked lysine products since labeled Bfr dimers dissociate upon heating at temperatures above 70 °C. Bfr constructs without the Qtag (HisBfr) were not labeled by the MTGase despite the exposed lysines and glutamines (lanes 5 and 9, Figure 3.3.1). However, the presence of dimer bands in these lanes may suggest lysine cross-linking in HisBfr. The absence of DC-labeled bands in HisBfr validates the design of the Qtag constructs. Even though there are surface glutamines on the HisBfr (three exposed per subunit, Figure 3.3.4 b), they may not be in a prefered TGase recognition sequence. Analysis of product fluorescence spectra and +ESI- MS data confirms that Qtag1-Bfr and the other Qtag constructs are MTGase substrates (Figures 3.3.2 and 3.3.3). The labeled peak masses are only 1 or 2 amu off from the expected masses (M-1 and M-2 peaks) which could result from the loss of 1 or 2 protons during ionization.
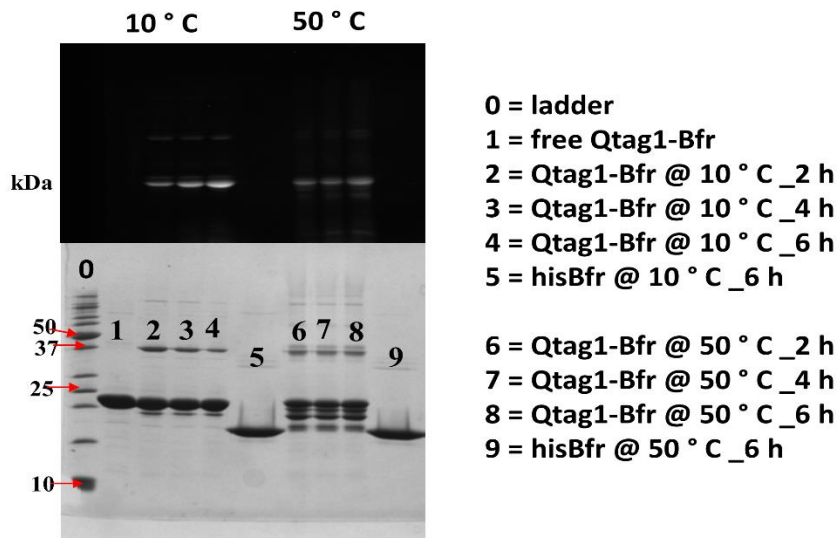
Figure 3.3.1. SDS-PAGE gel (15 %) showing the modification of Bfr-QTag1 substrate. The top panel is the fluorescent image of the unstained gel, indicating the presence of dansyl-labeled subunits as well as labeled cross-linked products. The lower panel shows the stained gel with a protein ladder (lane 0, units kDa).
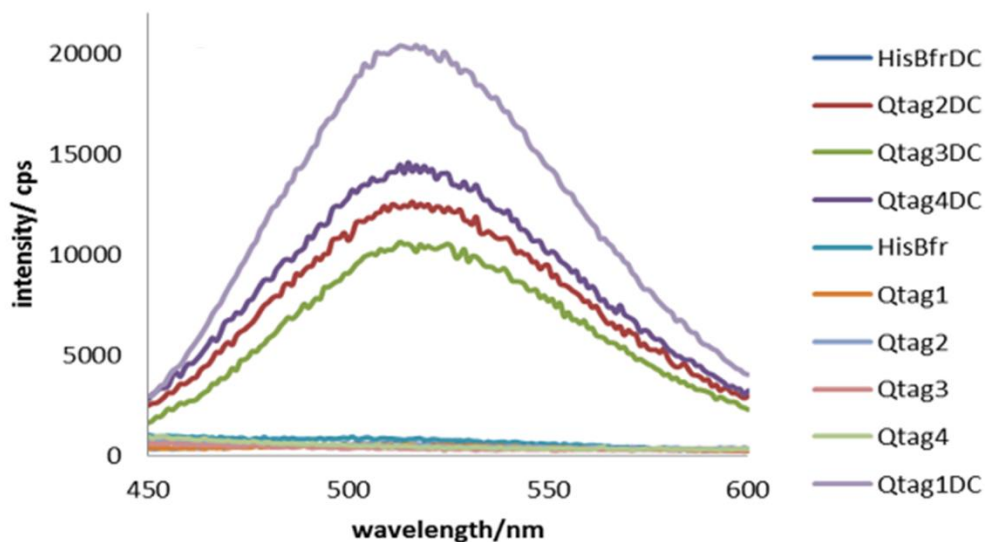
Figure 3.3.2. Emission spectra of dansylated QTag-Bfr constructs. Fluorescent measurements showing all DC labeled constructs except HisBfr are fluorescent (at the dansylcadaverine emission wavelength) when excited at 325 nm, excitation wavelength of dansylcadaverine. Excess DC from the reaction was removed with spin column before analysis.

Construct	Subunit mass (kDa)	Expected labeled subunit mass (kDa)	Observed labeled subunit mass (kDa)
HisBfr	19560.150	19819.460	19559.170
Qtag1-Bfr	21884.700	22204.160	22202.570
Qtag2-Bfr	21650.000	21969.460	21967.881
Qtag3-Bfr	20715.400	21034.860	21033.041
Qtag4-Bfr	20772.500	21091.960	21090.205
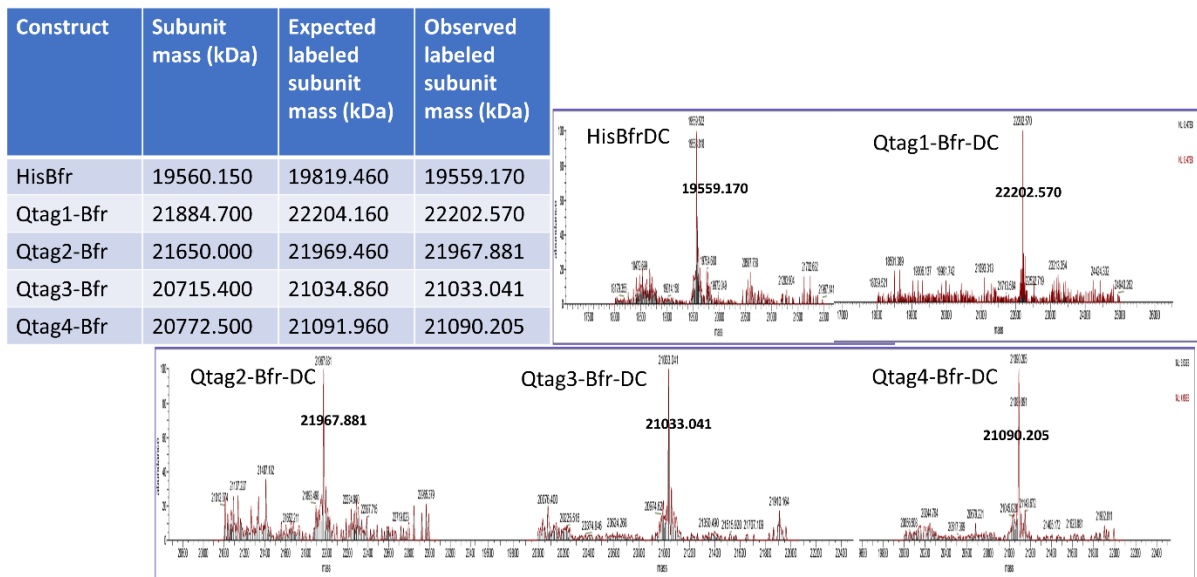
Figure 3.3.3. Positive ion ESI data for the MTGase catalyzed reactions. The table shows the masses of the unlabeled and labeled subunit and plots of abundance vs. mass for dansylated HisBfr (HisBfr-DC), Qtag1-Bfr(Qtag1-Bfr-DC), Qtag2-Bfr (Qtag2-Bfr-DC), Qtag3-Bfr (Qtag3-Bfr-DC), and Qtag4-Bfr (Qtag4-Bfr-DC).
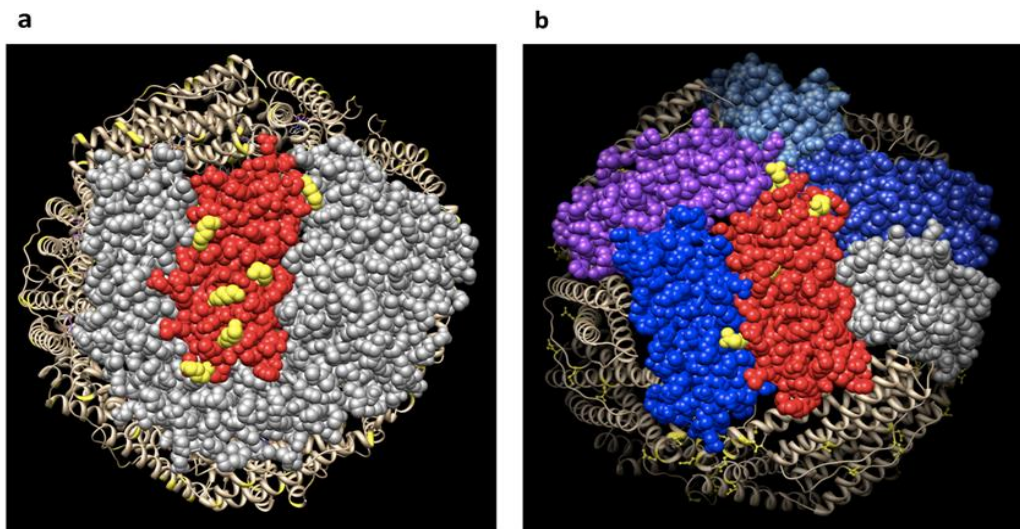
Figure 3.3.4. Exposed lysines and glutamines on a Bfr subunit. Panel a shows the surface lysines, in yellow, present on one of the Bfr subunits, whereas Panel b shows the exposed glutamines per subunit (also in yellow). Overall, there are nine lysines, six glutamines, and 19 glutamic acids per subunit. The images were generated in Chimera using the 1Bfr crystal structure.

94

3.3.1. Cyanogen Bromide (CNBr) Reactions of Dansylated Qtag-Bfr

The products of MTGase catalyzed reactions were treated with 50 mg/mL CNBr dissolved in 70 % formic acid. The formic acid solution is expected to unfold the protein to enable effective fragmentation of the protein. CNBr cleaves a protein by (i) The generation of a cyanosulfonium salt, via a nucleophilic attack on the cyanide group (of CNBr) by the sulfur atom of methionine. (ii) The cyanosulfonium intermediate is nucleophilically attacked by the methionyl carbonyl oxygen to produce methyl thiocyanate, byproduct, and an iminolactone. (iii) The thiocyanate is then hydrolyzed into C-terminal peptide and N-terminal peptide fragments (Figure 3.3.5).[143] The number of fragments that can be generated from Qtag-Bfr was estimated with the PeptideCutter tool [144] to obtain the fragmentation pattern, which resulted in nine peptide fragments with even distribution of all the glutamines in the Bfr subunits and the tags (Figure 3.3.6.)
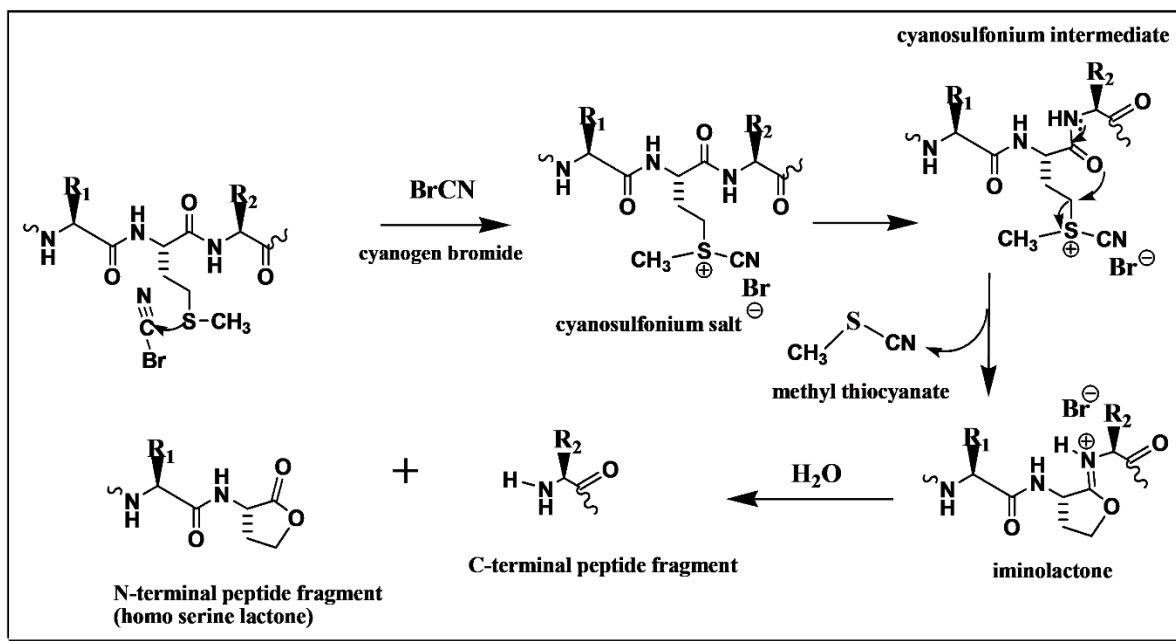
Figure 3.3.5. Mechanism of CNBr fragmentation. Figure was drawn with ChemDraw based on the example from: https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/protein2.htm
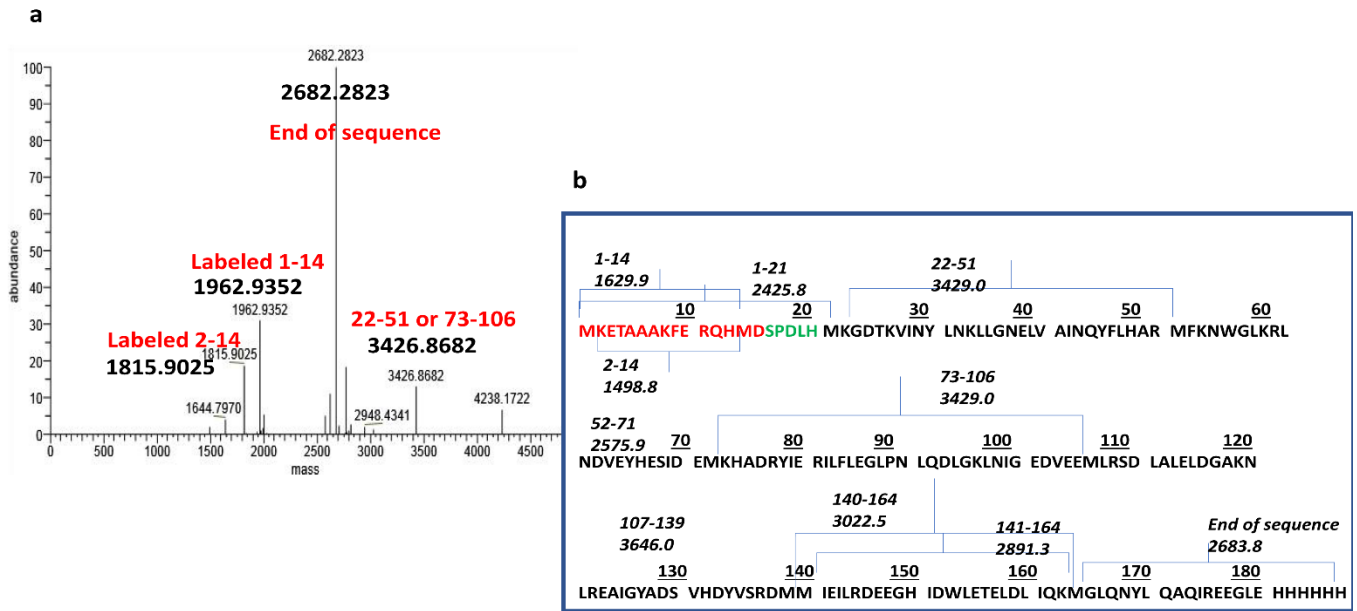
Figure 3.3.6. Cyanogen bromide fragments of Qtag1-Bfr-DC. (a) +ESI-MS results showing the end of sequence mass as the most abundant peak and three fragments from the N-terminus. (b) The nine estimated fragments obtained from the ExPASy peptide cutter tool.

3.3.2. Structural Effects of Bacterioferritin Surface Tags

The introduced peptide surface tags (Qtags, for example) have successfully converted Bfr into a transglutaminase substrate as illustrated above. However, it is also important to be able to reconstitute the intact 24-mer in the presence of these tags. The intact 24-mer makes the platform more versatile, and it potentially could allow for 24 copies of a protein/enzyme/antibody to be built on the outer surface of a single Bfr capsule. Modified Qtag surface peptides could allow for a versatile strategy to provide targeting molecules to decorate such a capsule, such as EphB4 targeting peptides which could target the Bfr to tumor cells expressing the EphB4 surface receptor.[145] Also, the modified capsule will be capable of hosting a guest molecule(s) within the hollow interior. The presence of the heme cofactor plays a significant role in quaternary structure of the Bfr and decidedly in modified forms of Bfr. However, it was observed that the amount of endogenous heme present in Bfr after isolation from *E. coli* cells is not sufficient to convert all Bfr molecules into their 24-mer quaternary form. Hence it was necessary to supplement the Bfr with exogenous heme. A solution containing two molar excess of heme (per subunit) was mixed

with Bfr in a high salt buffer (1 M NaCl, 200 mM MES-pH 6.5) and incubated at 70 ° C for 30 min or 80 °C for 10 min. This protocol was reported to by Wong *et al*. to be a satisfactory approach to removing and reinstalling the heme cofactor into Bfr [101]. The high ionic strength ensured that the native structure is restored upon cooling to room temperature, thus making the declustering and reclustering processes reversible. The declustering temperature of 70 °C was used for these experiments since it was found that incubating Bfr at 80 °C for 10 min causes extensive precipitation of the surface tagged Bfr variants (Qtags and Sort tags) but was found not to be a concern for either the HisBfr or WTBfr proteins. To ensure that the subunits are well dissociated to expose the heme binding pockets, the incubation times were extended to 30 min at 70 °C. Centrifugation (to remove excess heme) was performed at a medium speed 5018 *g* remove excess unreacted heme molecules. The nonspecifically bound heme molecules were further removed by dialysis or with desalting column. Figure 3.3.7 below shows the formation of the 24-mer structure from Bfr subunits in the presence of heme and declustering agents. The approach was further used in all the Bfr encapsulation studies in this project.

A UV scan is utilized to access the extent of encapsulation; free Bfr solutions exhibit the 280 nm absorbance as well as a 418 nm absorbance peak indicative of the presence of heme associated with the protein capsule, depending on the expression and purification process. Bfr bound with heme exhibits a blue shift of approximately 418 nm ($\varepsilon_{418} = 1.09 \times 10^5 \, M^{-1}cm^{-1}$)[101] with minimal 280 nm absorbance from this cofactor, the absorbance ratio of 280 nm to 418 nm was calculated, and if it was greater than 0.7, then 24-mer structures were expected in solution. Since the 418 nm (Soret band) could correspond to any tertiary or quaternary Bfr intermediate conformation, size exclusion chromatography was used to identify the intact 24-mer structure. The 24-mer structure eluted at the retention times of horse spleen ferritin (also a 24-mer of similar size as Bfr) and other ferritins of similar size (see appendix 2 for calibration profiles). All the Qtag-Bfr constructs were analyzed for their ability to form intact 24-mers with heme, and the results were compared to HisBfr (Figure 3.3.7).

Analysis of the data shows that apoHisBfr is mainly dimeric and monomeric. Heme addition shifts the equilibrium towards 24-mer production (red curve). In the case of Qtag1-Bfr, which was our initial Qtag construct, the apoQtag1-Bfr solution contained high mobility monomers which associated into dimers in the presence of heme. This observation prompted a review of the amino acid sequence at the N-terminal since it was presumed that the Qtag sequence

would not interfere with the subunit association. The Qtag peptide was expected to lie on the surface of the folded Bfr, but that was not the case. Reviewing a space filled model of the Bfr surface (Figure 3.3.8) the starting Met1 residue is adjacent to two bulky residues, Lys2 and Glu 66, which might be expected to be fairly mobile in conformation and allow sufficient room for the incoming tag to be accommodated. In fact, molecular dynamics computations undertaken on solvated bacterioferritin (20 ns dynamics run) exhibited substantial conformational movement of the Lys2 and Glu66 residues; data not shown). However, even though the recombinant modification was successful, the purified Qtag1-Bfr only existed as dimers in the presence of the heme cofactor. The twenty amino acid residues of the tag could have a secondary structure which blocks the interface of the subunits, hindering the 24-mer formation. The lack of a 24-mer quarnary structure in solution led to the design of a new Qtag2-Bfr with a flexible linker (PDLH was changed to GGGG). The SGGGG linker was used in order to extend the tag sequence away from the subunit interface.

Furthermore, the Qtag4-Bfr was redesigned based on the sequence of Qtag 3 (which was determined not to form a 24-mer). The addition of one extra glycine to the linker in Qtag4-Bfr did result in the formation of the desired 24-mers structure. The combined results show that an optimal length of glycine spacer (containing at least four Gly residues) is required to permit 24-mer formation.

Based on the above observations, the final sortTag1-Bfr construct was designed with a five glycine linker to facilitate the 24-mer conformation (Figure 3.3.7). This construct is expected to convert Bfr into an acceptor ligand that ligates to the acyl donor sequence transiently attached to the sortase enzyme. The process is simplified in Figure 3.3.8.5. The Sort tag1-Bfr constructs were successfully expressed and characterized for quaternary structure composition and were found to behave similarly to Qtag2-Bfr and Qtag4-Bfr in terms of protein yield and 24-mer formation.

QTag1
MKETAAAKFERQHMDSP
DLH

QTag2
MKETAAAKFERQHMDSG
GGG

QTag3 MERLQQPTGGG

QTag4 MERLQQPTGGGG

SortTag1 MLVPRGGGGG

— **Tag** — **Linker**
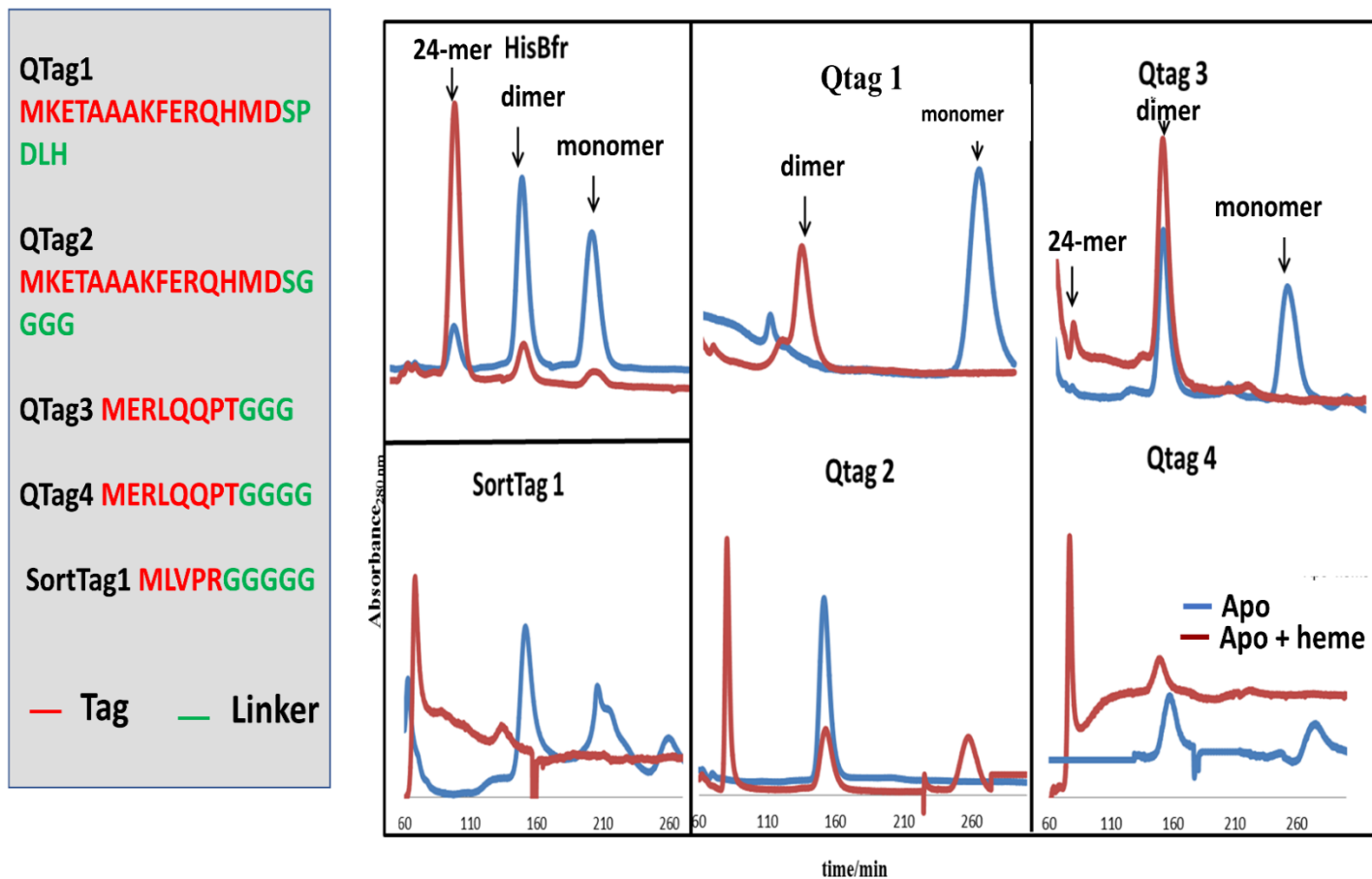
Figure 3.3.7. Size exclusion profile of Bfr heme binding and 24-mer formation. Sephacryl ™ S-300 HR (Hi Prep 26/60) column with a 1.3 mL/min flow rate was used to separate the various quaternary structures. The blue curves represent protein solutions without heme and the red curves are the heme reconstituted samples.

99
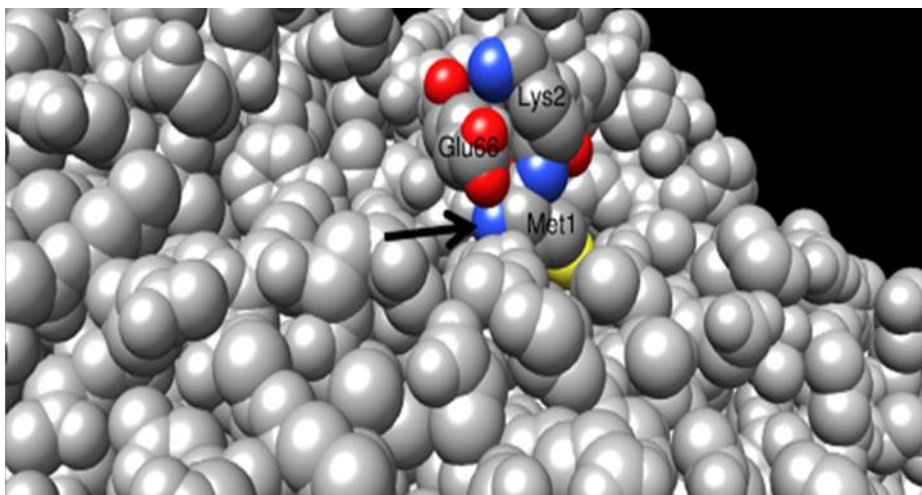
Figure 3.3.8. Space-filling model of bacterioferritin (PDB:1Bfr). Showing the N-terminal Met (arrowed) adjacent to Lys 2 and Glu 66.

* generated by proteolysis of N-terminal precursor sequence
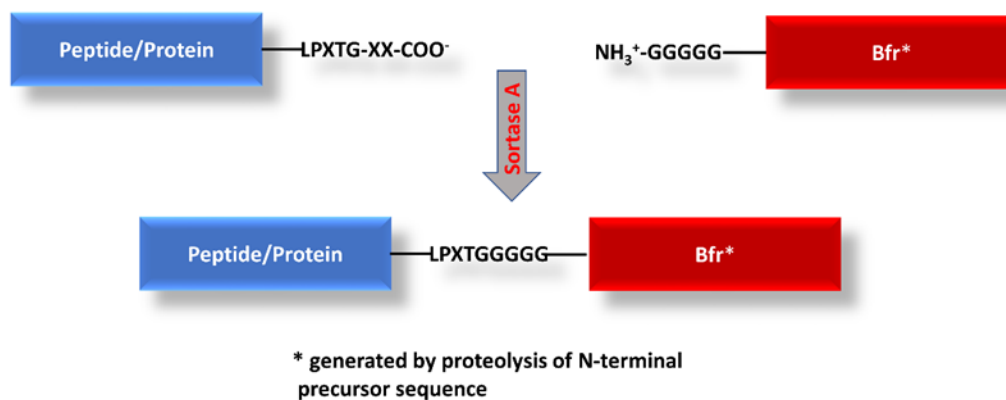
Figure 3.3.8.5. Sortase mediated condensation reaction. The sortTag1-Bfr having sequence MLVPRGGGGG serves as the precursor which, upon incubation with thrombin, provides the N-terminal sequence GGGGG shown above.

3.4. N-Terminal Site-directed Mutations

Two amino acid residues near the N-terminal methionine in Bfr were briefly investigated to attempt to understand further the factors that stabilize/destabilize the Bfr 24-mer quaternary structure. Two bulky residues that are close to the N-terminus were mutated via PCR based site-directed mutagenesis to form three new Qtag2-Bfr constructs: Qtag2E86A, Qtag2K22A, and Qtag2E86A-K22 keeping the C-terminal His-tag for purification and future modifications. Similarly, the expression and purification produced sufficiently pure (without any conterminant bands on SDS gel) proteins with consistent yields of approximately 25 mg/L for Bfr (Appendix 3). Heme binding was performed as described (at 70 °C) and 24-mer capsules were separated utilizing size exclusion chromatography as previously mentioned. The mutants were dimeric without the presence of exogenous heme (Figure 3.3.9 a). Analysis of the Qtag2-Bfr plot appears to indicate the presence of a 35:60 % monomer: dimer ratio under these conditions unlike Figure 3.3.7 above, where the solution lacking heme was 100 % dimeric in the absence of the exogenous heme cofactor. However, it was expected that the addition of heme should shift the equilibrium to the 24-mers form, as seen for Qtag2-Bfr. The heme binding results indicated dimer formation, as seen in Figure 3.3.9 b for both Qtag2K22A and Qtag2E86A-K22A mutants. Only the Qtag2E86A mutant showed a slight shift towards the formation of the 24-mer, suggesting that lysine 22 is useful for the electrostatic interactions between adjacent subunits and the resultant hydrophobic collapse of the subunits into a 24-mer cage. Also, the persistence of the dimer could indicate that the residues that are reported to stabilize the heme cofactor (Phe 26 and Met 52[63a]) are not affected by this surface modification. Analysis of this data suggests that the mutants are very stable to purification but may not be suitable for manipulation that requires intact 24-mer structures for any applications. Hence the original Qtag2-Bfr was used for further studies as described below.
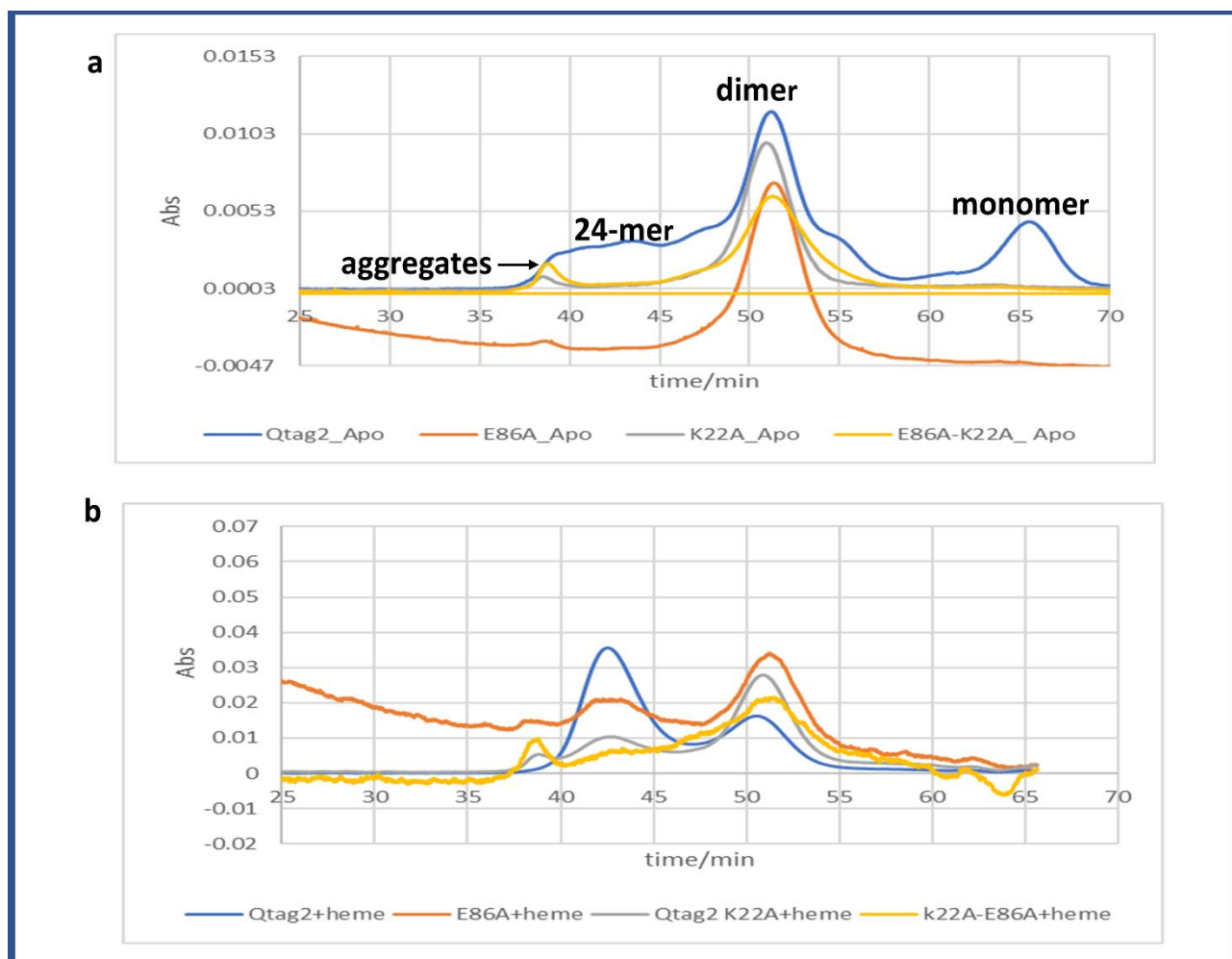
Figure 3.3.9. Size exclusion profile (Sephacryl ™S-300 10/300 HR) of Qtag2 and its variants. (a) the no heme (apo) forms after purification with absorbance at 280 nm on the y-axis also and time on x-axes. (b) separation of the 24-mer protein after heme binding.

3.4.1. Encapsulation of AuNP into Qtag2-Bfr (Targeted Guest Delivery)

The ability of Bfr to serve as a targeted cargo delivery system was investigated by encapsulating QTag2-Bfr with 5 nm gold nanoparticles (AuNP) and labeling the exterior surface with dansylcadaverine to yield the Qtag2-Bfr-DC nanoparticle, decorated with the dansyl fluorophore. This would be an approach to provide a proof-of-concept that a guest-loaded Bfr can be modified on its surface to attach molecules, which in future could be molecules such as antibodies, to help target these ferritins to cells or surfaces. The AuNP was encapsulated by clustering the Bfr subunits around the NTA-functionalized nanoparticle. The Bfr-C-terminal H_6 tags were coordinated by NTA in the presence of Ni^{2+} ions. A typical coordination complex is composed of a quadridentate NTA, and the two remaining valences on the Ni^{2+} are occupied by two histidine molecules from the H_6 tags as seen in Figure 3.4a below. Moreover, hemin solution was added to the reaction to lock the Bfr subunits into their 24-mer structure. Since the Bfr solution is a mixture of monomers, dimers, and 24-mers, the buffer was supplemented with 8 M GdnHCl to decluster all the higher oligomers and expose the internal H_6 tag to the NTA-guest. Hence the 8 M buffer system was used in all HisBfr-NTA mediated encapsulations. The 8 M GdnHCl did not affect any of the reaction components, and it was removed by dialysis after the encapsulation. Lastly, the encapsulated AuNP was separated from excess reagents (Bfr subunits, free AuNP) and other oligomers by size exclusion chromatography by monitoring the AuNP 518 nm absorption. Figure 3.4b below shows the encapsulated AuNP and free AuNP eluting at 43 min and 47 min, respectively. Also, Bfr monomers and dimers were observed without any 518 nm absorption peaks. The shift in elution time corresponding to the retention of the 24-mer structure indicated that the AuNP was encapsulated within the Bfr interior cavity. The AuNP was successfully encapsulated despite its observed 8 nm size on DLS (data not shown). The extended size is due to the proprietary linker and the NTA (corona) on the surface of the 5 nm AuNP. The encapsulation was facilitated by the tendency of HisBfr to associate around the guest molecule. Hence it is not clear from this analysis the number of Bfr subunits which surround the AuNP, and further experiments are required to validate this.

The Bfr-AuNP complex was further modified on its surface with transglutaminase-catalyzed covalent attachment of dansylcadaverine (DC) molecules. First, the complex was incubated with the DC substrate in the presence of 5 mM NaCl and microbial transglutaminase. The bioconjugation led to the covalent attachment of DC to the Bfr-AuNP surface. The labeling

of Bfr with DC was verified by employing fluorescence spectroscopy and SDS-PAGE to detect that the Bfr subunits were labeled by the fluorophore (Figure 3.4 d). Transmission electron microscopic (TEM) analysis was used to visualize the encapsulated AuNP (Figure 3.4 c). Analysis of the TEM image indicates AuNP encapsulation with a few empty 24-mer shells and free AuNP aggregates. All the encapsulated products were 1-2 nm larger than the empty Bfr 24-mer capsids. The results here show that Bfr could be repurposed for targeted cargo delivery where the cargo could be a drug or a magnetic resonance imaging (MRI) agent, and the exterior could be modified with cell recognition peptides or antibodies.
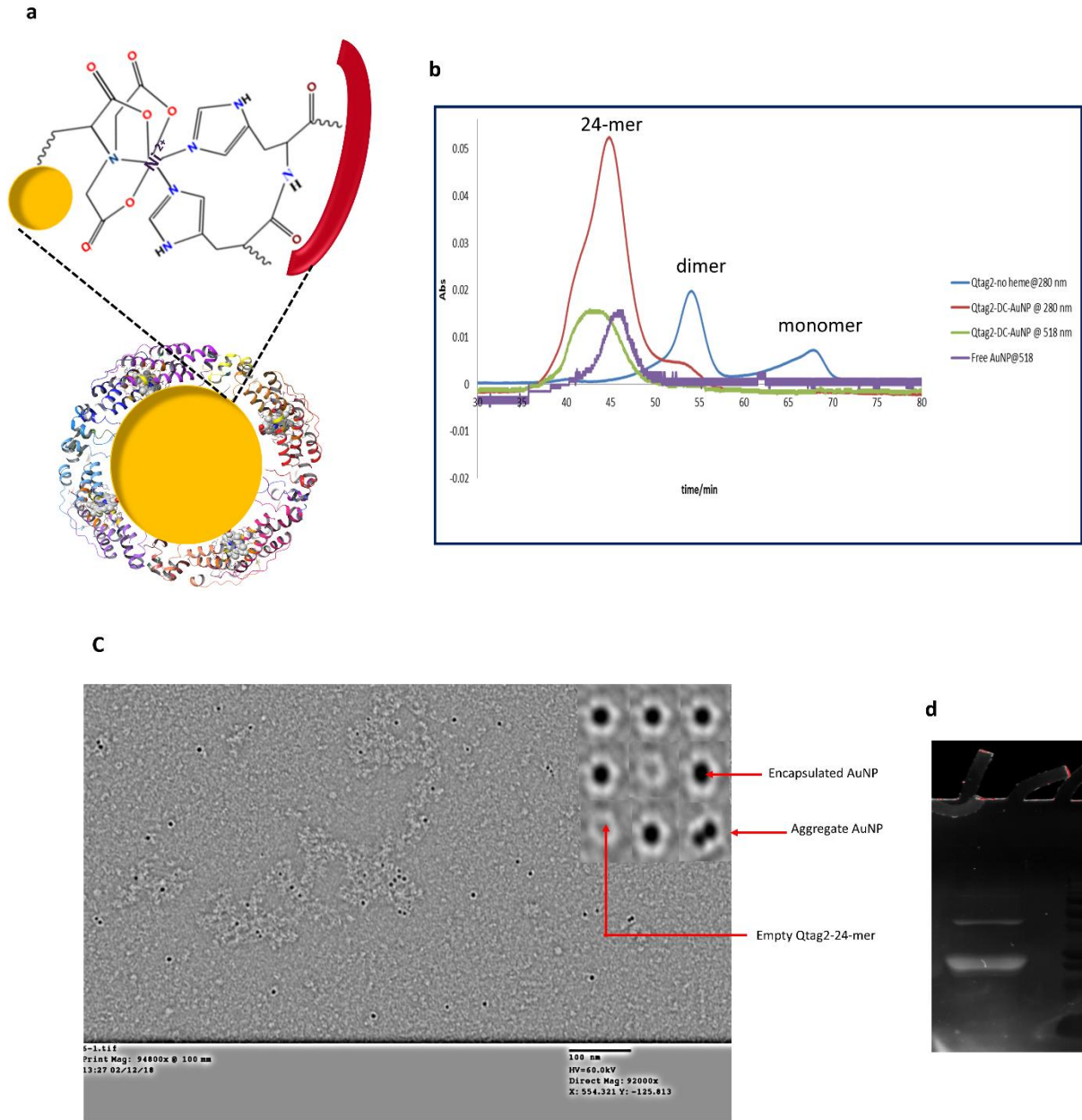
Figure 3.4. External and internal surface modification of Qtag2-Bfr. Panel a shows a schematic diagram of the His-Ni^{2+}-NTA interaction that stabilizes the AuNP. (b) SEC elution profile with the encapsulated AuNP (purple) corresponding to an intact 24-mer structure in size. (c) The TEM image of the pooled fractions corresponding to the 24-mer structure from SEC, indicating successful AuNP encapsulation (d) The SDS-PAGE gel illuminated to show the fluorescence of DC-labeled Qtag2-DC-AuNP, indicating that the AuNP-Bfr has been dansylated on the exterior surface.

3.4.2. Bacterioferritin RNase S Reconstitution

The second role of the Qtag1 and Qtag 2 Bfr construct is demonstrated by their tendencies to reconstitute RNase S activity. The Qtag1-Bfr contains the S-peptide of RNases S. The S-peptide is obtained from the subtilisin digest of RNase A followed by separation of the RNase S-peptide (residues 1-20) and S-protein (residues 21-124)[146] from the unreacted RNase A as seen in Figure 3.5 below. The S-peptide is employed as a fusion tag for protein purification on resins having S-protein attached to them. Also, the reassociation of the S-peptide and S-protein produces a fully functional RNase S enzyme complex with dissociation of 5×10^{-9} M or less (in the presence of 5×10^{-7} M S-protein).[146] Residues 1-15 have been reported to be necessary for the RNase S activity; hence, noncanonical peptide sequences can be used to reconstitute semisynthetic RNase S with high catalytic activity.[106]
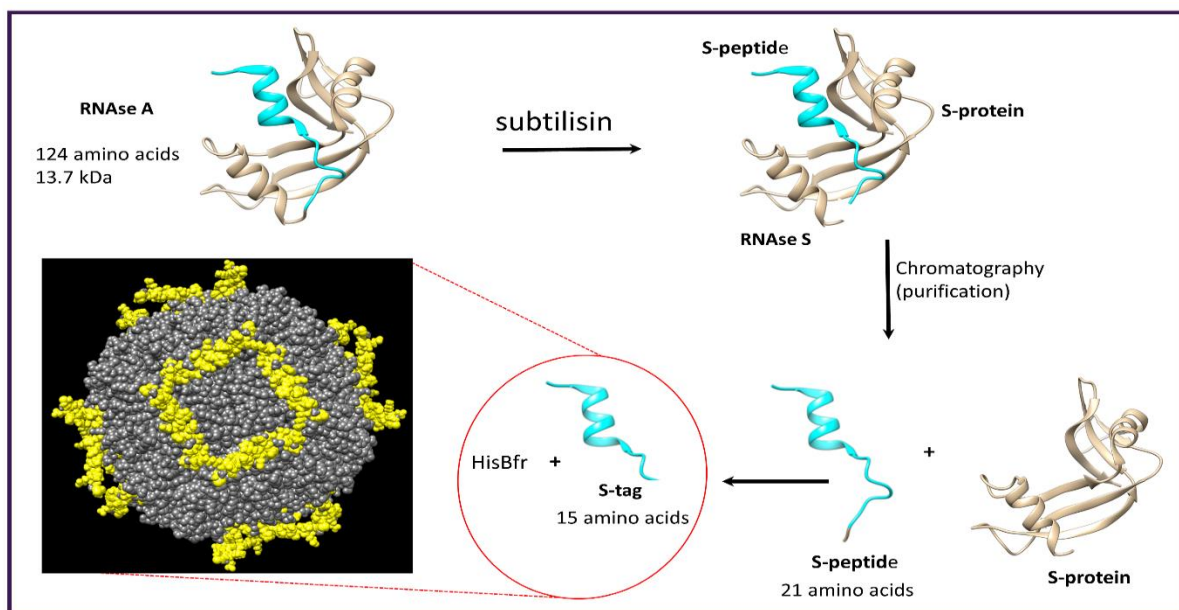
Figure 3.5. Representation of Qtag1-Bfr design. RNase A (2RNS) is cleaved into RNase S, which is purified to separate the S-peptide from the S-protein. The Qtag1-Bfr construct was designed with a plasmid coding for an S-tag peptide fused to the N-terminus of the HisBfr protein to produce Qtag1-Bfr (surface representation: HisBfr in gray and Qtag1 in yellow) which was later engineered to Qtag2-Bfr.

The Bfr surface was modified with the 1-15 oligopeptide (MKETAAAKFERQHMD), from the S-peptide (Figure 3.5). The commercial pET-29b vector, with S-tag, contained an SPDHL linker sequence between the Bfr N-terminal methionine sequence and the S-peptide. The plasmids were successfully transformed into *E. coli* BL21 cells and used to express the S-tag-Bfr. The sequencing results showed a Bfr with N-terminal S-tag and C-terminal H_6 tag, as seen in Figure 2 b and Table 1. The construct was renamed Qtag1-Bfr due to its ability to be utilized as a transglutaminase substrate, making it a double duty construct. Attempts were made (in collaboration with Taylor Urquhart, Honek Lab) to reconstitute RNase S activity from Qtag-Bfr and S protein with the Qtag-Bfr acting as the S-peptide. The S-protein was purified by following the protocol from Genz *et al.*[106]

The RNase assay was carried out with the RNase alert® lab test V2 kit. The RNaseAlert® proprietary substrate has a fluorescent label at one end and a quencher on the other end of a short oligonucleotide. The substrate has a low fluorescence in the absence of RNase due to the proximity of the quencher molecule, but this distance is increased upon RNase hydrolysis, which results in a bright green fluorescent signal upon excitation.

The tendencies of QTag1-Bfr and Qtag2-Bfr to hydrolyze this substrate were studied by incubating these constructs with the purified S-protein to form an expected Bfr-RNase S complex which would likely have RNase activity. The Bfr-RNase S was then incubated with the substrate and the emission spectra of the cleaved substrate recorded. The chromatographic step successfully separated the S-protein from the reaction mixture, peak 2, Figure 3.5.1 b. However, the assay showed residual RNase S activity in the S-protein when it was incubated with the substrate without any S-tag peptide. The observed S-protein activity (Figure 3.5.2 a) indicated that there might be residual uncleaved RNase A in peak 2 (Figure 3.5.1 b), but it has a lower rate constant (estimated 0.033 min^{-1}) than Qtag1-Bfr which was determined to be approximately ten-fold higher (approximately 0.33 min^{-1}) as seen in Figure 3.5.2 b. Conversely, the Qtag2-24-mer + S-protein sample showed low activity indicating that the Bfr surface (24-mer) may be affecting the assembly process and ultimately the hydrolysis of the substrate. Overall, the preliminary data suggest that the subunit Qtag1-Bfr can reconstitute RNase S activity, but the flexible linker region in Qtag2-Bfr may not be suitable for reconstitution of the RNase S activity. However, further studies would be warranted to explore the capability of reconstituting enzyme activity employing extended S-tag sequences on the Bfr surface.
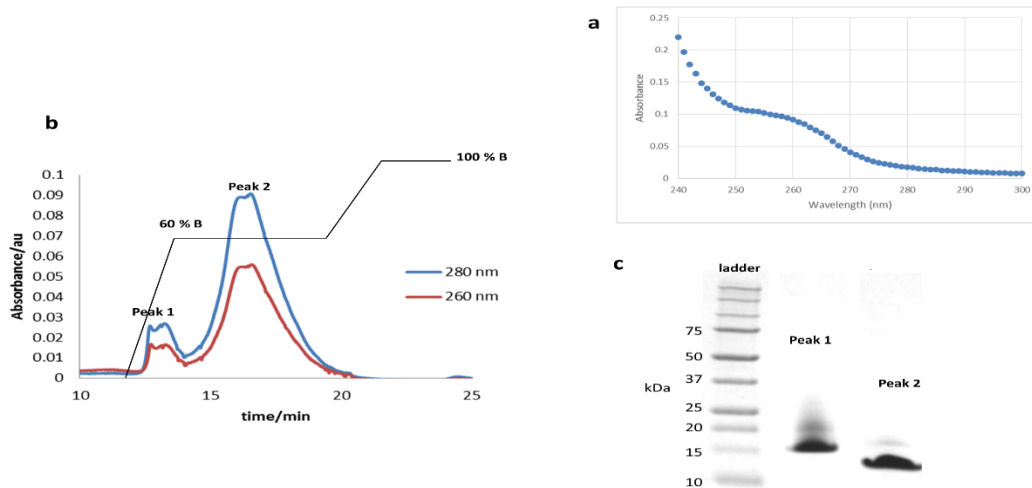
Figure 3.5.1. Purification and quantification of S-protein and S-tag peptide. (a) Absorption spectrum for the commercial S-tag peptide (KETAAAKFERQHMDS). (b) The elution profile from the Macro-Prep® High S 10/100 column showed two peaks; peak 1 corresponds to the RNase A while peak 2 corresponds to RNase S protein, and unresolved RNase A (faint band) as indicated in Figure c.
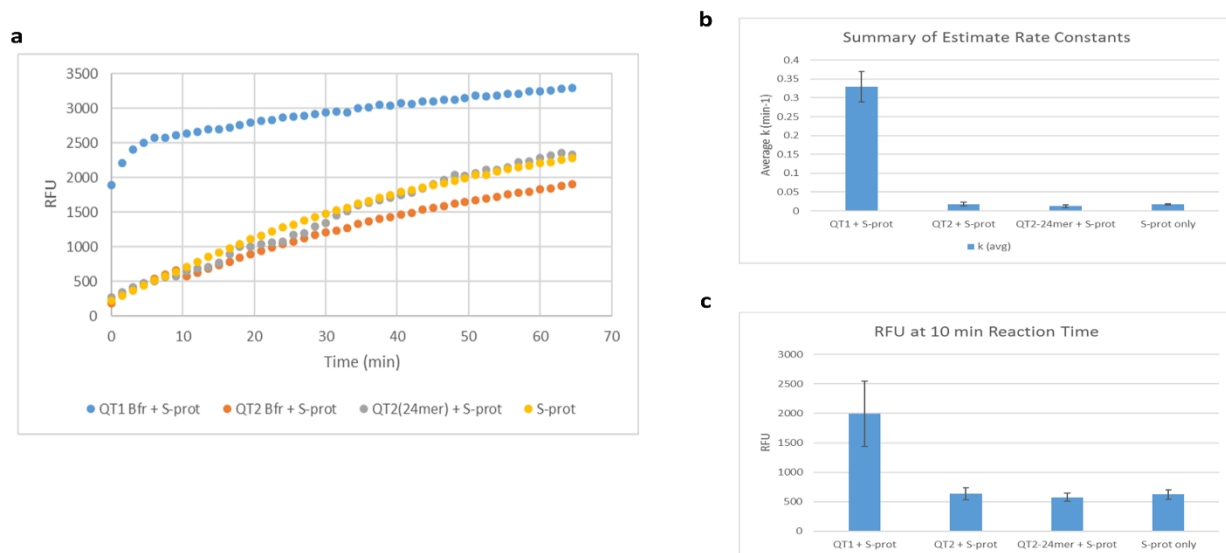
Figure 3.5.2. Preliminary rate of hydrolysis. (a) The fluorescence intensities associated with the RNaseAlert® substrate using Qtag1-Bfr-S-protein (blue), Qtag2-Bfr-S-protein (orange), S-protein (yellow), and 24-mer Qtag2-Bfr-S-protein (gray). The S-protein was added to the substrate to check any background hydrolysis from the reaction components. Figures b and c show the estimated rate constants and the relative intensities of hydrolysis within 10 minutes, respectively.

3.5. Encapsulation of His-tagged GFP (HisGFP)

The HisGFP molecules were encapsulated via a reverse polarity approach wherein the cargo was His-tagged, and the host Bfr contained the NTA functionality. The reverse polarity approach shows that Bfr is very versatile and can be engineered to encapsulate a wide variety of His-tagged proteins and enzymes. The HisGFP variant used here is the enhanced GFP with a single S65T mutation. The Ser65 residue becomes dehydrated to form a vinyl side chain during the post-translational chromophore formation (as seen in Figure 3.6) and the resultant chromophore has longer excitation and emission wavelengths of 490 nm and 510 nm, respectively.[147]
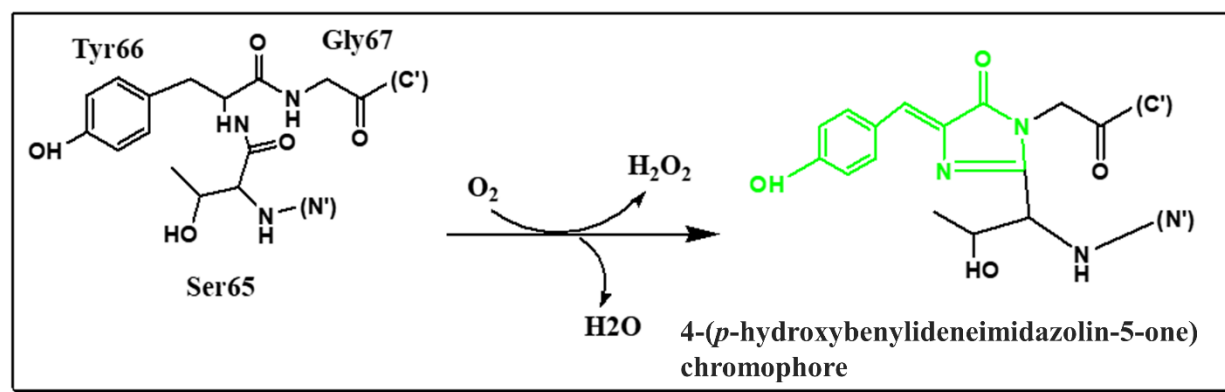
Figure 3.6. Formation of enhanced GFP chromophore. Figure was generated in ChemDraw.

Preliminary HisGFP encapsulation into Bfr was undertaken by Anton van der Ven (M.Sc.) and Sorina Choirean (undergraduate student) from the Honek laboratory. Firstly, the heme cofactor was activated by utilizing EDC/NHS to produce the bis-NHS ester of heme followed by reaction with Nα, Nα-bis(carboxymethyl)-L-lysine to produce the bis-NTA analog (Figure 3.6.1, insert), and it was reinserted into the dimer interface (at 80 °C for 10 min) to form the BisNTA-heme-Bfr(BisNTA-Bfr) complex. The BisNTA-Bfr complex was further incubated (60 °C for 30 min) in the presence of HisGFP molecules to produce a mixture of encapsulated and unencapsulated HisGFP. Secondly, the resultant mixture was separated on a Superose 6™ 10/300 GL column, at 0.5 mL/ min using an eluent composed of 150 mM NaCl, 300 mM KCl, 50 mM Tris-pH 8. The increased ionic strength (addition 300 mM KCl) further assisted in locking the Bfr subunits, initial trials with 150 mM NaCl, 50 mM Tris-pH 8 buffer produced an equilibrium

mixture of 24-mers and subunits in solution. The emergence of a 490 nm absorbance peak eluting at a fraction which also corresponds to the 24-mer capsule (as seen in Figure 3.6.1) was consistent with HisGFP (Sequence in Table 1) encapsulation since the free HisGFP eluted in the region associated with monomer and dimer sized proteins. Also, the presence of the GFP chromophore absorbance showed that the heat treatment at 60 °C mediated HisGFP encapsulation without affecting its stability. Even though both the spectroscopic and nonspectroscopic analysis indicated there could be some level of encapsulation, it was not clear whether the pooled fractions contained 100 % encapsulated HisGFP or a mixture of empty and encapsulated Bfr capsules. Hence transmission electron microscopy was used to visualize the fractions. The encapsulation of the HisGFP did not swell the WTBfr capsule (no increase in hydrodynamic radius) and the likely complex was eluted at the same retention time as the free 24-mer WTBfr.

Also, the addition of increasing amounts of HisGFP in a series of encapsulation experiments did not enhance the 490 nm absorbance associated with the 24-mer Bfr during SEC chromatography but did produce increasing amounts of HisGFP eluting at the monomer/dimer position. The appearance of excess HisGFP was likely due to the inability to encapsulate more than one HisGFP molecule within the WTBfr 24-mer. To show the relative sizes of the host and guest proteins, the crystal structure of a GFP (PDB:1C4F) was visualized and its dimensions determined (4.5 nm long and 3.5 nm wide; Figure 3.6.2). These dimensions would indicate that the size of GFP should be within the interior capabilities of the Bfr (1Bfr) cavity, but the presence of the His-tags and loop regions may increase the overall space the HisGFP occupies. The structure and dynamic nature of HisGFP may prevent the loading of two HisGFP molecules into the Bfr (also having Heme-NTA groups that take up some space). Control studies that were also undertaken by using unmodified heme did not exhibit any HisGFP encapsulation (Appendix 14). The results herein provide evidence that a modified heme cofactor can be used to encapsulate non-native guest molecules.
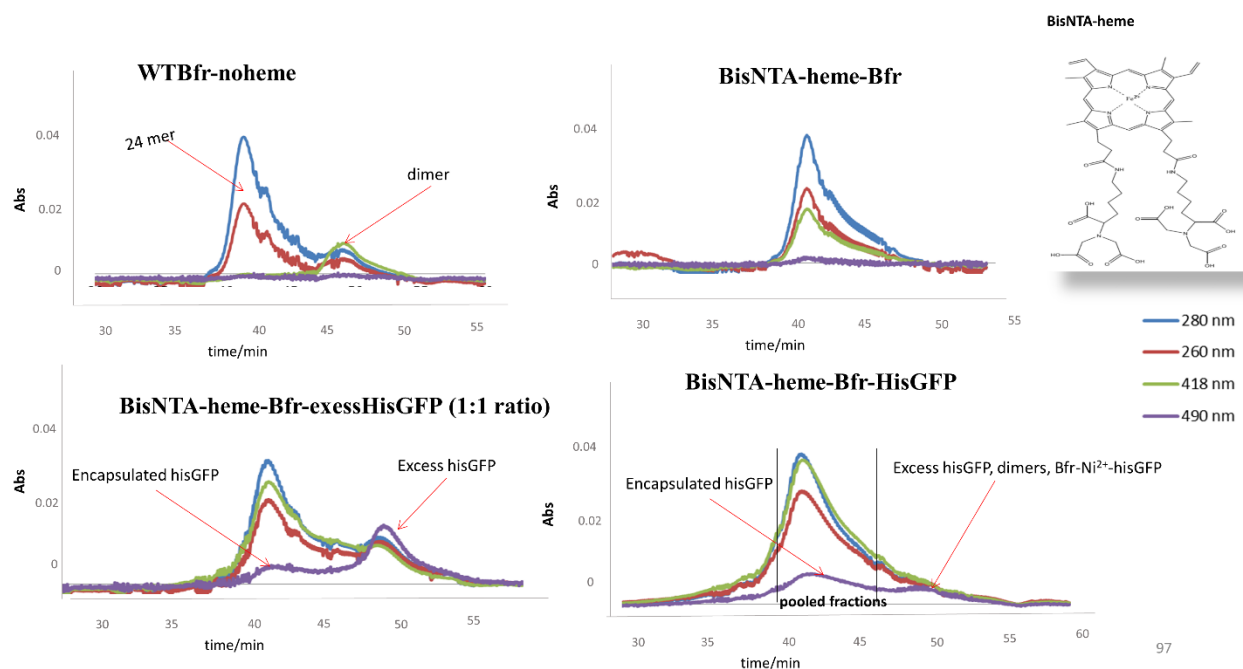
Figure 3.6.1. Size exclusion profiles of HisGFP encapsulation. This figure shows the elution profiles (on Superose 6™ 10/300 GL) of the starting material (WTBfr-noHeme) and its holo form, BisNTA-Bfr as well as the encapsulated product. The insert, top right, is the structure of BisNTA-heme molecule. The two bottom figures show that WTBfr has a limit to the number of HisGFP (490 nm) that can be encapsulated.
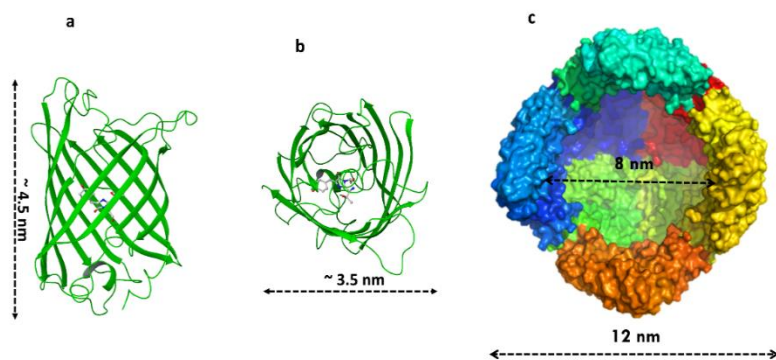
Figure 3.6.2. Dimensions of host and guest proteins. A and B show the length and breadth of enhanced GFP (PDB:1C4F), respectively. C shows the interior and exterior dimensions to Bfr (PDB:1Bfr) to illustrate that the GFP molecule can fit into the Bfr cavity. The dimensions were estimated with the pyMOL measure tool.

111

3.5.1. Characterization of the Encapsulated HisGFP

The pooled fractions from SEC (bottom right, Figure 3.6.1) were analyzed by employing UV-visible spectroscopy, 20 % SDS-PAGE electrophoresis, TEM (Figure 3.6.4 c) and mass spectrometry. The UV-vis scan of the encapsulated HisGFP was compared to the free HisGFP. The spectrum showed a Soret band at 415 nm for the 24-mer bisNTA-heme-Bfr as well as a 490 nm shoulder peak for the WTBfr-HisGFP (Figure 3.6.3a). The SDS-PAGE analysis (Figure 3.6.3 b) was used to verify that there were indeed two different proteins in the pooled fractions, and the 490 nm peak was not associated with the bisNTA-heme molecules. The gel image showed that there were two distinct proteins in the sample, which corresponds to HisGFP (approximately 27 kDa) and WTBfr (approximately 18 kDa). The exact masses of these proteins were obtained from +ESI MS analysis on the encapsulated sample (Figure 3.6.3c). However, the +ESI data showed the intact mass of WTBfr subunit,18494.33 Da, and cleaved HisGFP at 27741.051 Da. The mass loss was estimated to be Ni^{2+}-mediated His-tag cleavage of Ala-Ser-His from the N-terminal and addition of two K^+ ions (fragment ii, Figure 3.6.3e) and further investigated with +ESI-MS analysis. Free HisGFP was incubated with Ni^{2+} and bisNTA-heme (all at encapsulation conditions) for 60 minutes. The resultant mixture was washed with water and analyzed for the intact HisGFP peak. Surprisingly, no intact HisGFP was observed (Figure 3.6.3d). The HisGFP appeared to be cleaved into several species; the most abundant peak was 26473.101 Da (fragment v, Figure 3.6.3e). These fragments were assigned to C-terminal cleavages labeled ii to vi in Figure 3.6.3e, possible due to metal-mediated cleavage.

Finally, transmission electron microscopy was used to compare the WTBfr to the HisGFP encapsulated WTBfr (Figure 3.6.4b and c). The WTBfr without any guest molecules showed stain penetration, through the shell, into the interior cavity (Figure 3.6.4b) whereas the encapsulated HisGFP excluded the stain from the WTBfr interior cavity (Figure 3.6.4c). Figure 3.6.4c showed that there are a few empty WTBfr capsules and about 60-80 % WTBfr capsules filled with the HisGFP molecules, which means that the SEC column could not resolve the two forms of Bfr. The estimated particle size (with ImageJ software[148]) was about 12 nm for both empty and filled WTBfr.
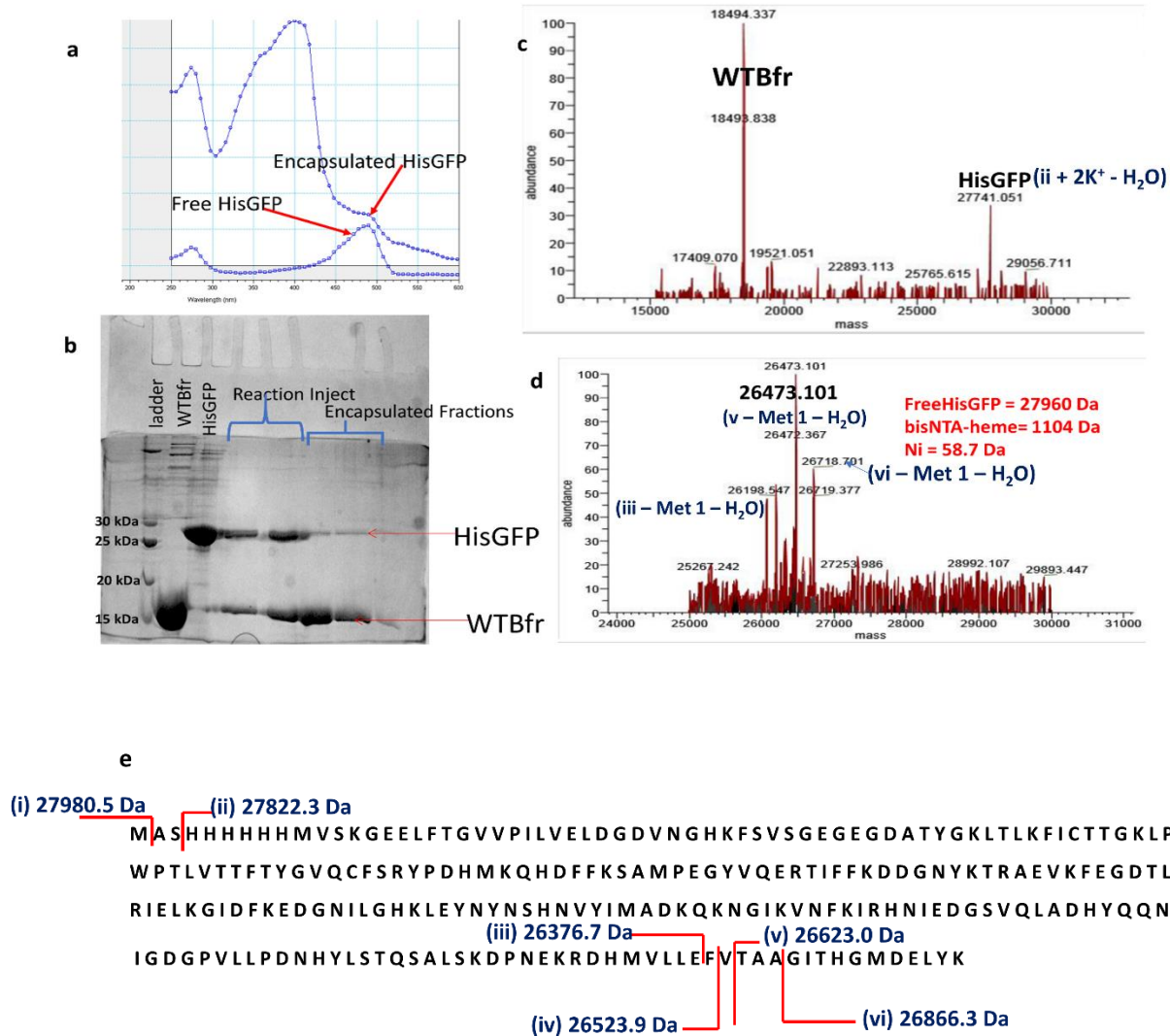
Figure 3.6.3. Analysis of encapsulated HisGFP. (a) the UV-Vis spectrum of the encapsulated HisGFP compared to the free HisGFP, both showed the characteristic 490 nm absorbance. (b) The SDS-PAGE image of the encapsulation showed that the pooled fractions contain both proteins. (c) Positive ESI-MS analysis for the encapsulated HisGFP with peaks corresponding to intact WTBfr and cleaved HisGFP. (d) Mass spectrum of HisGFP under the encapsulation conditions in the presence of bis-NTA-heme but lacking the presence of Bfr. (e) The deduced fragmentation pattern for HisGFP. The mass of purified HisGFP is **27960 Da**, which resulted from fragment (i) and the loss of H_2O during chromophore formation. Additionally, C-terminal fragments (iii and vi) were observed and assigned on Figure d.
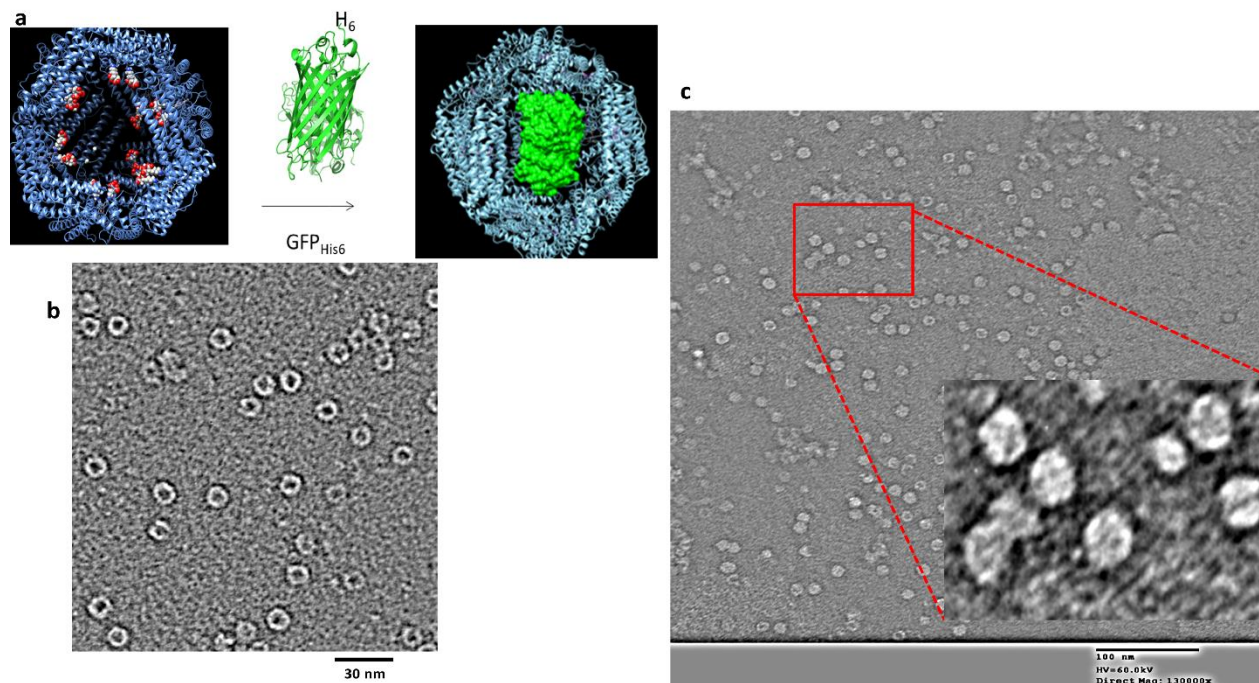
Figure 3.6.4. The TEM analysis of Encapsulated HisGFP. (a) The left image shows how the heme cofactors (in space fill) can be modified and used for encapsulation of HisGFP (green). (b) The TEM image of WTBfr (1Bfr), showing interior 0.5% molybdate stain penetration. (c) Image c indicated stain exclusion in the presence of GFP (1C4F) guest molecules. The protein structures were generated with Chimera

3.5.2 Modification of HisGFP Cysteines

In 1994, Inouye and co-workers reported that the addition of strong reducing agents such as $Na_2S_2O_4$, $FeSO_4$, 2,2'-azido-bis (3-ethylbenzothiazoline-6-sulphonic acid) (ABTS), and many others (at 5 mM, 2 mM, 0.2 mM, respectively) when added to recombinant GFP could extinguish the GFP fluorescence within a few minutes. [149] It has also been reported that DTNB, 5,5-Dithiobis (2-nitrobenzoic acid), could react with and quench GFP fluorescence. HisGFP fluorescence quenching was attempted therefore with DTNB, 5,5- Dithiobis (2-nitrobenzoic acid), also known as Ellman's reagent. Ellman's reagent (DTNB) reacts with free cysteines (under

alkaline conditions) to form colored solutions containing the thionitrobenzoate anion and the thionitrobenzoate-protein adduct (protein-TNB) as seen in Figure 3.6.4.1. The reduced TNB^{2-} absorbs at 412 nm (extinction coefficient of 14,1500 M^{-1} cm^{-1})[150] and this coefficient is used to estimate the amount of free cysteines versus cystines in solution. The HisGFP construct used in this work contains two cysteines (C58 and C80), which remain reduced after purification. Hence incubating with DTNB could produce two protein-TNB adducts which could be analyzed by either absorption spectroscopy or fluorescence emission (monitoring the effect on GFP emission).
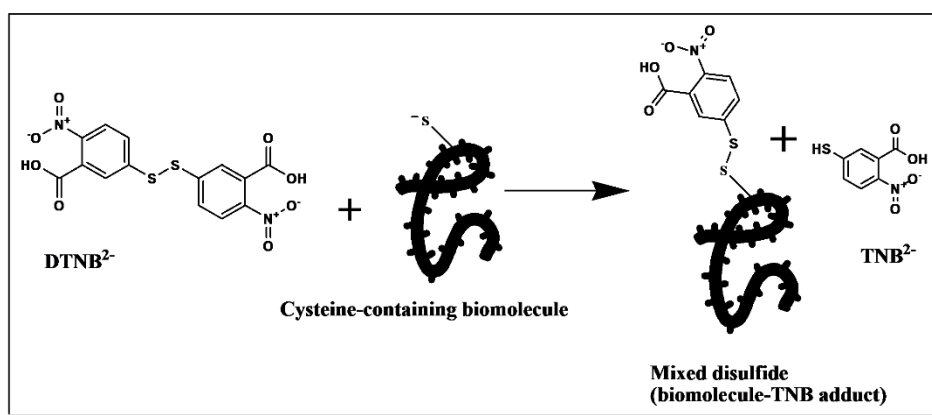
Figure 3.6.4.1. The reaction of Ellman's reagent (DTNB) with the cysteine sulfhydryl group of proteins. Figure generated with ChemDraw.

Firstly, the DTNB was solubilized in water and analyzed employing mass spectrometry, in negative ionization mode, to obtain the mass of 395 Da that corresponds to the M-1 ion (Figure 3.6.4.2a). A reaction was set up by incubating 0.4 µM of HisGFP with 1 mM of DTNB (both in 50 mM Na_2HPO_4, pH 8.0) at room temperature for an hour in the absence of Bfr. A few microliters of the reaction solution was then washed with DDH_2O and analyzed by +ESI-MS. Analysis of the +ESI -MS data in Figure 3.6.4.2c showed that the protein mass of HisGFP had increased by 396.842 Da, corresponding to HisGFP with both Cys residues modified by TNB.

The fluorescence of the HisGFP was analyzed for GFP chromophore emission, as seen in Figure 3.6.4.3 (Excitation = 490 nm, Emission = 525 nm). HisGFP was then reacted with varying amounts of DTNB (1 mM, 2 mM and 3 mM DTNB; blue, red, green, and purple curves respectively in Figure 3.6.4.3a). The sample was incubated for 30 minutes at room temperature before each measurement. The intensity of free HisGFP continued to decrease with increasing

DTNB concentration. The optimal incubation time was estimated from a freshly prepared 0.4 um

HisGFP and 1 mM DTNB mixture and incubated at room temperature for different time points

(Figure3.6.4.3 b). Again, the fluorescence intensity of the GFP continued to decrease with

increasing incubation times, but a significant decrease, 50 % decrease, was observed after 2 hours

of incubation unlike the "few minutes" reported by Inouye *et al*.[149]

Finally, a solution of 0.1 µM HisGFP (500 µL) and 0.1 µM encapsulated HisGFP (500 µL)

were separately titrated with DTNB (100 µM and 200 µM). The free HisGFP showed a decrease

in fluorescence intensity in response to increasing DTNB concentrations and increasing time,

respectively. However, the encapsulated HisGFP did not show a decrease in intensity for the 100

µM DTNB and a slight decrease for the 200 µM DTNB addition. The two solutions behaved

differently due to the protection afforded by the protein capsule. The results suggest that the

encapsulated HisGFP molecules are shielded from the negatively charged DTNB either due to the

physical barrier afforded by the protein shell and its overall negative electrostatic potential. Yeng

et al. have reported this charge-selective permeation previously for ferritins.[151] These researchers

utilized both human H-type ferritin and horse spleen ferritin to show that the permeation of

nitroxide spin probes is largely controlled by the selective porosity of the ferritin shells. They also

reported that other factors like temperature, incubation time, and concentration might have minor

contributions on the permeation.[151] However, the observed decrease in fluorescence intensity

upon addition of 200 µM DTNB also suggests that in addition to the charge selectivity, the

concentration gradient may also play a role in the diffusion of DTNB molecules to the interior.

Calisti *et al*. made a similar observation in their work entitled: "Probing bulky ligand entry in

engineered archaeal ferritins" where they observed DTNB modification of internal cysteines in

fully caged (closed pore) *Archaeoglobus fulgidus* ferritin, using DTNB concentrations of 200 µM

to 700 µM.[130]

Overall, the reverse polarity approach to control encapsulation was successfully used to

encapsulate HisGFP. Even though the +ESI-MS data of encapsulated HisGFP (Figure 3.6.3 c)

showed the presence of N-terminall-cleaved HisGFP the cleavage did not affect the process of

posttranslational chromophore formation in the HisGFP molecule or its binding via the His-tag

that had one fewer His hence the encapsulated HisGFP fluoresced. The preliminary quenching

experiments showed a level of chromophore protection by the WTBfr shell (up to a particular

concentration). To define the exact factors that diminished the protection, more data points

(varying concentrations of DTNB) could be collected as well as separate experiments could be set up to study the effects of temperature and incubation time on DTNB permeation into the WTBfr shell.
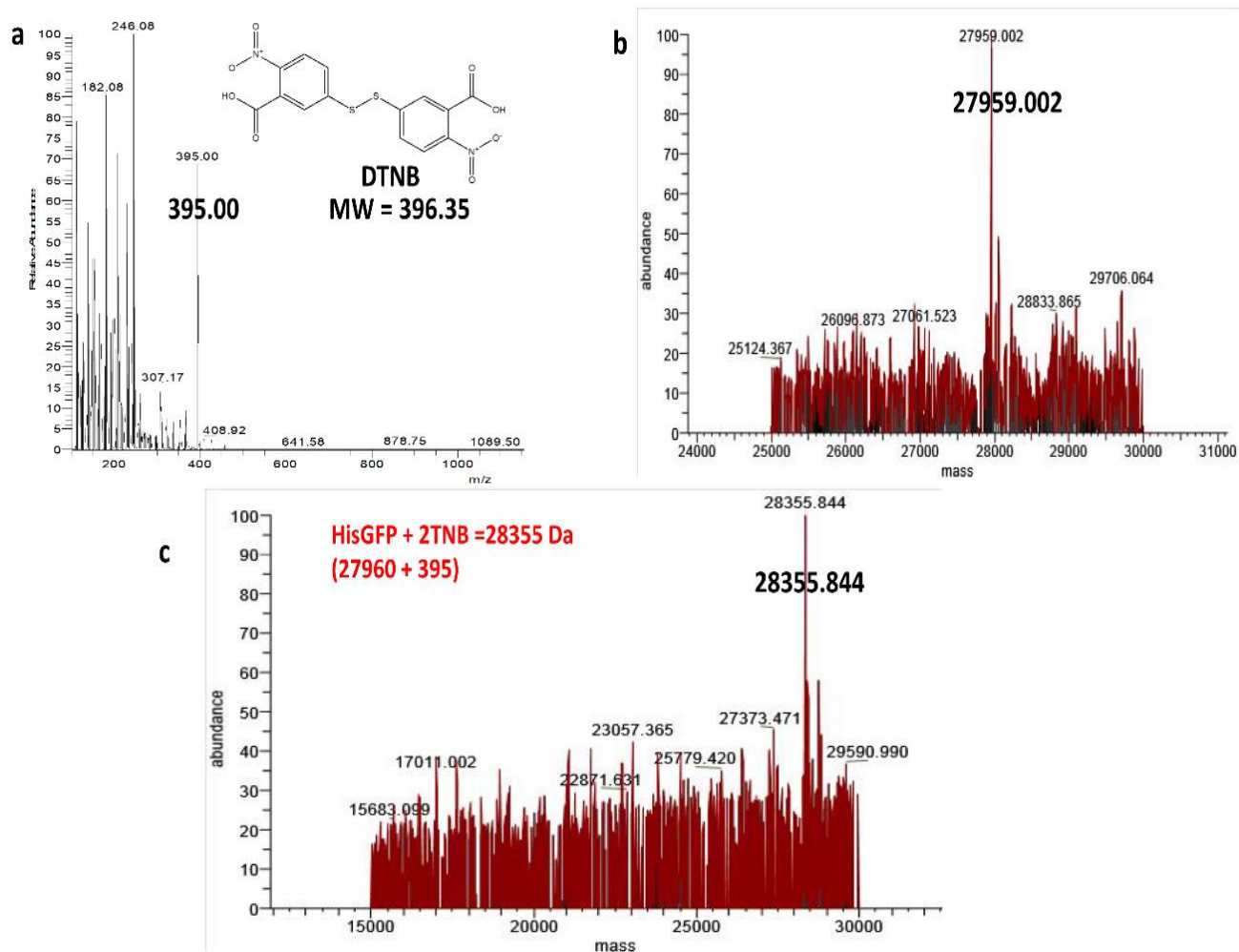
Figure 3.6.4.2. HisGFP-DTNB reaction. (a) -ESI-MS data of DTNB reagent (b) +ESI-MS spectrum of free HisGFP. (c) +ESI-MS data of 0.4 µM free HisGFP + 1 mM DTNB
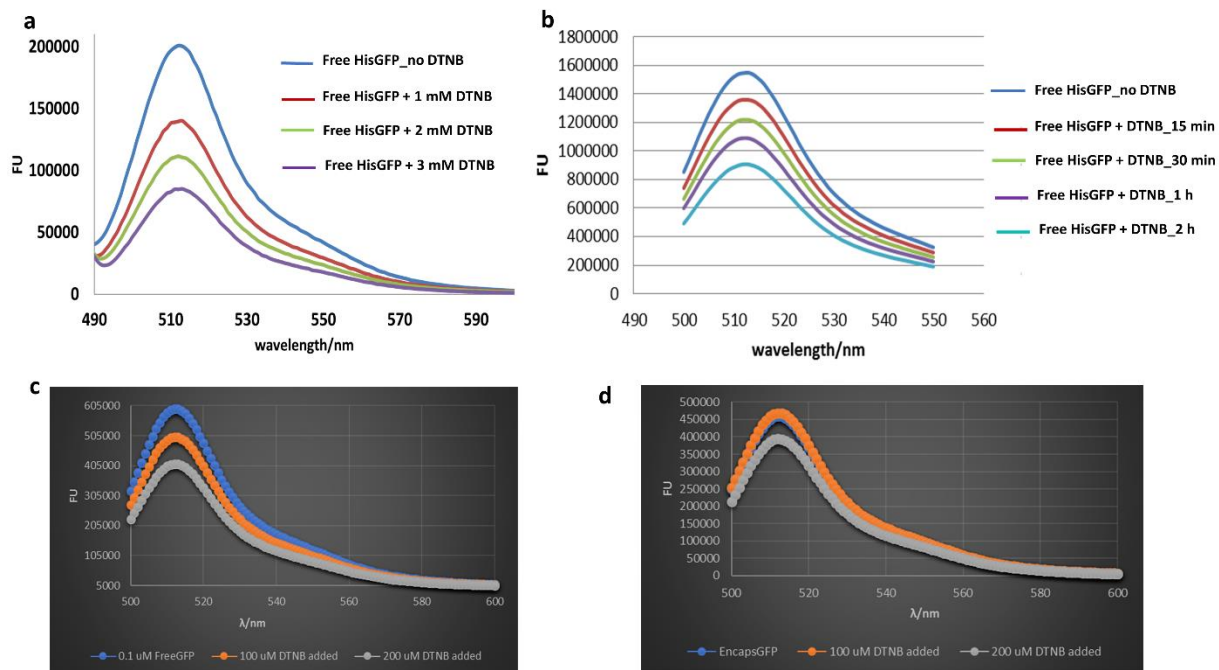
Figure 3.6.4.3. DTNB-mediated HisGFP fluorescence quenching. (a) A solution of 0.4 µM free HisGFP titrated with varying concentrations of DTNB (and incubated for 30 min) (b) Free HisGFP (0.4 µM) incubated with DTNB (1 mM) at room temperature for different time points. (c) Free HisGFP (0.1 µM) separately incubated with 100 uM and 200 µM DTNB for 30 minutes at room temperature. (d) WTBfr-HisGFP (0.1 µM) separately incubated with 100 µM and 200 µM DTNB for 30 minutes at room temperature. All readings were in triplicate using excitation and emission of 490 nm and 525 nm, respectively.

3.6. Hydrophobic Engineering of Bfr

The process of engineering a hydrophobic core within bacterioferritin capsule (Bfr with lipophilic interior surface) could be useful in transporting hydrophobic drugs and perhaps provide new insights into an entrapped hydrophobic environment that is locked within the dimensions of a cavity. Another advantage of hydrophobic engineering is the tendency of the composite (example, Bfr-hydrophobic polymer) to load more hydrophobic drugs and retain them better than their porous unmodified counterpart.[152] However, the slice through the Bfr shell shows that it is hydrophilic on the interior surface, just like the exterior surface, as seen in Figure 3.6.5. So how could one engineer a hydrophobic interior given this water-filled interior? One way is to turn all the internal amino acids hydrophobic, which could destabilize the quaternary structure of the capsule. An alternative route, which has been employed on viral capsid (not proteins), is derivatizing the subunits with hydrophobic molecules and causing them to assemble to create the hydrophobic core.[153,154] For example, Kwak *et al*. employed Cowpea Chlorotic Mottle Virus (CCMV) as a model system for loading virus nanocarriers where they used DNA amphiphiles. The DNA amphiphiles aggregated into micelles with a negatively charged DNA corona (that forms viral capsid) and hydrophobic core.[154] Even though this approach is indicated as a general method and could be expanded to other assemblies, it is of limited value since it does not control the complete cavity, just the internal surface of the cavity.

Here, we present a novel approach to hydrophobic engineering of capsule protein by making a micelle and trying to entrap this using our already developed His tag-NTA strategy. The Bfr capsule was previously engineered by Omaima Ben Krayem in the Honek laboratory to encapsulate lipid aggregates such as preformed micelles. Two types of lipids with NTA functionalities were employed to explore the possibility that a hydrophobic environment could be engineered into the cavity of an assembled capsule protein. The first lipid, synthesized in the Honek laboratory, was nonodecanoyl-NTA (C19-NTA) and the second one, obtained commercially, was 1, 2-dioleoyl-*sn*-glycero-3- [(N-(5-amino-1 carboxypentyl) iminodiaceticacid) succinyl] (DGS-NTA) (Figure 3.6.5.1).[121] Hence the established NTA-Ni^{2+}-H_6 interaction was employed. The HisBfr construct contains C-terminal hexahistidine tags that are exposed to the interior cavity. The His-tag exposure allows HisBfr to interact with the NTA functionalities of the lipids, in the presence of Ni^{2+} and a declustering agent, as seen in Figure 3.7.
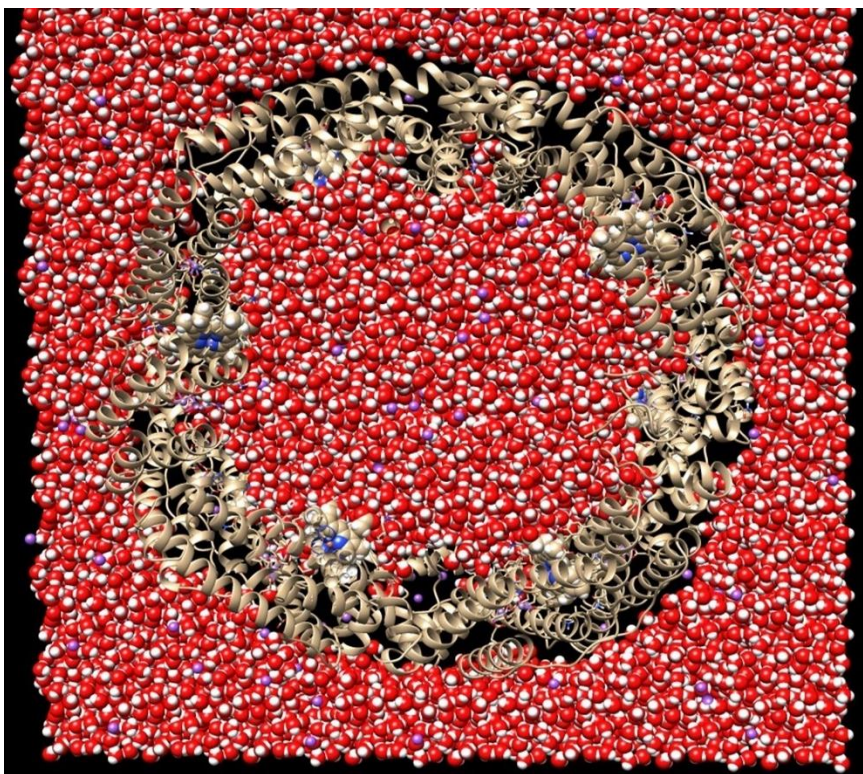
Figure 3.6.5. A static representation of the water-filled interior of Bfr. This image was generated by first preparing and optimizing the 1Bfr structure in the Maestro Prep Wizard, using default settings. The Desmond subprogram for aqueous solvation was then employed, which solvates and minimizes water molecules both within the protein cavity and also on the outside surroundings of Bfr.
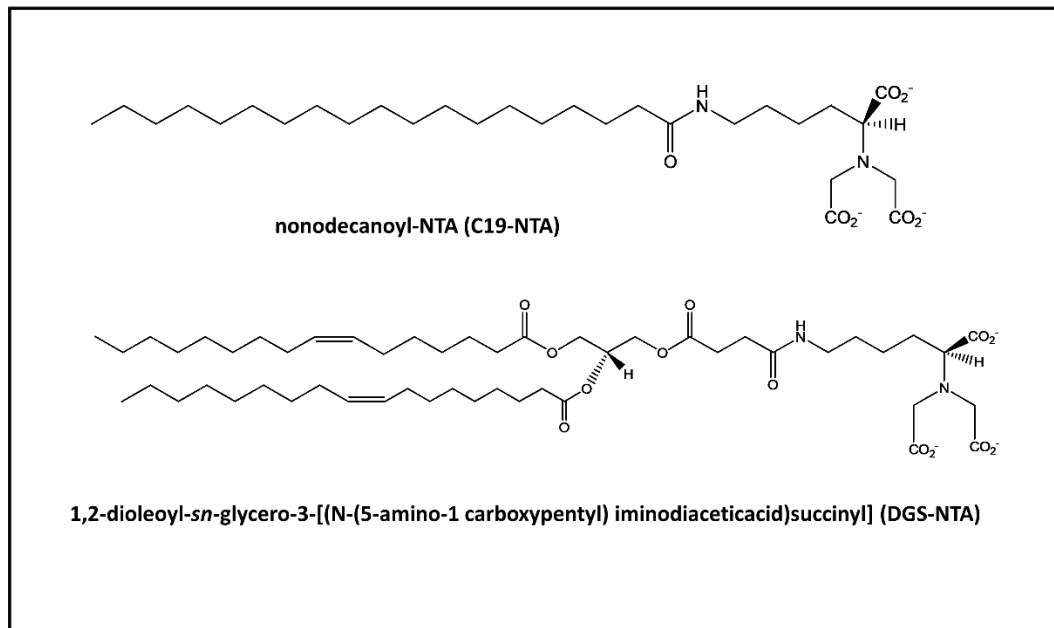
Figure 3.6.5.1. Structures of NTA-lipids used by Omaima Ben Krayem. The C19-NTA was synthesized in the Honek laboratory (by Omaima Ben Krayem) while the DSG-NTA was obtained from Avanti® Polar Lipids.

The results from these encapsulations were analyzed employing size exclusion chromatography, mass spectrometry, and transmission electron microscopy, similar to the approaches used previously in the characterization of guest molecule incorporation. Analyses of the SEC, TEM, and mass spectrometry results for the C19-NTA lipid showed that the Bfr capsule contained lipid, but the capsule had a larger diameter than the native 24-mer. The C19-NTA encapsulation produced capsules that were approximately 21 nm in diameter, almost twice the size of Bfr while the DGS-NTA was closer to the "normal" diameter of Bfr, 14 nm wide (Figures 3.7.1 a and 3.7.2a) (Omaima Ben Krayem). The ability to entrap lipid molecules in the form of micelles or aggregates within the dimensions of Bfr could be useful in potential applications in hydrophobic transport similar to liposomes, lipid emulsions and drug-polymer conjugates with added size homogeneity and enhanced permeability and retention (EPR) effect. Even though the initial encapsulations were successful, it was necessary to explore their reproducibility and

eventually to be able to control the dimensions of Bfr for future practical applications in drug delivery.

In order to determine the factors involved in controlling the final diameter of the Bfr-lipid complex, various experiments were conducted which included variation in the concentration of the lipids, their composition, and the amount of His-tagged: wild type Bfr subunits that were used. The proximity of the hydrophobic fatty acid chains could result in micelles or hydrophobic pockets within the interior of Bfr. This approach stems from previous encapsulations of DGS-NTA at surfactant levels (below the Critical Micelle Concentration, CMC) (by Omaima Ben Krayem), due to its ability to form lipid bilayers in solution, and this guest resulted in 14 nm diameter Bfr capsules being formed. The 2 mg/ mL lipid, employed previously by the Honek group, was decreased to 1 mg/ mL and 0.5 mg/ mL for two separate encapsulations to observe concentration effects on the Bfr diameter. The resultant mixtures were analyzed on SEC and overlaid with the 2 mg/mL sample to compare their retention/elution times. Both of the 0.5 mg/mL and 1 mg/mL samples eluted within the expected 24-mer Bfr retention time, unlike the 2 mg/mL sample, which eluted in the void region. The data indicated that the 2 mg/ mL preparation was broader than a 24-mer Bfr, as seen in Figure 3.7.1 a. Positive ESI-MS further analyzed the composition of the SEC fractions. Analysis of the MS data (Figure 3.7.1 b) showed a mass for the HisBfr subunit as well as the HisBfr-Ni^{2+} - C19-NTA for both 0.5 mg/mL and 1 mg/mL samples (Appendix 4) indicating that there are HisBfr subunits with lipids attached to them and the micelles (or aggregates of lipids) and their location were confirmed by TEM experiments. The TEM image in Figure 3.7.2 c revealed that a fraction of Bfr interior excluded the molybdate stain unlike the 2 mg/ mL encapsulation (Figure 3.7.2 a) which filled the entire Bfr capsule and prevented interior molybdate staining. However, the dimensions of the Bfr capsules were about 12 nm for both 0.5 mg/mL and 1 mg/mL (the 1 mg data is not shown). Thus, the subunit-mediated micelle encapsulation (encapsulation at surfactant concentration) may retain the architecture, but may not lead to as highly a hydrophobically dense Bfr capsule.

Another approach to forming an internally hydrophobic Bfr without expanding the capsule diameter was investigated. The lipids were encapsulated at micellar/ aggregation concentration similar to the previous conditions, but the composition of the Bfr cluster was changed. The Bfr subunits that were employed in this additional approach were composed of a mixture of 50 % HisBfr and 50 % WT Bfr (50:50 Bfr) subunits to decrease the sterics associated with the

employing 100% His-tagged subunits. The assumption here was that the declustered Bfr subunits would pack around the preformed micelle into a 24-mer (or some oligomeric) cluster, but not all the 24 H_6 tags would be required to stabilize the entrapped micelle, resulting in additional space being available in the cavity to accommodate the entrapped micelle.

Alternatively, shorter chain lipids might be encapsulated into the normal sized Bfr even if 100 % HisBfr subunits were employed. These chains were expected to form smaller diameter micelles, which might be better accommodated within a normal sized Bfr interior without a significant increase in cage dimensions. Three experiments were therefore undertaken: (1) C19-NTA with 50:50 mixed subunits (WTBfr: HisBfr); (2) C16-NTA employing 100 % HisBfr subunits; and (3) C14-NTA employing 100 % HisBfr subunits. All three reactions were undertaken with the steps in Figure 3.7. The MS data and SEC profiles (Appendix 5) were consistent with successful lipid encapsulation. The encapsulated samples were visualized using TEM with 1 % molybdate stain (Figure 3.7.3a-c). The size distribution of the Bfr-lipid complexes for the C19-NTA in the 50:50 subunit mixture (Bfr: HisBfr) was within the dimensions expected for "normal" Bfr capsules (10-13 nm, Figure 3.7.3a). The size analysis showed that the His-tags take up internal space, and not all 24 His-tagged subunits are required for successful lipid guest stabilization. Also, the C14-NTA and C16-NTA encapsulation experiments employing 100 % HisBfr subunits produced Bfr capsules without any significant increase in size. However, their sizes were slightly larger than the 50:50 mixture (WTBfr: HisBfr) capsules produced utilizing C19-NTA lipids which are in agreement with the presence of the 24 His-tags that are not compensated for by the shorter lipid chains (smaller diameter micelles for C14 and C16).

Finally, based on the level of stain penetration/exclusion, it would appear that the C19-NTA (employing the mixed subunit approach) produced the most hydrophobic (filled) 12 nm Bfr followed by C16-NTA and lastly the C-14 NTA. The results herein highlight the ability to control the exact hydrophobic character of the internal cavity of Bfr by employing a combination of the exact choice of lipid-NTA as well as the number of HisBfr subunits utilized for the encapsulation.
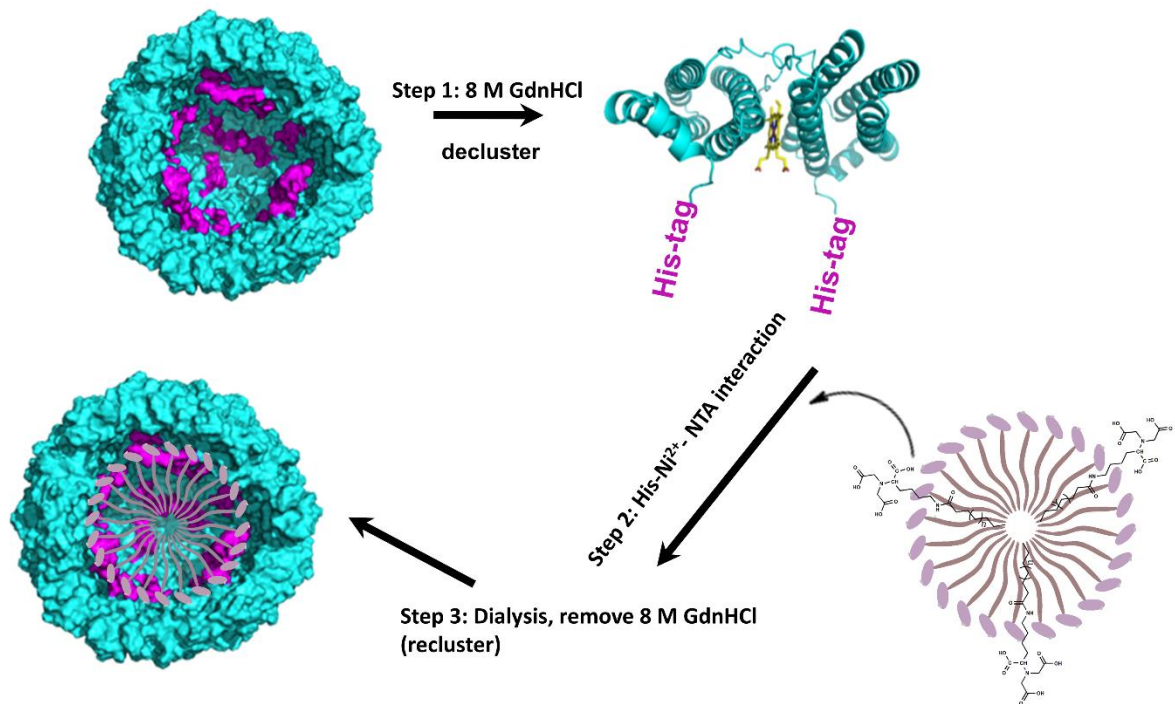
Figure 3.7. Generalized lipid-NTA encapsulation steps. The HisBfr capsule is dissociated to expose the His-tags (in purple) which then interact with the NTA moieties on the lipid in the presence of nickel (II).
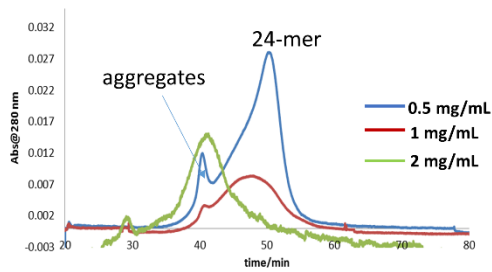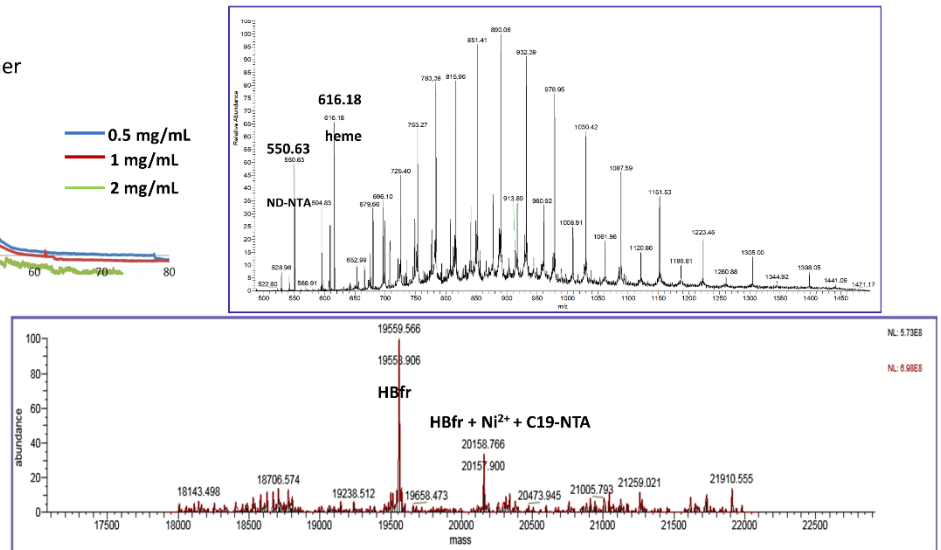
Figure 3.7.1. Separation and characterization of C19-NTA lipid encapsulations. (a) Size exclusion profile of the three lipids with 2 mg/mL (green) sample eluting in the void/aggregate region and the 1 mg/mL (red) and 0.5 mg/mL (blue) at the 24-mer retention time. (b) The mass spectrometry analysis on the fractions at the 24-mer region (Appendix 2) for the 1 mg/mL sample shows ions corresponding to C19-NTA, heme, and protein in the source spectrum (top) and the deconvoluted (bottom) spectrum providing the masses of the HisBfr subunit (1959 Da) and the HisBfr-Ni^{2+}- C19-NTA (20158 Da)complex.
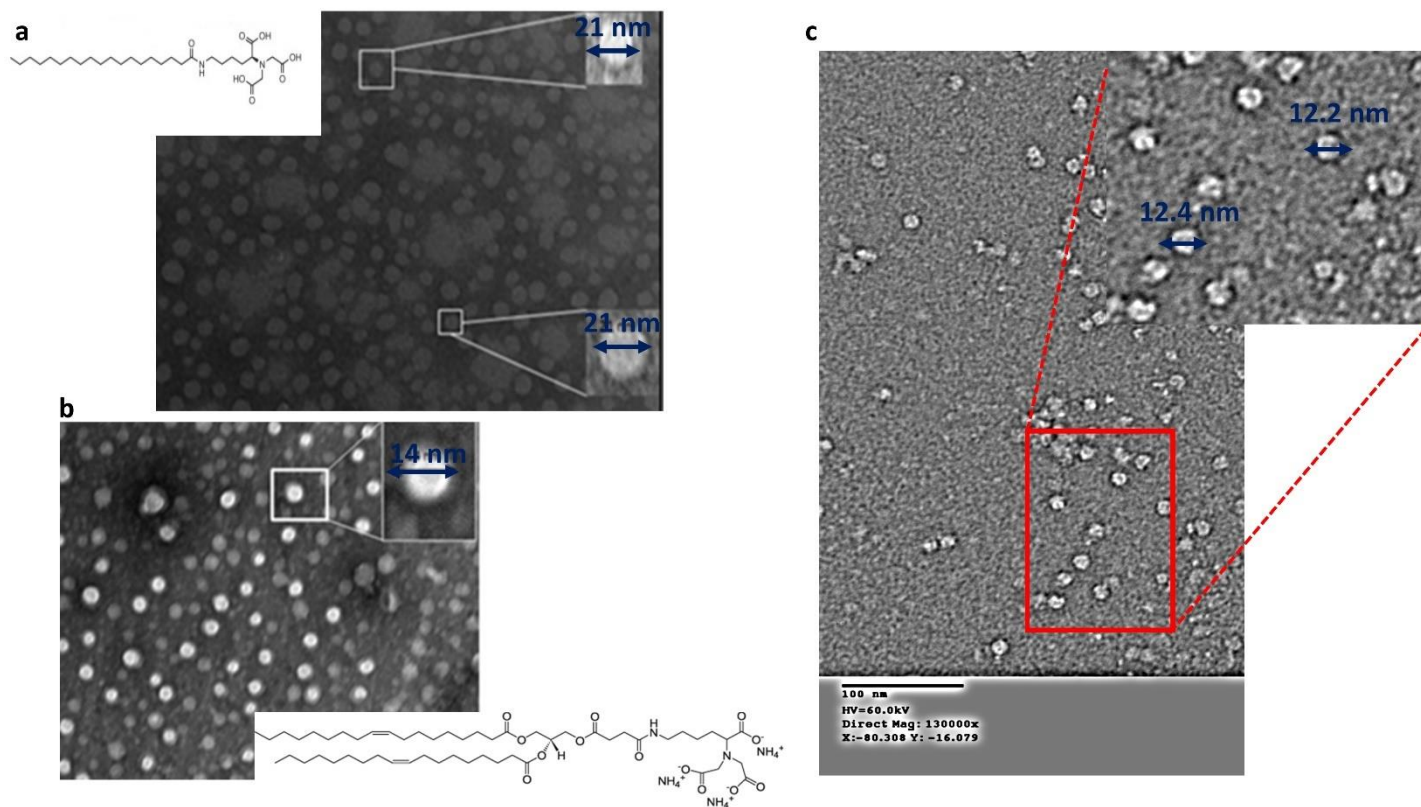
Figure 3.7.2. Preliminary Bfr hydrophobic engineering. (a and b) TEM images for the encapsulation studies performed by Omaima Ben Krayem, using 2 mg/ mL C19-NTA and 0.95 mg/ mL DGS-NTA respectively, in 100 % HisBfr (c) The TEM image of the encapsulation of 0.5 mg/mL C19-NTA with 100 % HisBfr subunits employed.
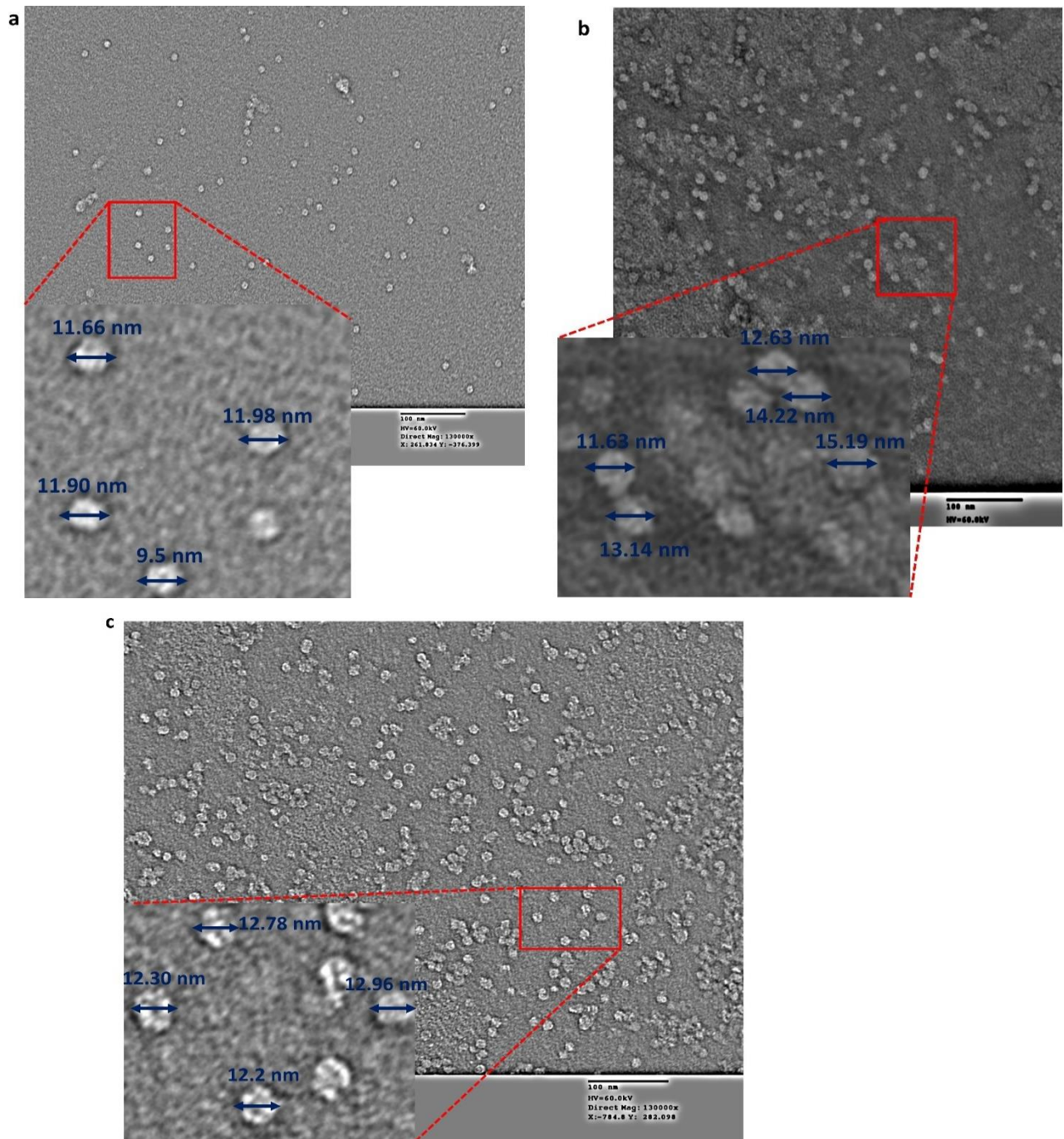
Figure 3.7.3. Optimized Lipid NTA encapsulation. (a) The TEM image of the complexes formed by employing C19-NTA in a 50:50 mixture of His-tagged and WT Bfr (b) TEM image of the complexes formed by employing C16-NTA in 100 % HisBfr (c) TEM image of the complexes formed by employing C14-NTA in 100 % HisBfr. All of the TEM samples were stained with 0.5 % molybdate stain solution.

3.6.1. Determining the Critical Micellar Concentration (CMC) of C19-NTA

To study the dynamic nature of the entrapped lipids as well as its tendency to trap useful cargo such as hydrophobic drug molecules, it is necessary to know the aggregation number and the CMC of the trapped micelles. Hence pyrene excimer excitation was used to obtain the CMC of free C19-NTA in the buffer (150 mM NaCl, 50 mM MOPS-pH 7.5) as described by Mènard *et al*.[118] Firstly, the pyrene solution (OD_{600} 0.04) was made in a buffer (150 mM NaCl, 50 mM MOPS-pH 7.5) and used to make a 2 mM C19-NTA stock solution. The stock solution was then used to make ten serial dilutions (from 1.5 mM to 0.02 mM, Figure 3.7.4 a). The emission spectra were collected from 350-600 nm (1 nm slit width), with 336 nm excitation wavelength (1 nm slit width). The local environments of the pyrene molecules were estimated from the I_1/I_3 (intensities at 374 nm/ 386 nm) and I_E/I_M (excimer/ monomer) ratios as seen in the Figure 3.7.4 b below. These experiments were done in collaboration with Abdullah Bassalem, a graduate student in the laboratory of Professor Jean Duhamel (Chemistry Department).

An analysis of the steady decrease in the I_1/I_3 ratio indicated that the pyrene molecules were migrating from a more hydrophilic macroenvironment (buffer) to a hydrophobic microenvironment (inside micelles). Also, the 0.2 mM sample produced the most intense excimer peak, and beyond that concentration, the emission began to drop. The drop in intensity could present the critical micellar concentration (CMC). Increasing the lipid concentration beyond this point produced additional micelles in solution which could entrap monomeric pyrene molecules and decrease the pyrene dimer excimer fluorescence (since pyrene concentration is kept constant) as seen in Figure 3.7.4 a; the C19-NTA CMC was estimated to be 0.2 mM, equivalent to 0.11 mg/mL.

The estimated CMC from this experiment is 0.11mg/mL. Tthis value is 8.3-fold the reported value (0.012 mg/mL),[118] the reported CMC was recorded in water unlike the MOPS buffer used here. Even though the experiments here were performed in buffer only, the results may be useful in explaining the nature of the hydrophobic environment within Bfr upon lipid/micelle encapsulation. There may be multiple pockets of micelles in the encapsulation mixture which were squeezed within the Bfr cavity upon subunit packing. Therefore, it is necessary to analyze the entrapped micelle further to obtain information about the size and how the surfactant chains are packed within the new micelle within Bfr.
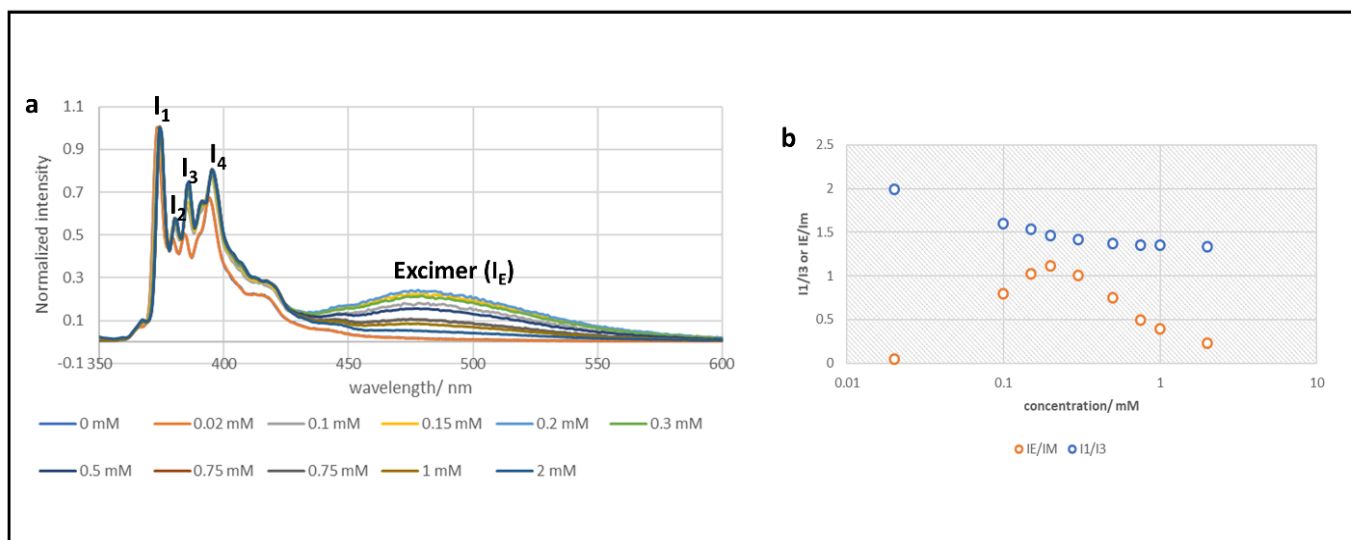
Figure 3.7.4. Pyrene emission spectra. The CMC of C19-NTA was determined by exciting pyrene solutions (fixed concentration, with 0.4 OD), containing varying concentrations of C19-NTA, at 336 nm (a) and plotting the intensities at 374 nm/ 386 nm (I_1/I_3), in blue circles, or ratio of the excimer to monomer (I_E/I_M) in red circles to obtain the CMC (b). The CMC was obtained from the concentration at the maximum (I_E/I_M) or the concentration where I_1/I_3 begins to plateau.

3.6.2. Sedimentation Velocity Analysis on Lipid-encapsulated Bfr

Similar to the velocity analysis described in section 3.2.5, two lipid-encapsulated Bfr samples (a. 100 % HisBfr + heme + C16-NTA, b. 50 % HisBfr + heme + C19-NTA) were analyzed.Unlike the results in Figure 3.2.1 which showed a lower symmetry for mixtures of Bfr subunits (1.2 for 100 % HisBfr and 1.17 for 50 % WTBfr + 50 % HisBfr) the introduction of lipid guest molecule lowers the degree of symmetry to 1.67 and 1.44 for the lipid within 100 % HisBfr and 50 % WTBfr + 50 % HisBfr, respectively. Since the f/fo ratio is related to frictional coefficient this observation may indicate that the surface exposed to the solvent may be less globular than the empty Bfr capsules. Moreover, the observed molecular weights were mostly lower than the expected molecular weights of the complexes (a. 100 % HisBfr + heme + C16-NTA, 509.4 kDa, b. 50 % HisBfr + heme + C19-NTA, 487 kDa) as seen in Figure 3.7.5. The C19-NTA sample (Figure 3.7.5a.) showed most populated species with sedimentation coefficient

of 4.4 S and 8.5 S with molecular weights 95 kDa and 253 kDa, respectively and two more species at 495 kDa (13.3 S) and 994.1 kDa (21.4 S). Conversely, the C19-NTA sample (Figure 3.7.5. b) exhibited a major peak with 4.7 S and molecular weight of 83.9 kDa and a smaller population with average molecular weight of 589.5 kDa having sedimentation coefficient of 17.1 S. While the larger species could result from the entrapment of several micelles within the Bfr cluster as suggested by the pyrene experiments, the observed lower molecular weights could likely be due to incompletely formed Bfr capsules or micelles with few Bfr subunits around them such as a dimer interacting with formed micelles in solution. The results shown here warrant a further look into these samples to estimate the actual aggregation numbers of the surfactant molecules interacting with the Bfr subunits.
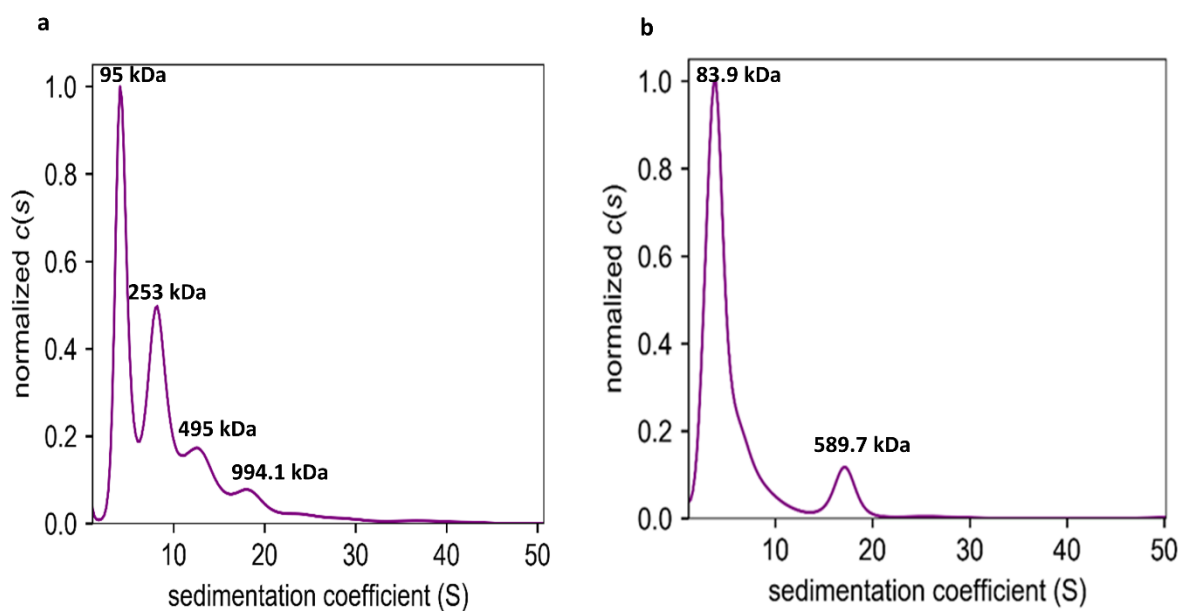
Figure 3.7.5. Sedimentation velocity curves for lipid encapsulated into Bfr. (a) 100 % HisBfr + heme + C16-NTA. (b) 50 % HisBfr + heme + C19-NTA. The various populations are labeled with their average molecular weights. Sample a and b were recorded at 20, 000 rpm and 30,000 rpm, respectively.

3.7. Characterization of Purified *Archaeoglobus Fulgidus* Ferritin (AfFtn)

The AfFtn constructs investigated in this work are similar in dimension and subunit organization to Bfr, yet they contain four 4.5 nm triangular pores. Three constructs were designed and prepared by custom gene synthesis (Genscript Inc): (1) wild-type AfFtn (WTAfFtn) without any modification to the native sequence; (2) hexahistidine-tagged AfFtn which contained $C-H_6$ tags which are exposed on the interior cavity; and (3) a cysteine mutant (M54C) AfFtnC54 which was modified with the reagent maleimide-X-Nitrilotriacetic acid (Maleimide-X-NTA, X is a proprietary linker) and used for His-tagged protein encapsulations. The cysteine location was designed to position the cysteine on one of the helices and able to point towards the middle of the protein cavity (Figure 3.8). Subsequently, an independent report appeared that utilized the same mutation to study the effects of protein structure on DTNB chemical modification. [130] Both the WT and Cys mutant were purified by heat shock, 80 % ammonium sulfate precipitation, and size exclusion chromatography. However, the His-tagged construct was purified by heat shock and immobilized metal affinity chromatography (IMAC, discussed in section 3.1). The SEC step was used to remove DNA contamination from the ammonium sulfate precipitation step as the resulting solution showed a high 260/280 ratio, seen in Figure 3.8.1a.
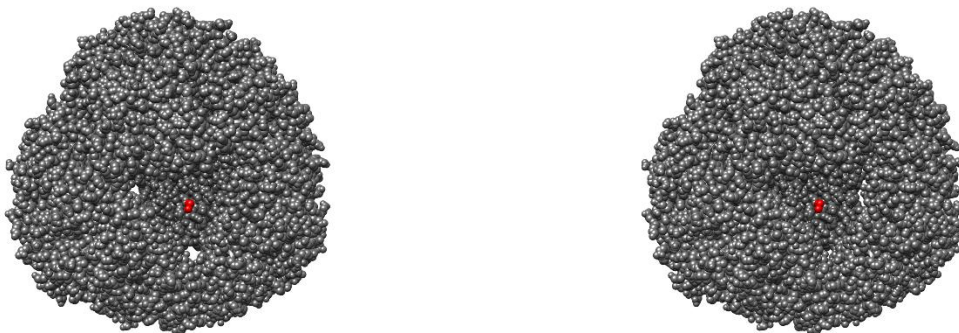
Figure 3.8. Cross-eye stereo image of the AfFtn (1S3Q) showing the position of Met54. The Met 54 (which is the Cys54 mutant position) in red shows that the residue is in the middle of the protein and points to the inside of the cavity. Overall there are 24 subunits with 24 M54C mutations, however only one MC54 mutation is selected in this example. Chimera was used to generate the image.
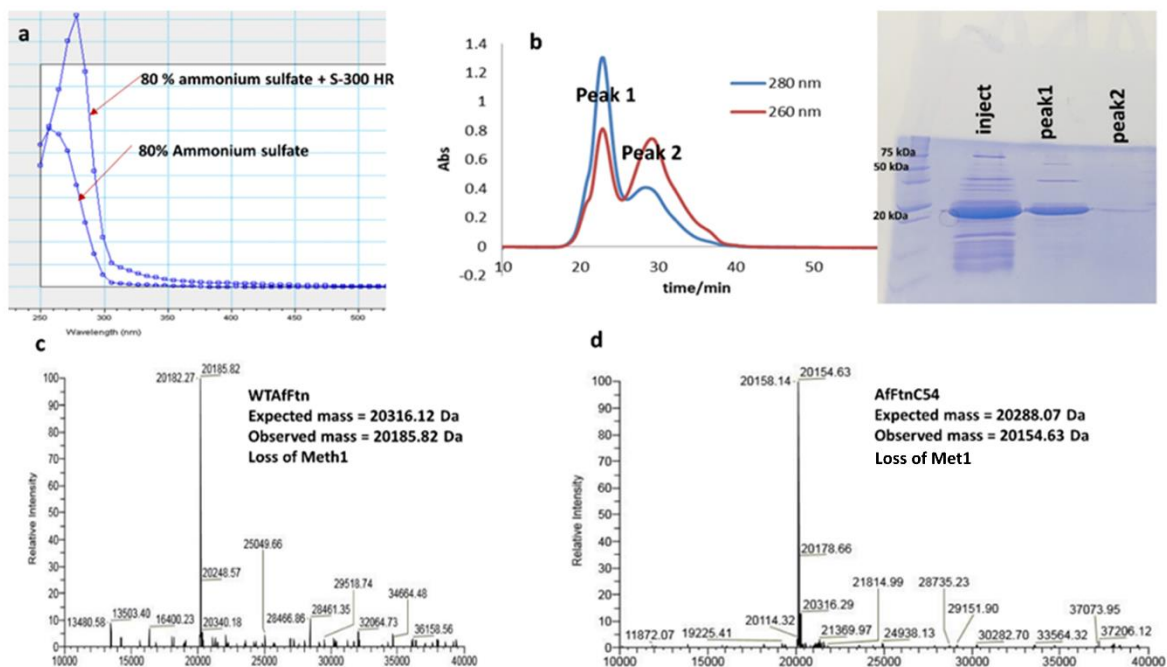
Figure 3.8.1. Characterization of WTAfFtn and AfFtnC54. (a) Comparison the absorbance spectrum of WTAfFtn from the ammonium sulfate step and subsequent SEC. (b) Size exclusion chromatogram for WTAfFtn from ammonium sulfate precipitation step, the 15 % gel (b, insert) used to identify AfFtn containing fractions. (c and d) The deconvoluted +ESI- MS spectra for WTAfFtn and AfFtnC54, respectively. The AfFtnC54 was also precipitated by 80 % ammonium sulfate and run on SEC to remove DNA contamination before acquiring the +ESI data in Figure 3.8.1d

The purified AfFtn constructs were analyzed for their quaternary structures in solution by loading them onto the Sephacryl™ S-300 10/300 column and eluting with varying salt concentrations. The apo-AfFtn is reported to exist as dimers at salt concentrations of 20 mM and below and as 24-mers at 150 mM and above.[8] The salt-mediated association was tested initially on the HisAfFtn construct with buffers containing 50 mM, 150 mM, and 300 mM NaCl. Analysis of both the SEC and DLS experiments, in Figures 3.8.1 a and b, showed the HisAfFtn in 50 mM and 300 mM salt buffers eluted at a retention time corresponding to the size of a 24-mer structure (approximately 45 min) and the size distributions for 150 and 300 mM salt samples were found to be the same (overlap around the average 16 nm), suggesting that they may be forming 24-mers.

132

Despite the slight shift in the dimensions of the 50 mM sample, the protein is still considered to exist as a 24-mer with a few dimers similar to the 300 mM sample in the SEC profile (Figure 3.8.2 a). Additionally, the HisAfFtn was dialyzed into 20 mM NaCl buffer and analyzed on the SEC and was found to contain a few 24-mers (data not shown). Based on the above experiments, the declustering buffer for guest encapsulation was chosen to be 10 mM NaCl, 25 mM HEPES, pH 7.5 to ensure that all the subunits were somewhat dissociated to expose the C-terminal H_6 tags. Even though the AfFtn is reported to be more stable, in 24-mer form, at very high salt concentrations, up to 600 mM,[8] the optimal reclustering and SEC buffers contained 150 mM NaCl, 25 mM HEPES, pH 7.5. Using physiological amounts of salts is not only beneficial for AfFtn association, but it also prevents salting out of future guest molecules.

Similarly, the WTAfFtn construct existed as a 24-mer in 150 mM NaCl (data not shown) as was expected for the M54C mutant (AfFtnC54), but it was dimeric at 150 mM NaCl. The salt-mediated subunit association was affected by the mutation, as was observed by Calisti $et\ al.$[130] Calisti and coworkers successfully showed that the charge distribution at the opening of AfFtn pore screened DTNB from reacting with the reduced SH groups on AfFtnC54. Also, they reported that low concentrations of divalent cations like Ca^{2+} and Mg^{2+} (5-20 mM) more efficiently shift the oligomerization towards the 24-mer form than high NaCl concentrations. Hence the AfFtnC54 in 150 mM NaCl was buffer-exchanged into a buffer containing Mg^{2+} (20 mM $MgCl_2$, 25 mM HEPES-pH 7.5) and run on SEC with the same buffer. It was observed that the Mg^{2+} ions in the buffer were indeed capable of shifting the protein to form 100 % 24-mers in solution unlike NaCl, which resulted in the protein existing predominantly as a dimer (Figure 3.8.2 c, red curve). Further DLS analysis on both samples, in Figure 3.8.1 d, agreed with the SEC results, where the 150 mM NaCl sample was polydispersed unlike the monodispersed 20 mM $MgCl_2$ sample.

Finally, the AfFtnC was modified with an NTA analog into AfFtnC54-NTA and dialyzed with a buffer containing $MgCl_2$ to form the 24-mer, and the 24-mer AfFtnC54-NTA was studied by AUC. Analyis of the results show that the species in 20 mM $MgCl_2$, 25 mM HEPES buffer-pH 7.5 are mostly intact 24-mers as seen the Figure 3.8.2.1. The sedimentation pattern showed mostly one population of 471 kDa molecular weight at 17.7 and a minor peak of 932.5 at 27.9 S. The symmetry (fo/fo =1.21) very similar to Bfr in Figure 3.2.1.
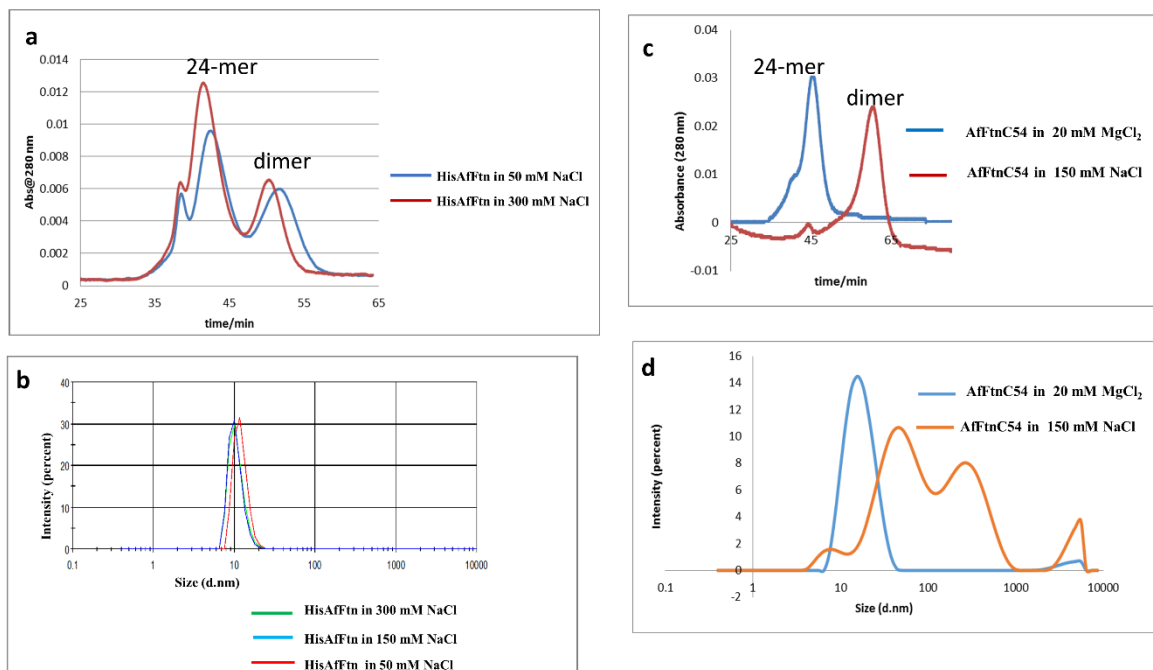
Figure 3.8.2. Effects of sodium chloride and magnesium chloride on AfFtn. (a) Elution profile of HisAfFtn at 50 mM NaCl (blue) and 300 mM NaCl (red). (b) The DLS analysis on HisBfr at 50 mM NaCl (red), 150 mM NaCl (blue), and 300 mM NaCl (green). (c) SEC profile of AfFtnC54 in 150 mM NaCl (red) and 20 mM $MgCl_2$ (blue). (d) DLS of samples in AfFtnC54 in 150 mM NaCl (blue) and 20 mM $MgCl_2$ (orange).
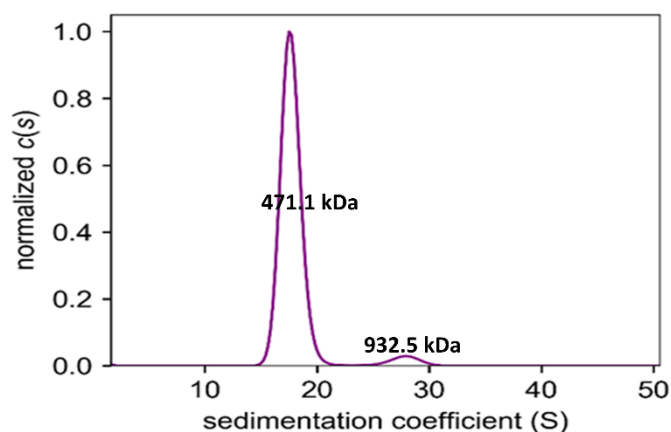
Figure 3.8.2.1. Sedimentation velocity analysis on AfFtn-NTA. The figure here was obtained from the residuals and size distribution curves in Appendix 12. Speed used; 20, 000 rpm.

3.7.1. *Archaeoglobus Ferritin* **Encapsulation Studies**

3.7.1.1. Encapsulation of a 5 nm Gold Nanoparticle (AuNP)

      Similar to the Bfr, the AfFtn construct with C-terminal (interior) polyhistidine tags was used to encapsulate a model inorganic guest, a AuNP via the H_6-Ni^{2+}-NTA interaction. This was to determine whether this unusual ferritin having multiple large pores in its structure would behave in a similar manner to Bfr in regards to encapsulation strategies. For example, it could be possible that the larger pores complicate the encapsulation process in some way. Most of the reaction conditions employed were the same as those utilized for the formation of the Bfr-AuNP host-guest complex, but the declustering of AfFtn subunits was mediated by low ionic strength. The declustering buffer used in these studies was 10 mM NaCl, 25 mM HEPES-pH 7.5. First, the AfFtn subunits were allowed to interact with the guest molecules at room temperature for 90 minutes with slow stirring. The second step involved AuNP encapsulation/entrapment, where the AuNP was entrapped within the AfFtn cluster by increasing the ionic strength of the buffer (to 150 mM NaCl, 7.5pH). The encapsulated AuNP was isolated from the reaction mixture by SEC. Again, the SEC profile showed the presence of the AfFtn having the size of a 24-mer structure with 518 nm absorbance, indicating that the AuNP was successfully encapsulated within the AfFtn cluster as seen in Figure 3.8.3a. The TEM images of free HisAfFtn (Figure 3.8.3b) were compared to the AuNP encapsulated HisAfFtn (Figure 3.8.3c). The AuNP were encapsulated within the AfFtn imaged without added contrasting agents as a very dark interior with low contrast soft material (protein shell) around it. The size distribution of the TEM images was estimated to be around 12 ± 1.8 nm using the ImageJ software.[148] Analysis of these results is consistent with the conclusion that AfFtn can be successfully engineered to encapsulate large non-native guest molecules within its interior cavity. These results should serve as important information to help guide more complex encapsulation studies, such as the entrapment of enzymes or micelles within the protein cavity. The additional large pores available in this type of ferritin might then be employed to selectively screen substrate specificity of enzymes that work on substrates of varying length, such as a glycohydrolytic enzyme. Additionally, the dynamics of fatty acid transfer into and out of such porous ferritin might be of interest in the control of the micellular composition of an entrapment micelle.
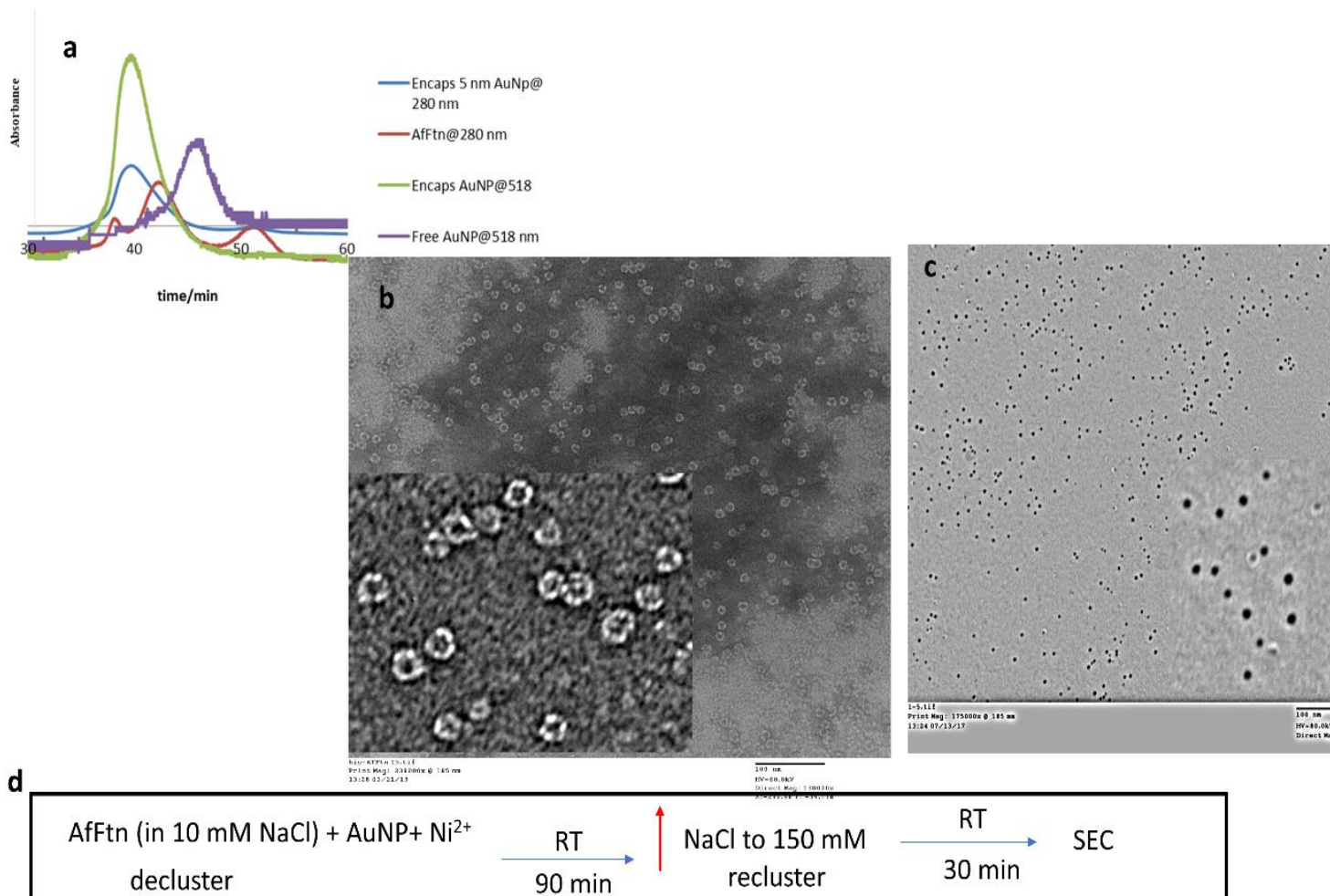
Figure 3.8.3. Encapsulation of AuNP within AfFtn. (a) Size exclusion profile showing that AuNP having NTA labels coelute with the 24-mer AfFtn. (b) The TEM image of HisAfFtn without any guest molecules. The molybdate stain percolates into the interior of the 24-mer cluster. (c) The presence of high electron density AuNP within AfFtn cluster without any molybdate staining. (d) The simplified reaction scheme for the AuNP encapsulation

3.7.2. Encapsulating His-tagged Guest Molecules into the *Archaeoglobus Fulgidus* Ferritin

3.7.2.1. Cysteine Modification

In order to introduce the NTA functionality onto the Cys54 residue in AfFtn (AfFtnC54), a bioconjugation approach was followed which used tris (carboxyethyl) phosphine (TCEP) as a reducing agent to reduce the Cys sulfhydryl groups in the event that during purification and/or storage, two subunits would come together and form a disulfide linkage. The sulfhydryl groups might be expected to be oxidized under the modification/ storage condition (pH 7.5) if trace amounts of oxygen were present. Hence the TCEP incubation step and inert conditions would be required to reduce the disulfide bonds, preparing the free thiol required for the modification. Upon reduction, the thiolate anion (nucleophile) would then be able to attack a maleimide acceptor to form the Cys-S-maleimide product as seen in Figure 3.8.4 for the simple maleimide, N-ethylmaleimide.
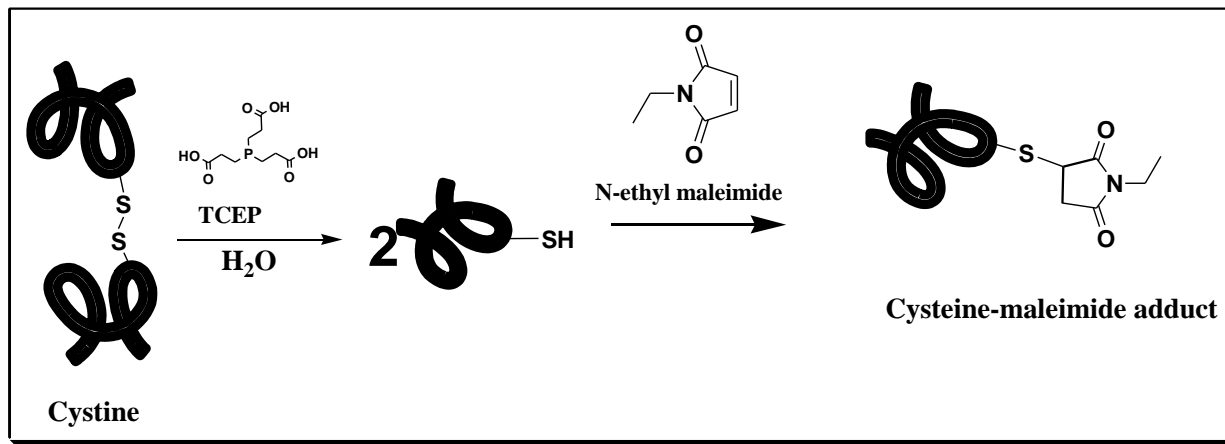
Figure 3.8.4. Representation of cysteine modification with N-ethylmaleimide. Figure generated with ChemDraw.

The AfFtnC54 protein (in 20 mM NaCl, 25 mM HEPES-pH 7.5) was incubated with TCEP (100 equiv. at room temperature, for 15-30 min.) and then N-ethyl maleimide (1 equiv. AfFtn: 100 equiv. TCEP: 50 equiv. N-ethyl maleimide final ratios) was added to the incubation mixture. It was felt that if the protein partially or completely existed as the disulfide-linked dimer (no other cysteine are present in the sequence of the AfFtn), then the addition of TCEP would address this possibility and prepare the protein for subsequent reaction with the maleimide reagent. N-Ethylmaleimide was used as a simple model maleimide since it is less expensive than the commercial maleimide-X-NTA that was to be employed to attach an NTA functionality to the protein. The resulting mixture was further incubated at different time points (at room temperature) to obtain the optimal reaction time. Samples were collected, buffer exchanged into DDH$_2$O and analyzed by ESI mass spectrometry (at time points 1 h, 4 h and overnight, 19 h), and the results were compared in Figure 3.8.5. It was observed that most of the protein in solution remained unmodified (20154 Da) despite the length of incubation (Figure 3.8.5 a-c). This was initially unexpected and indicated that the ratio of the TCEP reagent to N-ethylmaleimide perhaps was important but also it seemed that TCEP might be neutralizing the maleimide due to the expected nucleophilicity of the phosphorus in the TCEP molecule. It was therefore decided that the reaction with N-ethylmaleimide should be attempted again, but without the TCEP reagent. Indeed, the omission of the TCEP incubation step led to 100 % formation of the expected N-ethyl maleimide-AfFtn adduct with a mass of 20281.06 Da (Figure 3.8.5d). The observed mass was the M-1 peak. These conditions were then used for the NTA maleimide as an acceptor molecule. Figure 3.8.5 e shows the modification in the presence of TCEP, whereas Figure 3.8.5 f is without TCEP. While the TCEP-containing sample showed peaks corresponding to both modified and unmodified AfFtnC54 at 20570 Da and 20158 Da, respectively, the reaction without TCEP showed complete modification without any traces of the AfFtnC54 unmodified subunit.

The reduction step with TCEP continues to be utilized for cysteine bioconjugation since it has been reported to not react with maleimide, to be odorless, and more effective than both dithiothreitol (DTT) and β-mercaptoethanol (BME). [155,156] A recent use of TCEP for ferritin modification was reported by Calisti *et al*.[130] They reported successful modification of a cysteine mutant of *Archaeoglobus fulgidus* ferritin with maleimide NTA in the presence of 3 mM TCEP. However, the reduction step has been reported to produce maleimide-TCEP adducts that significantly reduce the yield of Cys-S-maleimide. The addition of TCEP to the reaction buffer

containing the maleimide has indeed been found to react with the maleimide moiety forming an adduct which has been identified (Figure 3.8.6) and is in fact as rapid as maleimide-Cysteine adduct formation.[157] A recent IR characterization by Kantner *et al.* identified the adduct as the ylene (approximately 70 %).[157] instead of the previously proposed ylide/ylene resonance forms. Unfortunately, the ylene is reported to be very stable and not susceptible to nucleophilic attack by thiol groups over wide pH ranges.[158,159,160] The stability, therefore, completely hinders the Cys modification process.
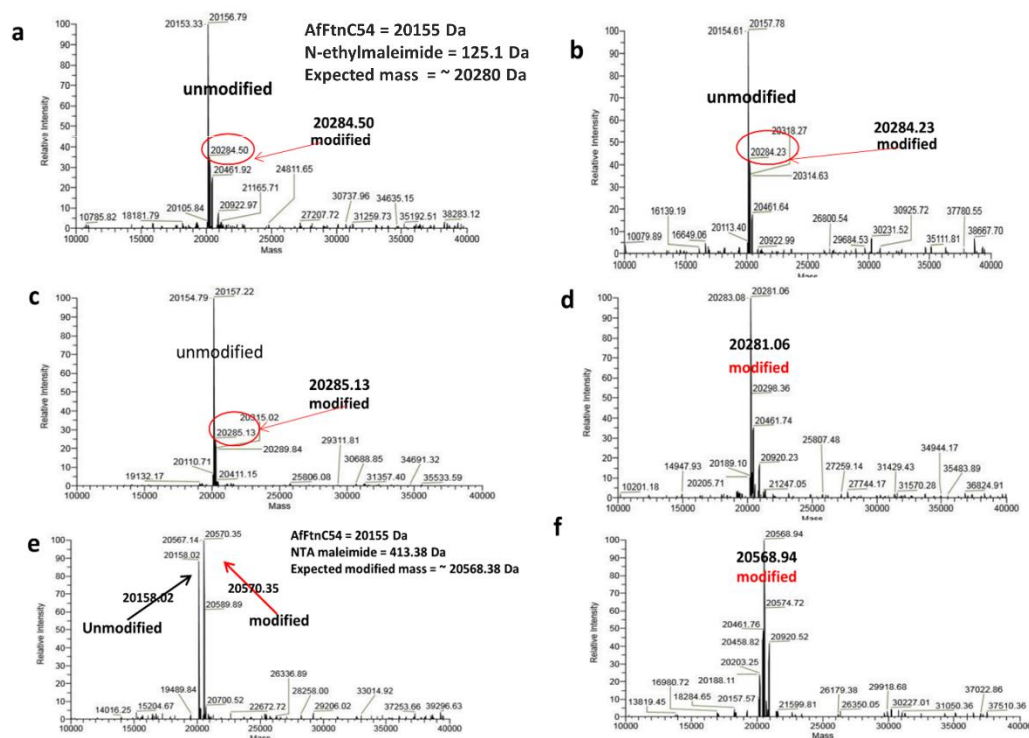
Figure 3.8.5. Optimization of the AfFtnC54 maleimide modification. Initial modification using N-ethyl maleimide, TCEP, and AfFtnC54 in ratios 50 equiv.: 100 equiv.: 1 equiv., respectively. (a) incubation for 1 h (b) 4 h (c) overnight (d) modification without TCEP; 1 equiv. of AfFtn: 50 equiv. N-ethyl maleimide (e) replacing N-ethylmaleimide with maleimide NTA ratios: 1 equiv. AfFtn: 100 equiv. TCEP: 50 equiv. Maleimide NTA (f) incubation without TCEP: 1 equiv. AfFtn: 50 equiv. Maleimide NTA.
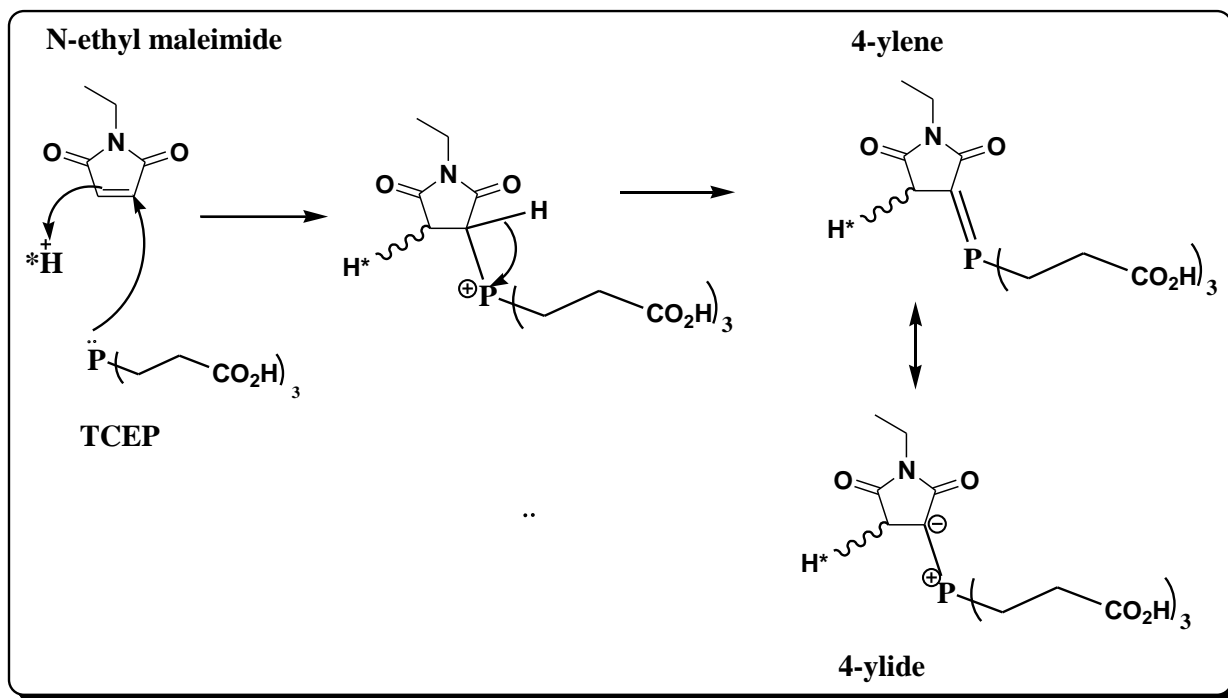
Figure 3.8.6. The reaction of TCEP on a maleimide. The formation of undesirable products (ylene and ylide) as demonstrated by Kantner and co-workers .[157] Figure made with ChemDraw.

The results herein agree with what was reported since the omission of TCEP led to increased maleimide modification both for N-ethylmaleimide and maleimide-X-NTA. However, the mass of ylene was not observed in the deconvoluted spectra (of Figure 3.8.5a-c and e) due to the mass range used for the spectral analysis that focussed more on the AfFtn subunit mass. The curved shape of individual AfFtnC54 subunits may not allow intersubunit disulfide bond formation. Hence the cysteines exist in the reduced form. Moreover, addition of Ellman's reagent (DTNB) to the protein solution resulted in a yellow solution that absorbed at 430 nm (data not shown). Finally, the optimal reaction condition was set to 1 equiv. AfFtn: 20 equiv. Maleimide-X-NTA incubated at room temperature for a minimum of 1h.

3.7.2.2. Encapsulation of His-tagged Proteins

Initially, HisGFP was used as a model to study the conditions for encapsulating His-tagged enzymes such as a glucohydrolase. Unlike the HisGFP encapsulation into Bfr, which was mediated by heme-NTA, the HisGFP was encapsulated by the mediation of the Cys-S-NTA functionality on the AfFtnC54 subunit as illustrated in Figure 3.8.7 below. The modified AfFtnC54-S-NTA was washed (in a spin column) three times with declustering buffer (10 mM NaCl, 25 mM HEPES -pH 7.5) to remove excess maleimide-X-NTA. The washed sample was then incubated with HisGFP (in the same buffer) in the presence of Ni^{2+} (1 equiv. AfFtnC54-NTA, 2.43 x 10^{-9} moles: 1 equiv. HisGFP, 2.43 x 10^{-9} moles: 20 equiv. Ni^{2+}, 4.86 x 10^{-8} moles). The reaction mixture was allowed to equilibrate at room temperature (approximately 23 °C) with slow stirring for 30 minutes. After equilibration, the mixture was divided into two vials and one was incubated at room temperature and the other at 50 °C (for 60 minutes each). The ionic strength of the resulting mixtures was increased by adding 250 mM NaCl and dialyzing against the reclustering buffer (150 mM NaCl, 25 mM HEPES-pH 7.5) overnight to allow the AfFtnC54 subunits to associate in the presence of HisGFP. The final mixtures were run on a Sephacryl™ S-300 HR column to isolate the encapsulated HisGFP.

The encapsulation results indicated that there was an association between the HisGFP molecules and AfFtnC54. However, the AfFtnC54-Ni^{2+}-HisGFP complex eluted at approximately 55 min instead of the 45 min retention time associated with the 24-mer AfFtn for these particular conditions with this specific column (Figure 3.8.8). This suggested that the HisGFP associated with the AfFtn subunit upon incubation since it eluted earlier than the free HisGFP (65 minutes). Also, the incubation temperature does not affect the interaction between the AfFtnC54-NTA subunits and HisGFP as the red (room temperature), and green (60 °C) curves overlap, as shown in Figure 3.8.8. Also, buffer containing only 10 mM NaCl was able to sufficiently expose the NTA functionalities on individual subunits for interaction.

Hence, the encapsulation protocol was optimized by (i) Pre incubating 1 equivalent AfFtnC54-NTA (in 10 mM NaCl, 25 mM HEPES, pH 7.5) with 50 equiv. Ni^{2+} at room temperature for 15 minutes. (ii) Mixed with 1 equiv. of HisGFP and allowed to equilibrate at room temperature for 30 minutes. (iii) Incubation at 60 °C for 30 minutes. (iv) Further incubation at room temperature for at least 4 (up to 12 h) with slow stirring. (v) Recluster via dialysis, buffer (20 mM MgCl2, 25 mM HEPES-pH 7.5) for at least 24 h at 5 ˚C. (vi) Size exclusion

chromatography to separate encapsulated HisGFP from the remaining components. The results of this optimized protocol are presented in Figure 3.8.9. The SEC profile (Figure 3.8.9 a) showed a peak eluted before the 24-mer retention time (40 min) with HisGFP absorption (490 nm, red) and another at the 24-mer retention time of 45 min without the 490 nm peak. They are indicating that there was some population of encapsulated AfFtnC54-NTA that are physically larger than the intact empty 24-mer. Hence, TEM analysis (Figure 3.8.9 b) was used to verify that the fractions (at 40 min) were indeed encapsulated AfFtnC54-NTA-HisGFP and not aggregates of proteins. The TEM image showed mostly encapsulated AfFtn (filled without stain penetration) and a few unencapsulated (empty with stain penetration) capsules. All the capsules, filled and empty, had an average diameter of 14.43 nm.

The observations illustrate that reclustering AfFtnC54 in the presence of guest molecules in a buffer contained low divalent ions followed by increasing salt concentrations is an effective method for guest encapsulation for the chemically modified AfFtn. Future work will explore the use of Mg^{2+} as another method to generate the 24-mer structure after encapsulation.
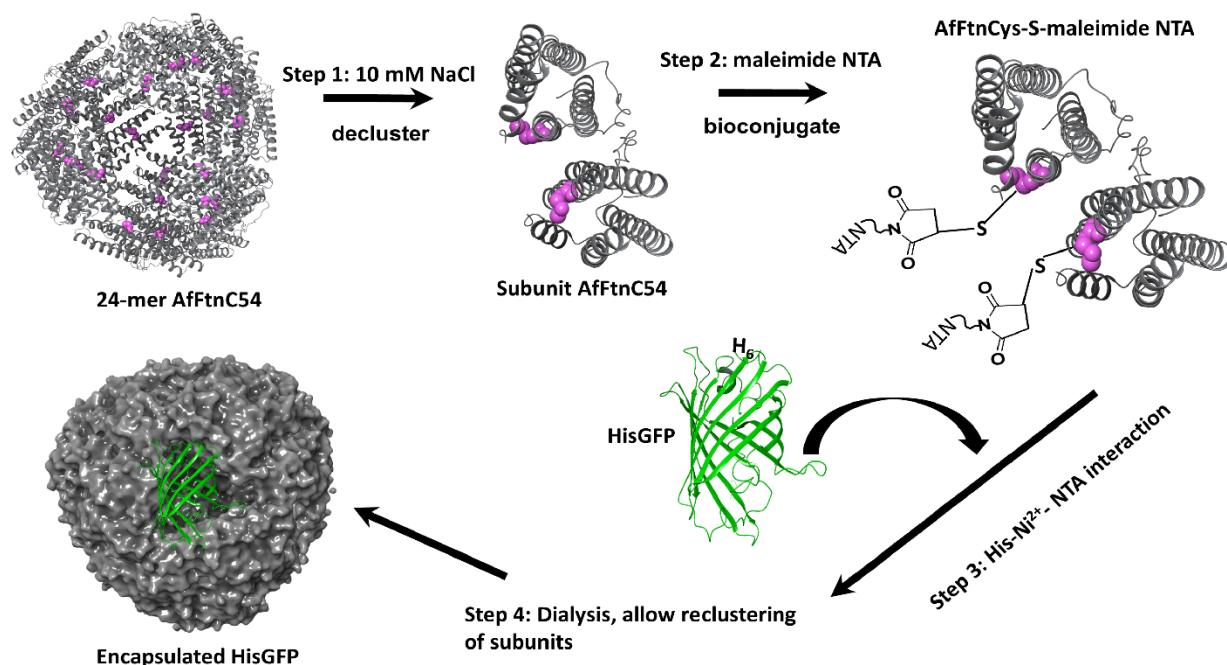
Figure 3.8.7. Schematic representation of HisGFP (1C4F) encapsulation into AfFtnC54 (1S3Q). The C54 mutations are illustrated in purple spheres. The protein crystal structures and the maleimide structurs were generated in Maestro (Schrödinger LLC) and ChemDraw, respectively.

142
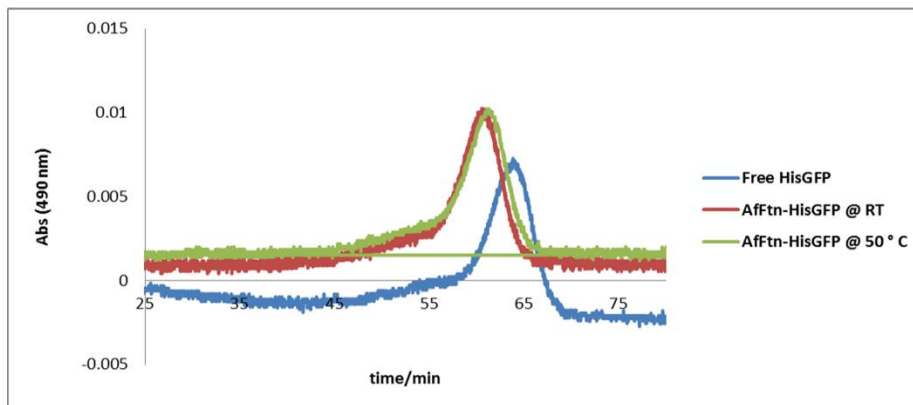
Figure 3.8.8. Encapsulation of HisGFP into AfFtnC54-NTA. The blue curve represents free HisGFP elution on the Sephacryl™ S-300 10/300 HR. The red and green curves represent HisGFP interacting with AfFtnC54-NTA subunits at room temperature (RT) and 50 °C, respectively. The shift in retention times from the free HisGFP indicates the formation of new AfFtnC54-NTA-Ni^{2+}-HisGFP complex
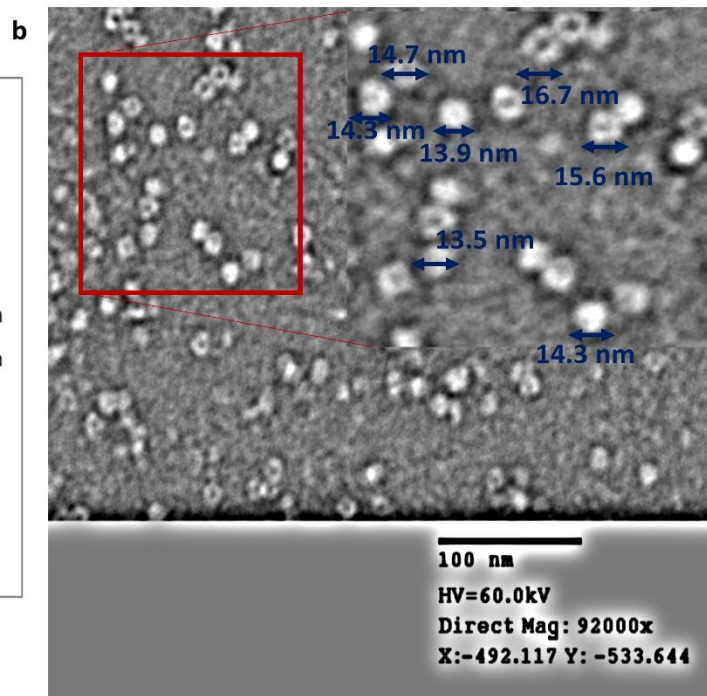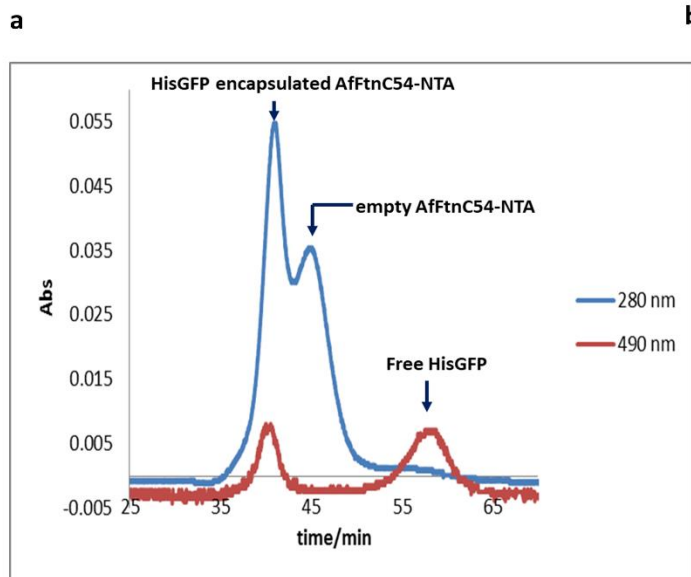
Figure 3.8.9. Optimized HisGFP encapsulation into AfFtnC54-NTA. (a) Sephacryl S™ 300 10/100 HR profile (at 0.5 mL per minute). (b) The TEM image of pooled fractions (retention times 35-42 min). The sample was stained with 0.5% of molybdate stain.

3.8. Encapsulation of Streptavidin within AfFtn

The strategy and protocols for the encapsulation of another model protein, streptavidin, were similar to those employed for encapsulation of the nanoparticle AuNP in terms of the affinity interaction that permits the important guest-capsule protein subunit interaction to occur before reclustering is initiated. However, preliminary experiments (in our laboratory) have shown that the composition of the host Bfr could not be 100 % His-tagged to obtain a successful encapsulation for this protein. A mixture of 60 % His-tagged to 40 % Wild-type (WT) Bfr subunits were used to decrease the amount of internal space that is occupied by the twenty-four His-tags of the shell. The streptavidin that was utilized in this work possesses a fluorescein fluorophore attached to the protein. The fluorescein modification exhibits 495 and 520 nm absorbance and excitation maxima, respectively.

Firstly, the streptavidin-fluorescein (SF) protein was made to quickly interact with biotin-X-Nitriloacetic acid (biotin-X-NTA) to produce the streptavidin-biotin complex seen in Figure 2.8. The complex has one of the strongest dissociation constants (4 x10^{-14} M)[161] and is stable under harsh conditions such as extremes of pH and temperature, high concentrations of chaotropes, and organic solvents.[162] The properties of the complex coupled with the robust ferritin structure allows the encapsulation to occur in either 8 M GdnHCl or at 50 °C for up to 90 min of incubation with neither significant denaturation nor precipitation. The 8 M GdnHCl declustering buffer or higher temperature was used to separate the ferritin subunits and expose the his-tags to the NTA functional groups on the biotin-streptavidin complex. The resulting mixture was separated on a size exclusion column, and the peak eluting at the 24-mer retention time with fluorescein absorbance was collected for further analysis.

Secondly, the extent of guest protection by the Bfr shell was studied by comparing SF encapsulated into Bfr with Sf encapsulation with the AfFtn ferritin. Previous work in the Honek laboratory (by Anton van der Ven) showed that the fluorescein fluorophore attached to the streptavidin, when encapsulated within the Bfr capsule, through use of the compound biotin-X-NTA, was protected against collisional quenchers such as iodine.[120] Here, The SF molecule was initially encapsulated into AfFn using the salt-mediated encapsulation at room temperature (RT).

144

Subunits (100 % HisAfFtn) were incubated with Ni^{2+} at room temperature, and SF was separately incubated with biotin-X-NTA at room temperature for 30 min. The two reactions were combined and allowed to further incubate (at RT) for 60 minutes and then the ionic strength of the solution was increased to 150 mM by the addition of NaCl. The resulting mixture was further incubated for 30 minutes and separated on a Sepahacryl™ S-300 10/300 HR column, as seen in Figure 3.8.9 a. The elution profile was compared to the free HisAfFtn to monitor the appearance of a 495 nm peak at the 24-mer retention time (approximately 45 min). The chromatogram for the HisAfFtn-SF (red and green curves) showed that the SF molecule was not encapsulated within the HisAfFtn cluster since the SF eluted at a later retention time (around 52 min), similar to where the free SF elutes. The results in Figure 3.8.9a indicated that the encapsulation was unsuccessful.

Due to the several components in the reaction that could be varied, it was reasonable to reproduce the previous work by Anton van der Ven to ensure that all the components were active and the protocol appropriate. SF encapsulation into Bfr employed a mixture of HisBfr and WTBfr at 60 % and 40 % respectively, to ensure that there were sufficient His-tags for NTA interaction while reducing interior steric interaction.

Also, encapsulation was mediated by heating at 40-60 °C to ensure sufficient subunit dissociation without any significant streptavidin-fluorescence. The SEC profile, in Figure 3.8.9 b, showed SF coeluted with Bfr-24-mer and the pooled fractions were analyzed further with UV-vis where a shoulder peak was observed for encapsulated SF at 495 nm (data not shown).

Experiments were undertaken to encapsulate SF employing 100 % HisAfFtn subunits with heating. The subunit association was mediated by low salt (10 mM NaCl), and the SF encapsulation was mediated by both low salt and heating either at 40 °C or 60 °C. However, no encapsulation was observed as no 495 nm peak was observed at 45 minutes in Figure 3.8.9c. Perhaps the use of HisAfFtn for SF encapsulation must be performed with a mixture of wild-type and His-tagged AfFtn subunits just as was found of being successful in employing Bfr. Hence, the WTAftn construct was engineered and purified (described in section 3.7). Finally, a 60:40 % mixture of His-tagged AfFtn to wild type Aftn subunits was used for SF encapsulation, the results, Figure 3.9 d, however, were not as expected. The SF guest did not coelute with the AfFtn host molecule, but SF freely existed in the solution (eluting at approximately 53 min). The encapsulation was tried at 40 °C and 60 °C and produced similar results to Figure 3.8.6 d. (data not shown)

The encapsulation of SF into AfFtn may require different encapsulation conditions which may or may not require subunit mixing. Alternatively, the $MgCl_2$ buffer could be used for reclustering and shifting the equilibrium to the 24-mer complex.
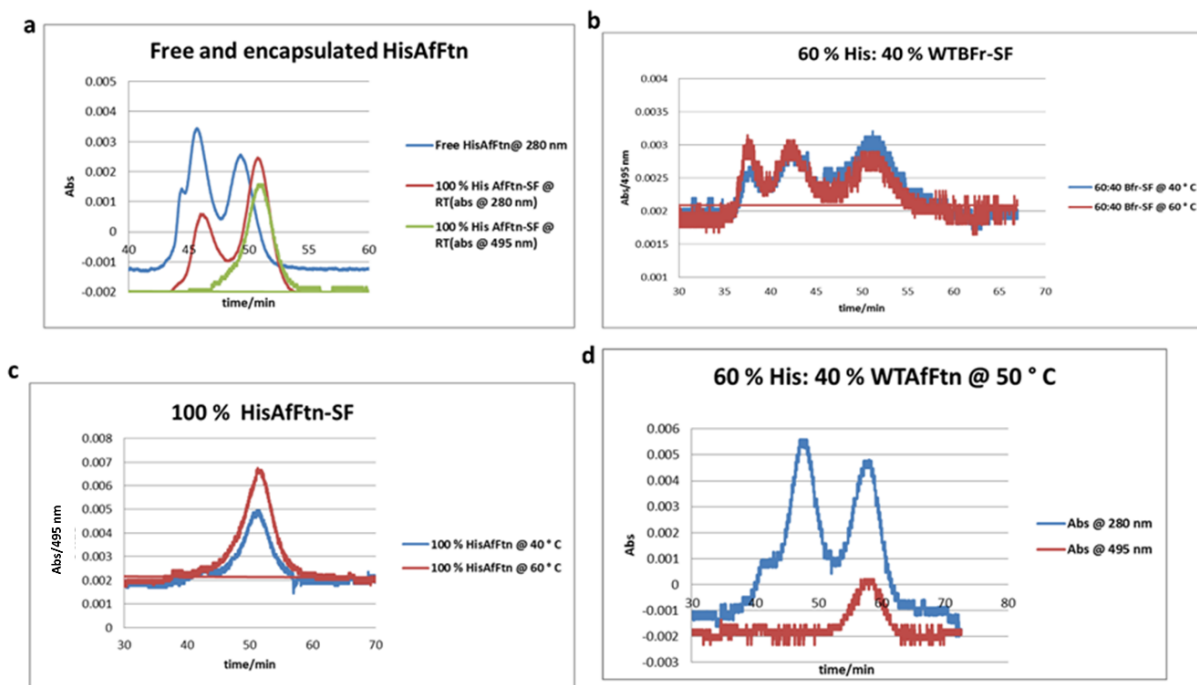
Figure 3.9. Encapsulation of Streptavidin-fluorescein (SF) into AfFtn. (a) The initial encapsulation trials with 100 % HisAfFtn at room temperature. The red line corresponds to the 280 nm absorbance of the HisAfFtn-SF complex, whereas the green is for the same mixture but specifically monitors the 495 nm associated with SF. The blue is the free AfFtn protein (at 280 nm) without SF guest molecules. (b) Reproducing SF encapsulation in Bfr. The encapsulation was tried at 40 °C (blue curve) and 60 °C (red). (c) Encapsulation of SF into HisAfFtn at 40 °C (blue curve) and 60 °C (red curve). (d) Encapsulating SF into 60 % His: 40 % WTAfTn at 50 °C.

CHAPTER 4: BFR PHARMACOPHORE MODELING

4. Current Objective

Current research on Bfr bioencapsulation in the Honek laboratory has focused on driving 24-mer formation via the dimer intermediate. The 24-mer Bfr is then used to encapsulate native proteins (such as streptavidin and His-tagged GFP), a 5 nm gold nanoparticle, nitrilotriacetic acid (NTA) functionalized-lipids, and NTA-labeled dyes. These modifications are all mediated by the heme 24-mer association (Figure 4). The heme cassette has also been employed in a reverse polarity guest encapsulation, where it was modified with NTA and reinserted into apoBfr. This modification allowed the heme to target polyhistidine-tagged proteins into the Bfr cavity specifically. The heme cofactor is not just a structural component, but a means to engineer the Bfr platform. Finding alternative molecules with different properties that are capable of binding into the heme site will be useful in future applications of the Bfr nanoplatforms. For instance, a molecule that locks the 24-mer conformation will be ideal for imaging and drug delivery applications.
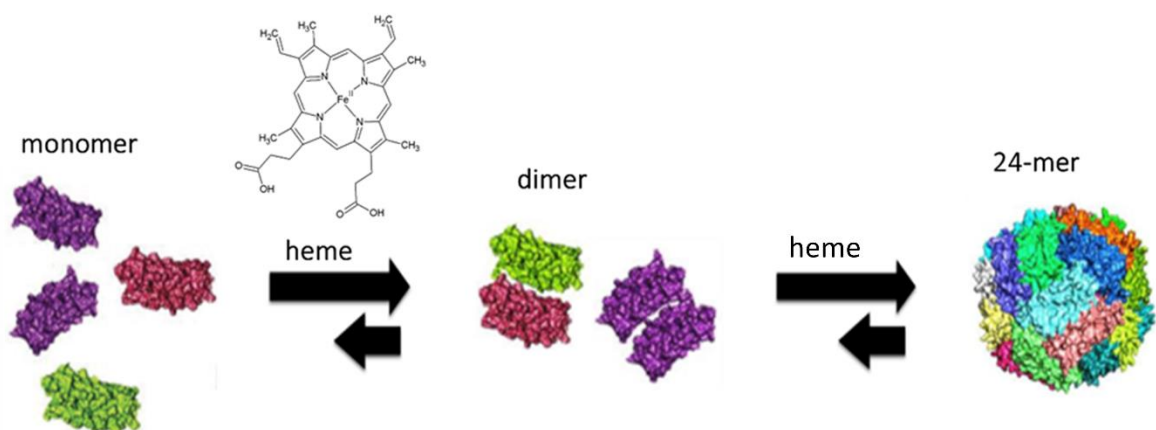
Figure 4. Heme-mediated 24-mer formation in Bfr. The different colors represent the individual subunits within the Bfr 24-mer. This Figure was generated in pyMOL using the 1Bfr crystal structure.

The primary objective of this section of the thesis research was to computationally identify novel molecules that could equally fit the heme binding site and drive 24-mer capsule formation. More specifically it was of interest to explore whether drug docking software, used extensively in medicinal chemistry to identify enzyme inhibitors, for example, could be utilized to identify non-heme molecules that could control the quaternary structure of the Bfr capsule.

Molecular modeling (MM) simulations can be used to screen small molecule databases and eventually narrow the possible molecules that could be experimentally tested on the protein in the laboratory. MM could also provide more information about the energetics of the binding between a possible analog and the heme binding site in the Bfr 24-mer. Insight into the important features required for a ligand to bind to this the site could be obtained from the identification of non-heme ligands. Furthermore, the identified ligand could then be synthetically modified to enhance its aqueous solubility (if it were too water-insoluble to be practically useful). Another advantage of this approach would be the amount of time and money that would be saved here compared to the traditional trial and error methods of experimentally screening small molecules.

4.1. Computing for Bfr Cofactors

4.2. Software

In 1990, Richard A. Friesner and William Andrew Goddard III founded the Schrödinger Incorporation. The software company focuses on developing simulation software for scientific research. The software packages are developed specifically for biotechnology, materials science, and pharmaceutical research. The Schrödinger computation package contains several software applications such as Prime, Glide, LigPrep, Jaguar, Strike, and many more.

Maestro 11 (and its recent update) is a versatile molecular modeling environment from the Schrödinger cooperation, which is not just a graphical user interface, but the portal to all Schrödinger's computational technology. The Maestro 11 interface is designed to assist the user in preparing, organizing, and analyzing data. Maestro supports several file formats (example. SD, PDB, Maestro). Also, the builder tool could be used to construct any other molecular model preferred by the user. Tasks can also be automated and customized via scripting. Another exciting feature of the software is the 3D visualization of objects in the workspace. Figure 4.1 shows the main graphical user interface (GUI) of the Maestro 11 software [123]
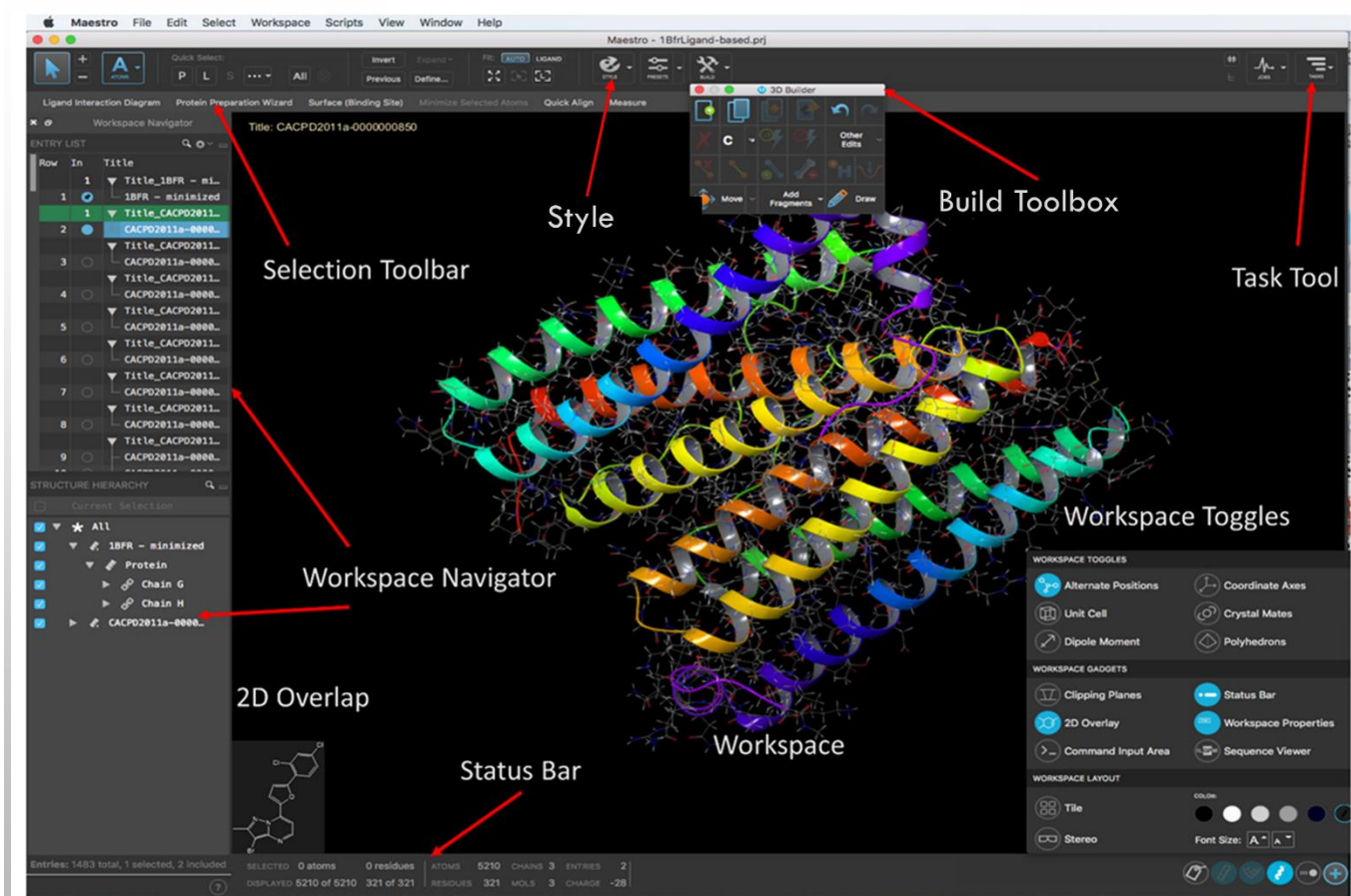
Figure 4.1. The Maestro 11 interface. This figure outlines the various computational features of the Maestro 11 software. The GUI is not different from recent updates, Maestro 12. This image was designed based on the Maestro Quick Reference Guide:

www.schrodinger.com/sites/default/files/maestro_11_qrc.pdf [163]

4.3. Maestro 11 Applications

The Maestro Molecular Modelling interface contains several applications located in the Task Tool. This section includes a detailed description of the primary applications used in this work. These applications are; The Protein Preparation (Prep) Wizard, Phase, Glide Docking, Prime, Ligand preparation, and Induced fit docking applications.

4.3.1. The Protein Preparation Wizard

This application allows the user to directly import Protein Data Base (PDB) structures from the RSCB site [164] or import a previously saved structure. A PDB structure may not be suitable for MM calculations due to the limited resolutions of X-ray crystal structures. The limited resolution results in part of the residues in connecting loops (flexible regions, high B factors) of the protein to be missing from the crystal structure in some PDB structural data files. Therefore, there is the need to bring the structure from its raw state to a more useful state for MM simulation.

The Protein Preparation wizard performs this task in three stages. The first stage involves importing the structure, followed by the performance of basic tasks of fixing the structure in the three main tabs. The first tab is the Import tab, followed by the Review and Modify tab and the Refine tab. The Import tab allows the user to include structures into the entry list. It also contains several sub-tabs that can be used to perform functions such as: align to reference structures, assign bond orders, Add /remove hydrogens, cap termini, create zero bond orders to metal ions, convert selenomethionine to methionine, fill in missing side chains, and fill in missing loops. After the structure has undergone preprocessing, it can be reviewed and modified. The solvent molecules used in crystallization can be removed. Also, parts of the subunits from multimeric proteins can be deleted if they are not be involved in the modeling and the remaining structure can be optimized and minimized in the refine tab. Figure 4.2 shows an example of Bfr preparation and optimization.
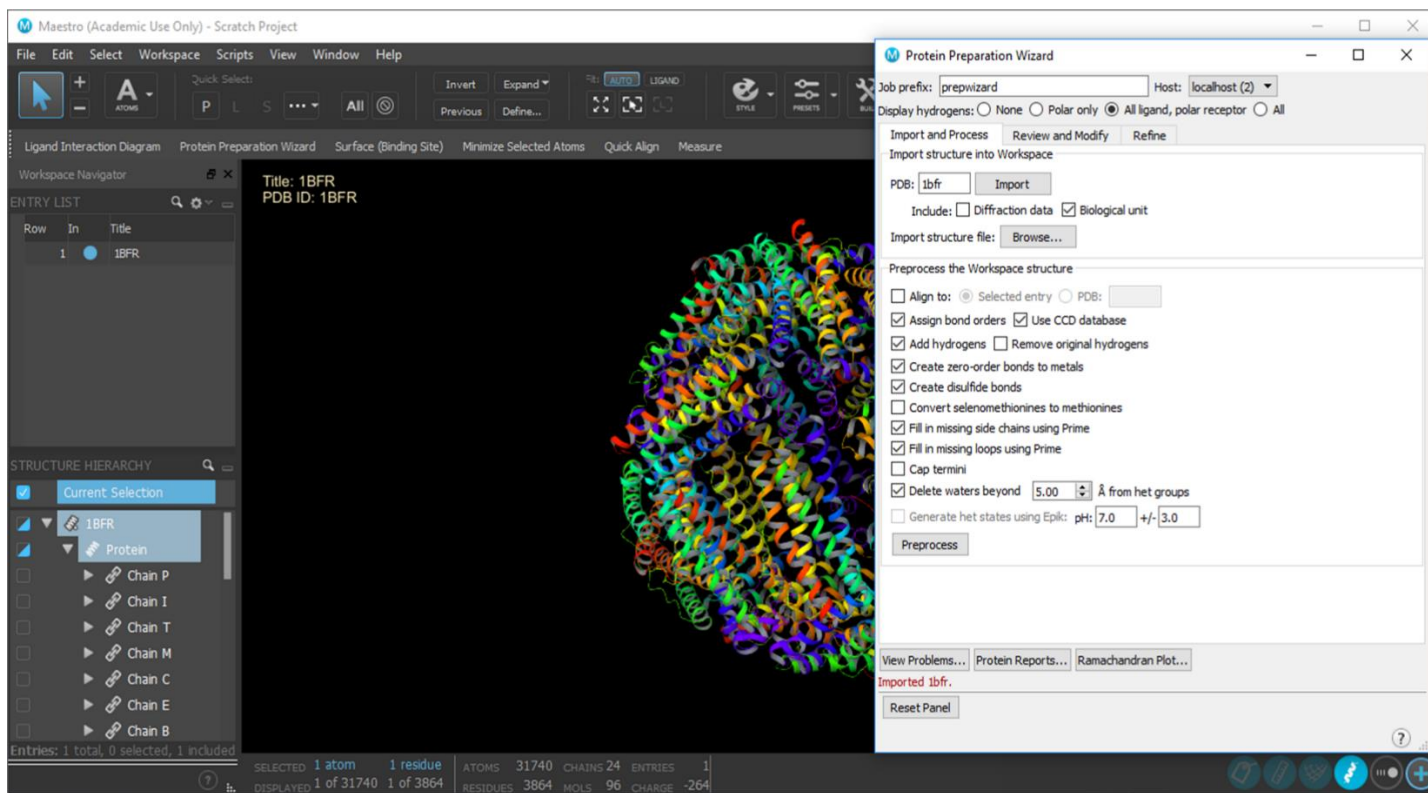
Figure 4.2. Preparation and optimization of the 1Bfr crystal structure. The Figure is a screen capture of 1Bfr imported into the workspace and the protein Preparation wizard highlighting a few of the options employed.

4.3.2. Phase Screen

The Phase Screen application is another computational program available from the Maestro program. The Phase screen application is a virtual screening process that is useful for the generation of a pharmacophore, which is subsequently utilized in database screening. Similar to other virtual screening applications, the Phase screen protocol can rapidly screen millions of molecules from a database and rank them based on their abilities to recognize molecular features of the active site and dock (as illustrated in Figure 4.3). These molecular features are termed a pharmacophore or a pharmacophoric pattern. Hence a pharmacophore is an abstract description of molecular features (both steric and electronic) that a particular molecule or a receptor must have for it to be recognized by a docked ligand. The general concept of a pharmacophore is that different molecules can have the same biological activity provided that they have similar chemical

groups in three-dimensional space. The chemical groups include hydrogen bond donors and acceptors, negative charges, and aromatic groups.[165,166]
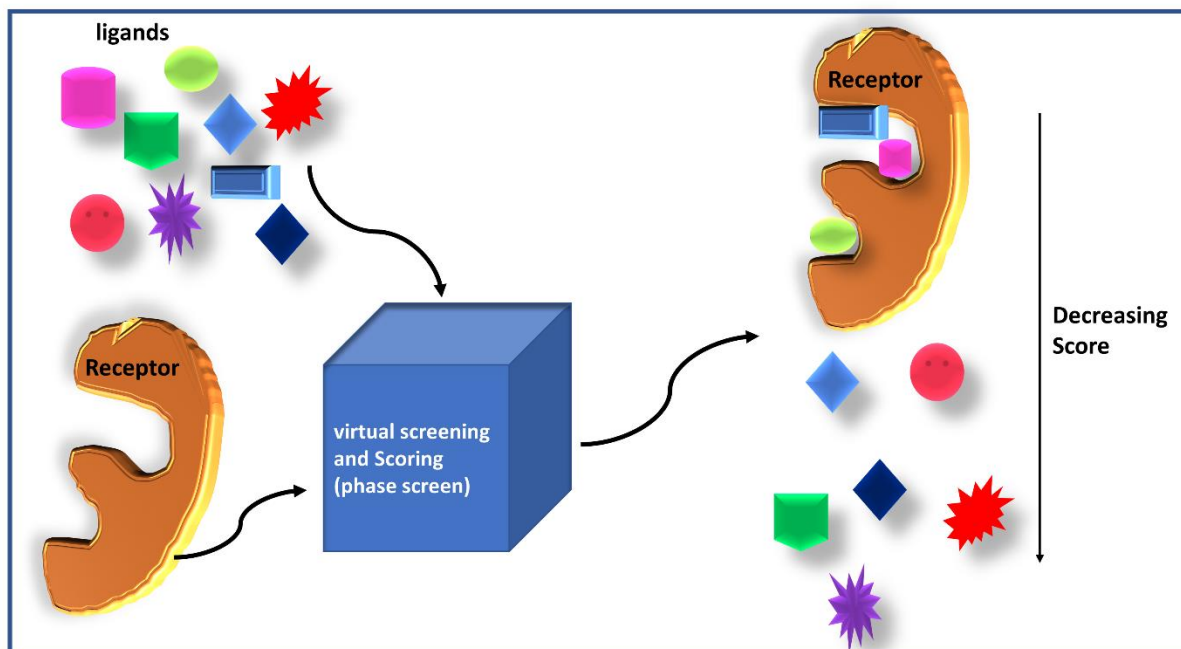
Figure 4.3. A schematic representation of Phase screen. Figure designed in Microsoft PowerPoint

A typical pharmacophore generated from the Phase program illustrates a spatial arrangement of the relevant chemical features that are required for binding. The spatial arrangement may be generated via ligand-based information or fragment-based information, or a mixture of both. The ligand-based approach generates a hypothetical pharmacophore from the structures of a known ligand and receptor, whereas the fragment-based strategy uses only a receptor molecule and its interaction with numerous small molecule probes which generate a pharmacophore based on this "mapping" approach. Alternative ligands could be generated with this approach in cases where there is limited receptor information but sufficient ligand information.[167]

The fragment-based (protein-based), on the other hand, generates a pharmacophore using only the receptor. The receptor molecule is screened with smaller fragment molecules to generate

a pattern of the relevant, interactive sites in the receptor molecule. The 3D arrangement may or may not be very similar to the spatial organization from the ligand-based approach. A library of ligands is virtually screened against these features to identify new ligands that could bind to the receptor molecule. The active ligands obtained from the virtual screening are ranked with the PhaseHypScore (PhaseScore) in the Schrodinger software. The score orders millions of molecules based on their performance in virtual screening as well as the quality of the ligand alignment. [168]

4.3.3. Glide Docking

The Glide high throughput docking screening program is used to screen several molecules for their biological activity.[169] Glide tends to identify active molecules from a smaller database of molecules identified as possible ligands based on the initial Phase screening process. Glide docks ligands to an active site and estimates their binding affinities. Additionally, the conformations of the docked ligands (ligand Poses) and the positions of correctly docked ligands can be obtained and visualized.[170] Virtual screening may be performed with standard precision (Glide SP) or in extra precision (Glide XP). The Glide SP docking employs functions which reduce the number of conformers and torsional refinements used in the docking process, making it less expensive, whereas Glide XP performs an extensive torsional and conformational sampling to eliminate false positives that are obtained from the SP docking. The docking process usually begins with a fast Glide SP docking, and the top ligands are further docked with Glide XP to obtain a docking score in terms of the Gibbs free energy. A typical XP score (equation 1) is made up of the principal terms that favor binding (equation 2) and those that hinder binding (equation 3).[171]

$$XP\ GlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \quad (1)$$
$$E_{bind} = E_{hyd_{enclosure}} + E_{hb_{nn_{motif}}} + E_{hb_{cc_{motif}}} + E_{PI} + E_{hb_{pair}} + E_{phobic_{pair}} \quad (2)$$
$$E_{penalty} = E_{desolv} + E_{ligand\_strain} \quad (3)$$

The E_{coul}, and E_{vdW} terms describe the coulombic and the van der Waals energies, respectively. The hydrophobic enclosure $(E_{hyd_{enclosure}})$ and the hydrophobic pair $(E_{phobic_{pair}})$ functions describe the interaction between the hydrophobic section(s) of a molecule that come into proximity to the receptor cavity as the ligands bind to the receptor. The displacement of the bound water molecules from the two molecules (ligand and receptor) in contact increases the

entropy of the system. The increased entropy favors the binding process. The Protein-hydrogen-Bonding interactions are described by the functions containing the hb terms ($E_{hb_{nn_{motif}}}$ + $E_{hb_{cc_{motif}}}$ + $E_{hb_{pair}}$) where the nn and the cc terms refer to the neutral-neutral motifs and the charge-charge motifs, respectively. The transfer of hydrophobic molecules from the bulk solution upon ligand binding could favor binding energy since there will be more hydrogen bonding between the water molecules that have been displaced from ligand and receptor surfaces into the bulk solution. The overall ionization state (E_{PI}) of the molecules can also enhance binding. Conversely, the desolvation energy (E_{desolv}) function describes the decrease in free energy. The inability of the bound ligand to freely rotate in solution (E_{ligand_strain}) leads to a significant decrease in entropy and leads to less favorable binding.[171]

4.3.4. Ligand Preparation (LigPrep)

Similar to protein structures obtained from the PDB which have to be refined in order to be utilized in a docking simulation, ligand structures from databases or vendors may require initial preparations and optimizations as well. In the Maestro software this is termed the LigPrep program. The LigPrep application converts 2D structures to low energy 3D structures. Also, LigPrep expands the input structure by the generation of various ionization states, tautomers, stereochemistry, and ring conformations which may be present for a potential ligand.

The application is briefly described using the color-coded (steps) in Figure 4.4. First, the ligand structures can be imported into the LigPrep panel by either browsing for the structure files in the top panel (red) of the application or by using existing structures in the entry list (or project table). Next, a criterion (orange) can be used to filter out unwanted ligands such as molecular weights or functional groups. The third step (teal) involves setting the ionization state, which generates all the possible ionization states for each ligand on file. Also, metal binding states can be added when docking to the active site that contains a metal (metalloproteins). Fourthly, the ligand can be desalted to remove unwanted water molecules and counterions from the crystal structure. Additionally, the last but one step (green) generates keto-enol tautomerization, sulfur and nitrogen tautomers, histidine, and DNA base tautomers. Also, chirality information can be retained or generated from the new 3D structures of the ligands. Here a limit can be set on the number of stereoisomers (the default is set to 32 stereoisomers per ligand). Lastly, the output

format (purple) SD or Maestro can be chosen from the output tab and the number of processors to be used for the computation can be chosen from the job settings icon (right of the job name). The prepared ligands, now in 3D with correct ionizations, can then be added to the project table and used for induced fit docking.[172]
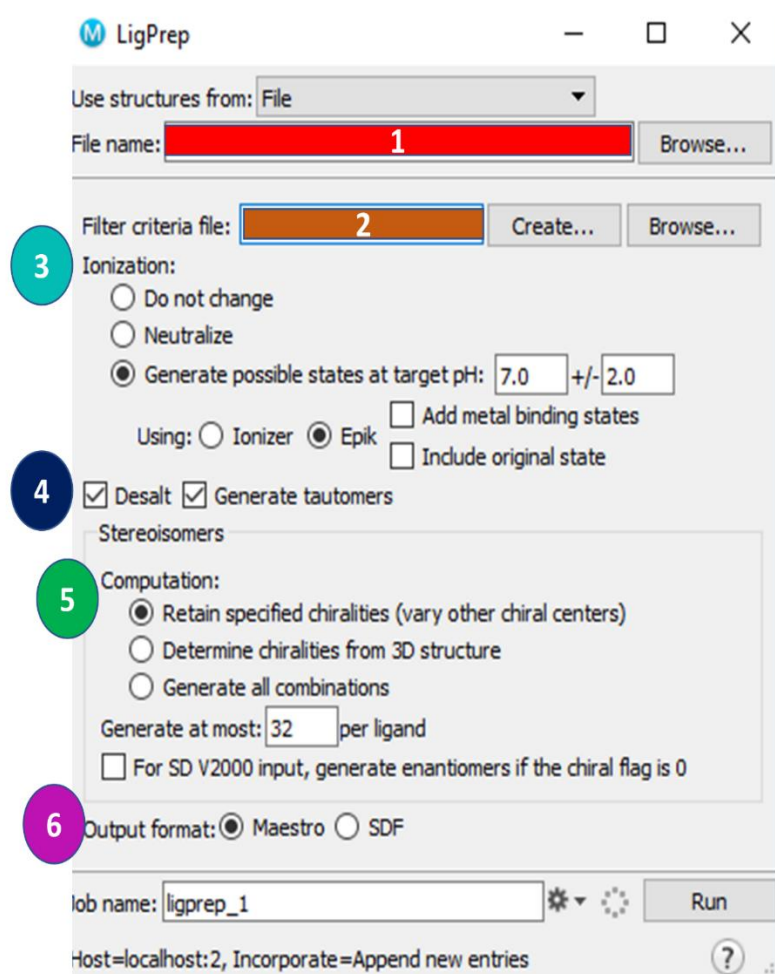
Figure 4.4. The LigPrep application menu screen. A screen capture from Maestro 11, LigPrep application.

4.3.4. Induced Fit Docking (IFD)

The rigid receptor docking (such as Glide) models the docking between a receptor that does not undergo significant changes upon ligand binding and the ligand itself. The modeling is performed by scaling the van der Waal's radii of the ligand binding site and freezing or deleting highly flexible receptor side chains to prevent close contacts[173] But, the process of deleting side chains and radii scaling may result in a wider active site that allows false ligands to dock and

hence produce false positives. The benefit of this approach is that a rapid throughput of ligands is the result which in practical terms allows for rapid screening of large ligand databases, providing a list of molecules from the original database which could potentially bind to the receptor.

To overcome this, and many other, limitations, the soft docking model was described by Jiang and Kim in 1991, which accommodated small changes in protein conformation. In their work entitled "Soft docking": matching molecular surface cubes," the authors modeled minor conformational changes by developing a "cube representation" of molecular surface and volume. This process, described by Jiang and Kim, used a fixed receptor molecule but used a "soft" scoring function to evaluate the interactions. Some of the interactions are charge-charge interaction, hydrogen bonding, van der Waals interaction, and the size and shape of surfaces (Figure 4.5 a).[174] The soft scoring functions are derived from the molecular mechanic's force field to allow for small overlaps between the ligand and the receptor that are used to estimate small plasticity in the receptor molecule within a short computation time. However, ligands with multiple functionalities and diverse sizes may not perform well under soft docking.[175]

Alternatively, ensemble/ multiple docking (Figure 4.5 b) can be used to account for receptor flexibility. In ensemble docking, multiple conformations of the receptor molecule are used to simulate large-scale conformational states. The ensemble docking approach was employed by Kua *et al*. to study the specificity of acetylcholinesterase with acetylcholine. The computed k(cat) and K(M) values were in general agreement with the experimental values. [176] Despite the success of this method, it suffers from the lack of evaluating wide-range receptor conformations. Also, since the receptor conformations are generated before docking and are not allowed to vary during the docking process, it does not correctly account for any ligand-induced conformational change.[177] Hence, the Indiced Fit Docking protocol can be applied to overcome these limitations.

Unlike Glide docking, which employs a rigid receptor to estimate the binding energies quickly, the induced fit docking application attempts to accurately predict the change in receptor conformation and its associated energy/ modes upon ligand binding within a reasonable time frame. The Induced Fit Docking protocol combines both Glide and Prime applications to account both ligand and receptor flexibilities (Figure 4.5c). The receptor is allowed to have some flexibility by decreasing the van der Waals (vdW) radii, and an increased Coulomb-vdW cutoff [169] Also, highly flexible side chains are temporarily mutated to alanine. Hence, the procedure

combines components of the Glide and Prime[178] applications to estimate the energies of the docked active ligand quickly.

The Glide application is used to generate the various binding modes for the ligands and receptors while the Prime application[178] samples the flexibility in the receptor. As mentioned, Glide docking is performed on the 3D ligands from LigPrep to generate diverse Ligand poses. The Ligand poses are then accommodated into the binding site by performing Prime structural prediction which reorients receptor sidechains and trims highly flexible chains to make room for the ligand and the resultant pose for the residues, and the ligand is minimized to obtain the low energy receptor conformations for each ligand. This step is the main difference between the IFD and the Glide docking. In Glide docking there is a single receptor conformation in which all ligands have to dock. The minimized receptors (for each ligand) is used to separately redock their corresponding ligands (by Glide docking) to obtain a final docking score.[173,179]
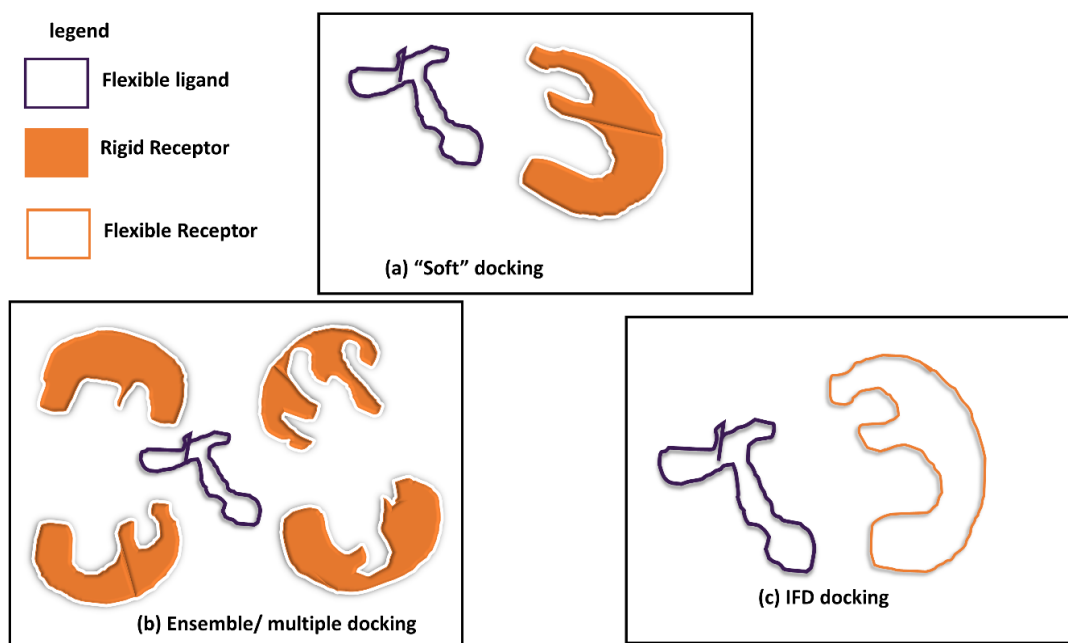
Figure 4.5. A representation of docking models. (a) the "soft" docking model with the rigid receptor and a flexible ligand. Also, the first figure (a) can represent rigid receptor docking in the Glide docking strategy. (b) ensemble docking having multiple rigid receptor conformations and flexible ligand. (c) the Induced Fit Docking approach, having both flexible receptor and ligand structures. This Figure was prepared in Microsoft PowerPoint.

4.4. Force Field

The Optimized Potential for Liquid Simulation All-atom (OPLS-AA) force field was developed by Jorgensen and Coworkers.[180] The OPLS mathematical functions are used to describe the atomic and molecular interactions in larger systems (protein and polymers). The torsional and nonbonded parameters of the force field were originally obtained from an *ab initio* molecular orbital calculation at the 6-31G* level. Initial parameterization involved 34 pure organic solvents (including: alkanes, alkenes, alcohols, ethers, amides, acetals, sulfides, disulfides, thiols, aldehydes, and ketones). These solvents were used to mimic the amino acids since peptide residues have organic functional groups. Rotational energy profiles from the *ab initio* calculation were then fitted to obtain the torsional parameters. Bond charge correction factors were also applied to all charge distributions (lone pair and sigma hole-charge) in aqueous solutions to obtain a high level of accuracy. On the other hand, bond stretching and bending motions were adopted from the AMBER (Assisted Model Building with Energy Refinement) [181] force fields. The overall OPLS function is similar to other potential functions, AMBER, and CHARMM (Chemistry at HARvard Macromolecular Mechanics)[182] force fields.

Hence, the first OPLS force field was a combined AMBER/OPLS[180] which was later updated to OPSL-AA. The OPLS-AA (all-atom)[183] force field. Subsequently, The OPLS-AA force field was improved into newer versions for diverse applications; (i) OPLS-AA/L(or OPLS-2001), reparametrized OPLS-AA for peptides and proteins.[184] (ii) OPLS_2005, an improvement of OPLS-AA/L to provide a larger coverage for organic functional groups.[185] (iii) OPLS2, to predict absolute solvation energies for diverse small molecules.[186] (iv) OPLS3[187] and the more recent the OPLS-UA (united atom) proposed by Patsahan *et al*.[188] to model curcumin molecule in water, methanol, and dimethyl sulfoxide solvents. The OPLS updates cover a broad range of molecules, decrease computation time, and sufficiently reproduce experimental data. The Maestro 11 software uses the new generation OPLS3 force field. This newer version is capable of fitting torsional parameters better than other small molecule force fields.[187]
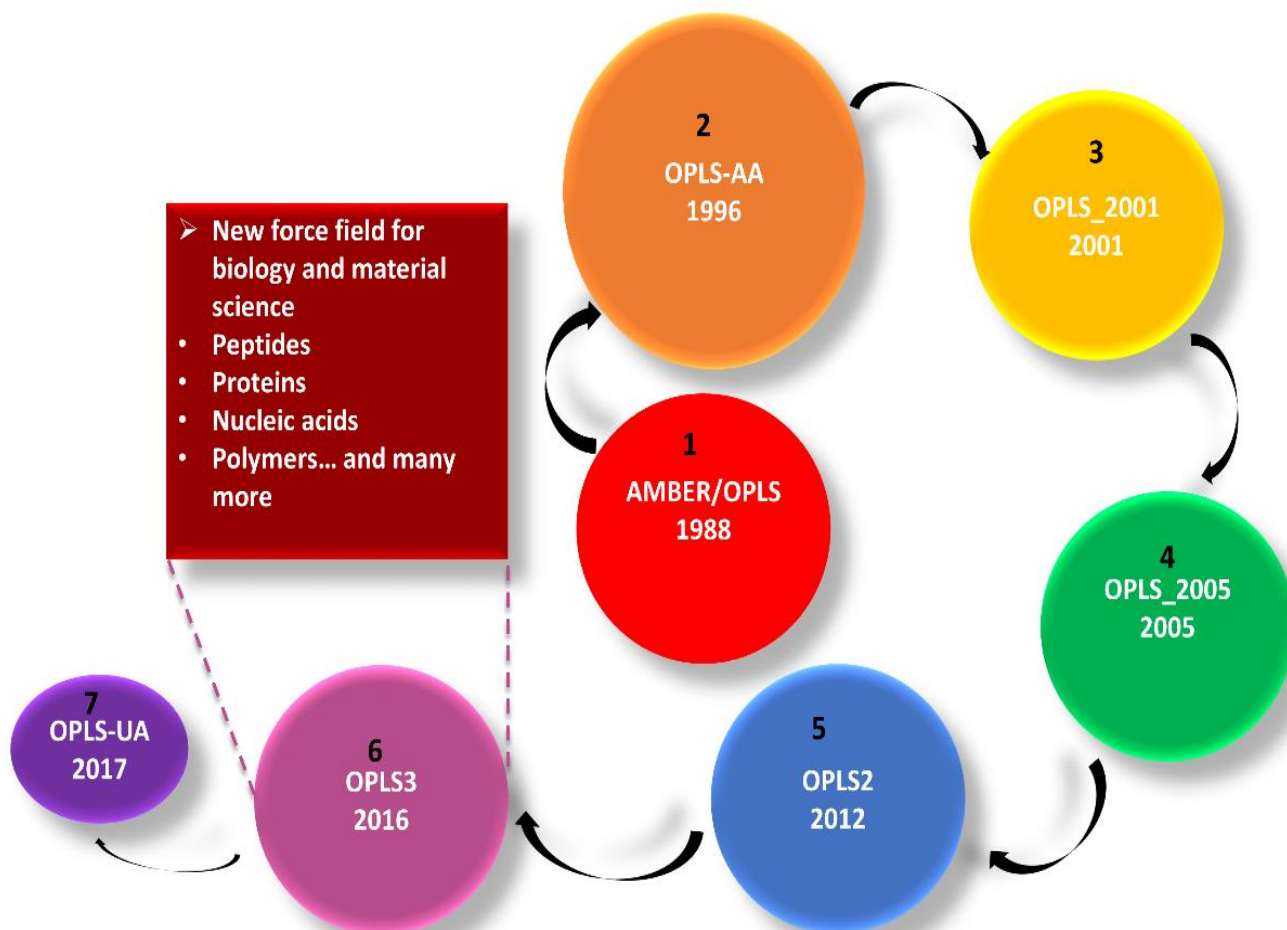
Figure 4.6. The evolution of the OPLS force field. Through the AMBER/OPLS to the recent OPLS-UA. All of the updates stem from the OPLS-AA, the largest sphere (orange). The Figure was generated in Microsoft PowerPoint.

4.5. Methodology

This section describes how the various types of pharmacophores are generated for the Bfr cage protein (PDB ID; 1Bfr) and how the screening process is undertaken. The computational workflow is as follows. A hypothetical pharmacophore was produced with the Phase program for virtual screening. Glide docking was performed for the ligands obtained from the Phase screening. Finally, IFD (not in the initial workflow) was to obtain the computed binding energies between the ligand and the receptor molecule, and they were used to categorize the identified ligands (ligand molecules) into strong binders and weak binders. The Figure 4.7 below shows the various steps leading to the hypothesis.
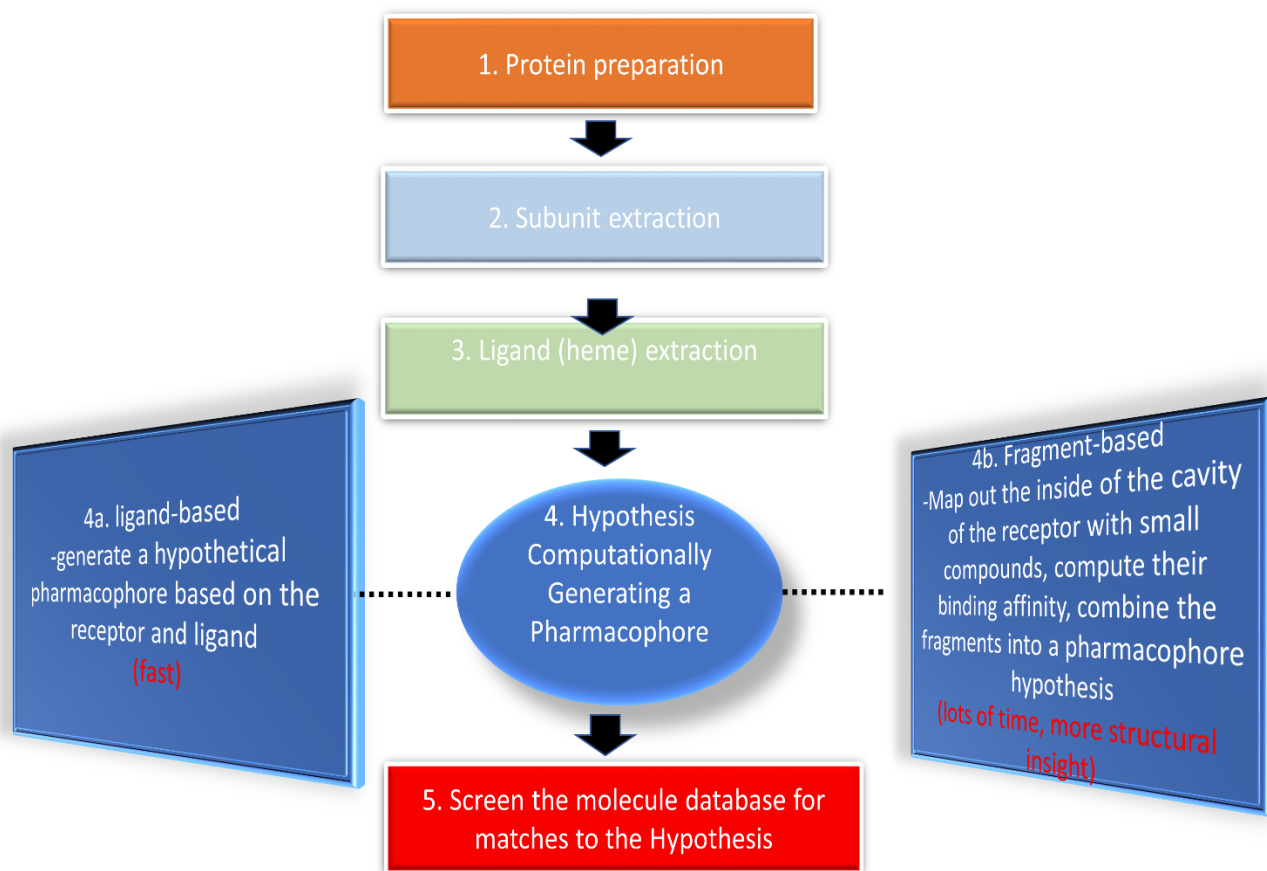
Figure 4.7. Initial computational workflow.

4.5.1. Protein Preparation

The *E. coli* bacterioferritin crystal structure (PDB: 1BFR) reported by Daurant and Coworkers [84] was imported with the Maestro protein preparation wizard tool. The native 24-mer structure was imported by checking the biological unit box. (Figure 4.8) The imported structure was loaded into the workspace, colored by PDB conversions to show areas that required attention (very high B factor regions with missing side-chain atoms and metal atoms). Bond orders were also assigned, but zero-order bonds were created to the heme Fe^{2+} ion. All water molecules that were within 5 angstroms from hetero groups were deleted. Steric clashes were removed with a brief relaxation in the OPLS3 force field with the Impact refinement module. Finally, hydrogens were added to all of the atoms in the system. The hydrogen network was optimized, and the overall structure was minimized.
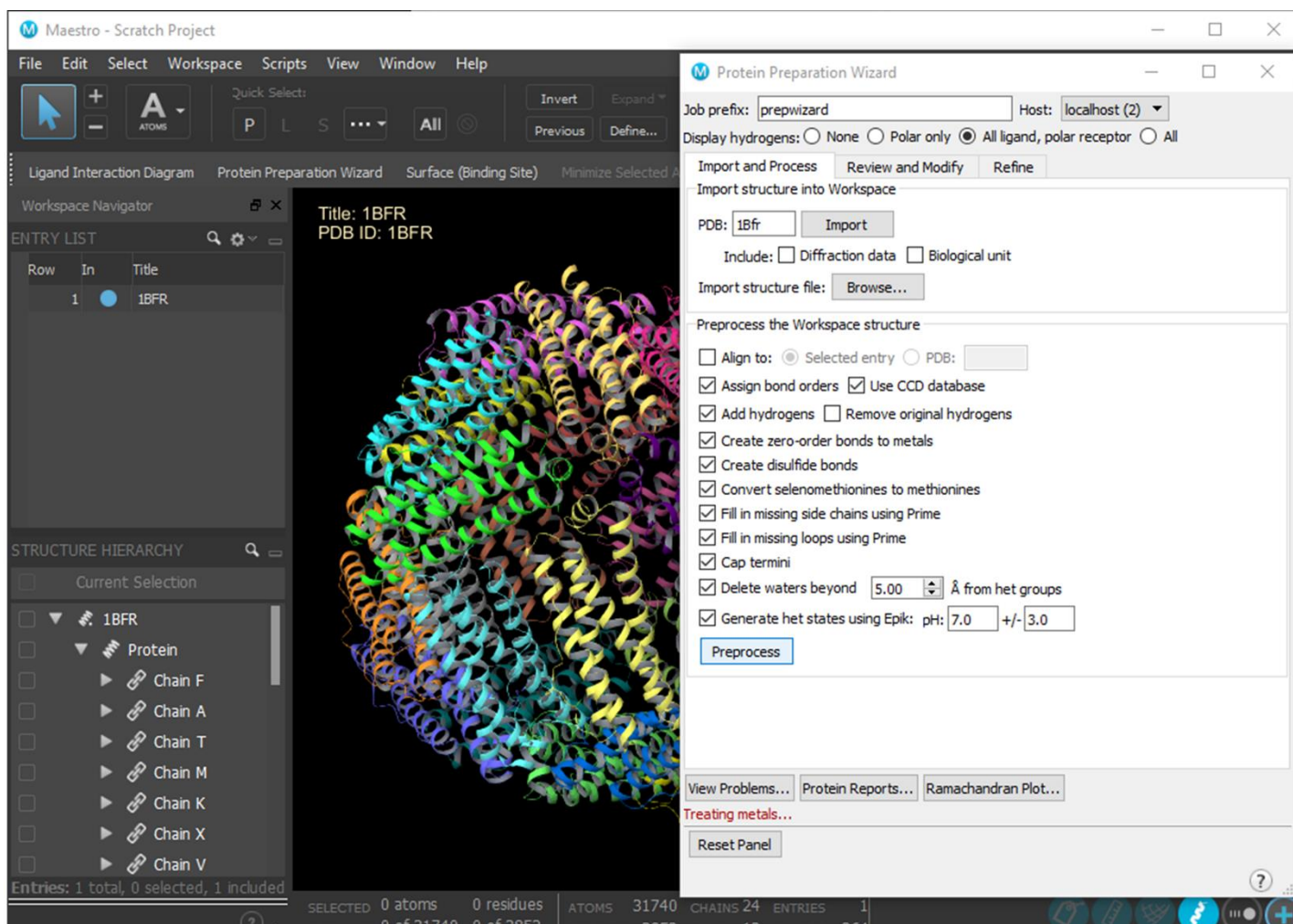
Figure 4.8. Optimizing the 1Bfr crystal structure for modeling. The Maestro Prep wizard on the right was used to import the 1Bfr biological unit and prepared with the options selected in the window. This Figure is a screen capture of the Maestro 11 Protein Preparation.

4.5.2. Subunit and Ligand Extraction

The Review and Modify tab (in the protein Prep wizard) was used to strip down the 1Bfr structure to two adjacent subunits (G and H) and their shared heme cofactor. The Fe atom associated with the heme ligand was also removed (to remove the bias towards the Fe^{2+}-methionine sulfur interaction), and the Mn^{2+} ions at the ferroxidase site were deleted to prevent active site blockage. The co-crystallized ligand (protoporphyrin IX) was extracted from the minimized subunit and labeled as Ligand, and the remaining subunit was labeled Bfr receptor. The ligand extraction was performed by selecting the molecules of interest and right-clicking and selecting Extract. The extracted structures then appeared as separate entries, and they were labeled as "heme ligand" (protoporphyrin IX) and "Bfr receptor" (only subunit, no heme). The process of subunit extraction is illustrated in Figure 4.9 below.
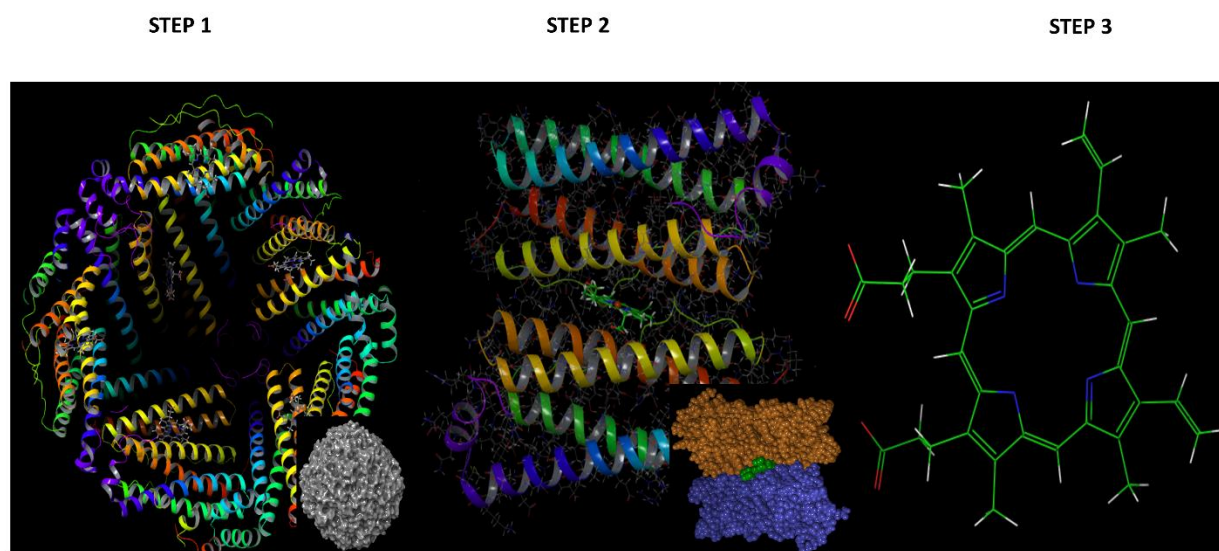
Figure 4.9. Bfr subunit and ligand extraction. In step 1, the optimized structure was loaded, and 22 chains were deleted leaving only chains G and H. The 2 chains were the labeled Bfr and the heavy atoms (Mn^{2+} and Fe^{2+}) removed in step 2. Finally, the protoporphyrin ix molecule was extracted and labeled as a ligand in step 3.

4.6. Hypothesis Generation

The Schrodinger Phase program [168] was used to generate a 3D mold (Hypothesis) of the interactive sites on the receptor. The atoms from the protein residues forming the binding cavity were treated as a set of atomic spheres with sizes determined by their van der Waals radius. Excluded volumes were added to the hypothesis to allow small overlaps between the ligand and the receptor, creating a kind of induced-fit effect. The excluded volume spheres for the receptor atom were positioned within a 5 Å shell around the ligand surface while ignoring all receptor atoms whose surfaces were within 2 Å or closer to the docked ligand surface. This restraint creates a loose excluded volume that forces the matched ligand to stay in the binding site most of the time. Two different hypotheses were generated, the first from the Ligand-based (using both the receptor cavity and heme ligand) and the Fragment-based which employed 667 fragment molecules to create the interactive sites in the receptor cavity (no ligand involved).

4.6.1. Ligand-based Hypothesis Generation

The Bfr dimer with a heme cofactor at the interface was imported into Maestro. The Phase application was selected from the task tab, and the "generate a hypothesis from a receptor and an active ligand" option was selected. The ligand was defined by clicking on an atom in the heme ligand in the workspace and the parameters described above were set as well as the maximum number of interactive sites. The generated pharmacophore was listed as "hypothesis" in the entry list and also appeared in the workspace. The seven interactive sites and their corresponding interaction energies were labeled in the hypothesis (Figure 4.9.1a)

4.6.2. Fragment-based Hypothesis Generation

The second hypothesis (Hypothesis1: Epharmacophore-Receptor cavity) was obtained by selecting Phase from the task menu and subsequently selecting: Develop a hypothesis; Receptor cavity in the workspace. This approach employs small fragment molecules (Figure 4.9.1d) to map out the various interactions in the active site. Before that, the active site region was defined by picking residues (Figure 4.9.1 c), which are expected to interact with the active ligands. The centroid of the cavity was set to x = -2.25, y = -27.83, z= 29.17, to ensure that the fragment molecules are confined to the center of the active site (marked in red, Figure 4.9.1 b). The number

of interactive sites was limited to 7 and run on 2 of 8 processors, which took about 4 hours to complete.

Chain	Residues
H	Phe 26, Lys 53, Met 52, Ala 55, Ile 49, Asp 56, lys 2, Leu 71, Glu 60, Ile 59, Asn 23, Lys 33, Leu 74, Glu 72, Leu 24, Asp 56, Tyr 45
G	Asp 56, Arg 30, lys 53, Ile 49, Met 52, Asn 23, leu 19, Phe 26, Leu 71, Leu 15, Ile 22, Ala 55, Asp 73, Ile 59, Leu 73
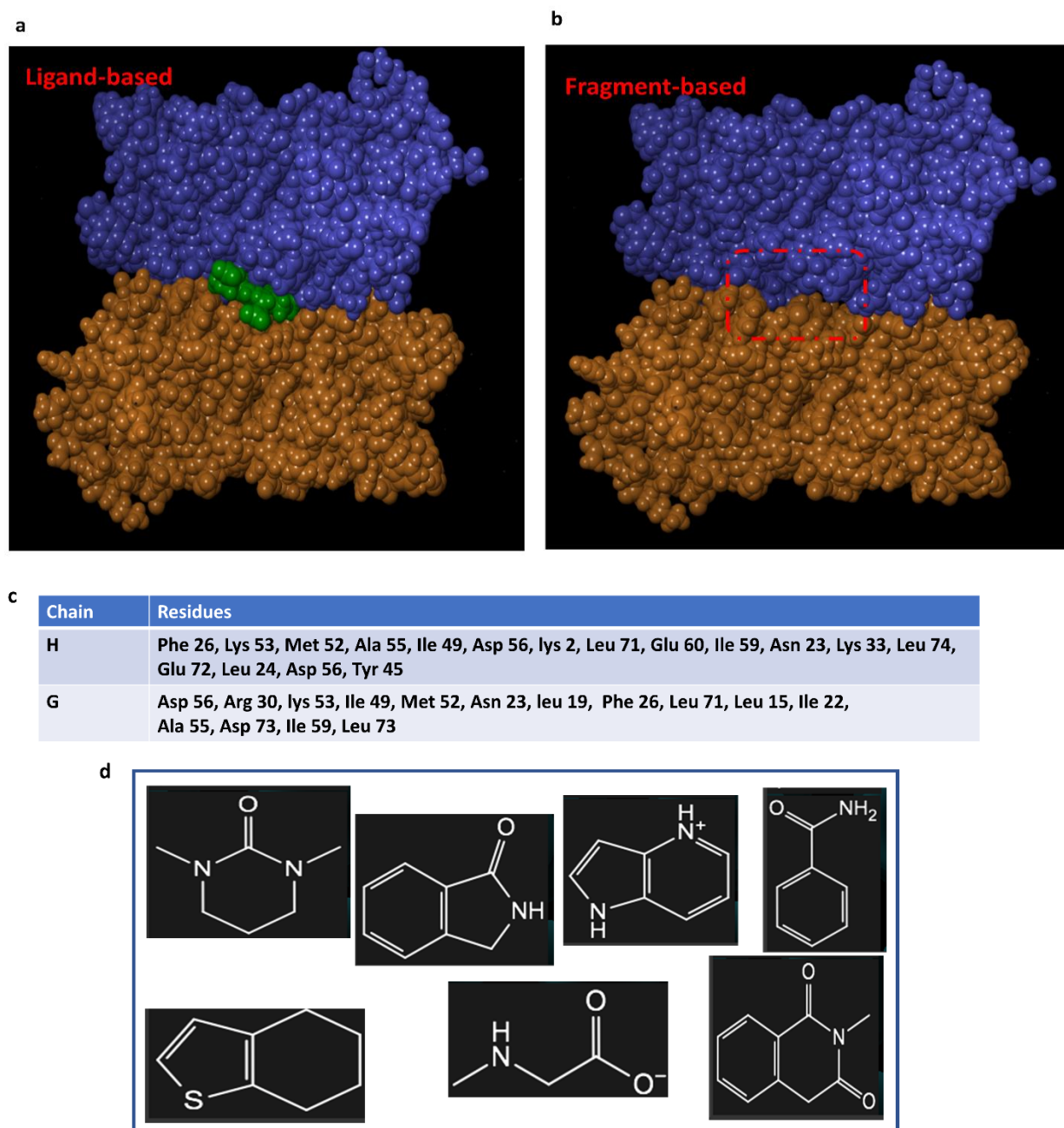
Figure 4.9.1. Generating Bfr pharmacophore. (a) Shows the dimer (receptor) subunit with its reference heme ligand (green) for the Ligand-based hypothesis (b) The receptor without a reference ligand showing the centroid where the fragments must dock. (c) A table of all the residues in the highlighted region of panel b. (d) examples of the 667 fragment molecules used to mold the interaction at the receptor active site.

4.7. eMolecule Screening

The Phase screen is a rapid but less refined strategy to virtually screen millions of molecules from a database or imported ligands to obtain matches (hits) for the generated hypothesis. The Molecules are sorted by the number of interactive sites, volume score, and their overall effective interactions. The databases used in this work were the CAP2011 and the Aldrich Market Direct®. The CAP2011 contains about 2.3 million (up to 4.8 million including all conformers) ligand molecules with average molecular weight not higher than 650 Da. Similarly, the Aldrich Market Direct® contains 7.3 million (up to 19.1 million conformers) ligand molecules with molecular weights between 250 to 500 Da.

Both of the databases were separately screened with two different hypotheses (ligand-based and from fragment-based). Each of the hypotheses contained seven interactive sites, and one of the conditions of the hypothesis screening was that a screened molecule must match at least four out of the seven interactive sites of the pharmacophore before it can be classified as a hit. The job was divided between 7 processors which took up to 2 weeks to successfully screen the entire CAP2011 database and up to one month for the Aldrich Market Direct® database. The top 500 hits were retained for each screen (fragment-based and ligand-based) and incorporated into the Entry list as well as the project table. The entry list showed a specific hypothesis with its corresponding 500 ligand candidates arranged in the order of decreasing Phase score.

4.8. Docking

The Phase screen hits were docked into the Bfr receptor molecule using the ligand docking protocols in Maestro. Figure 5 outlines the various steps in the docking process.
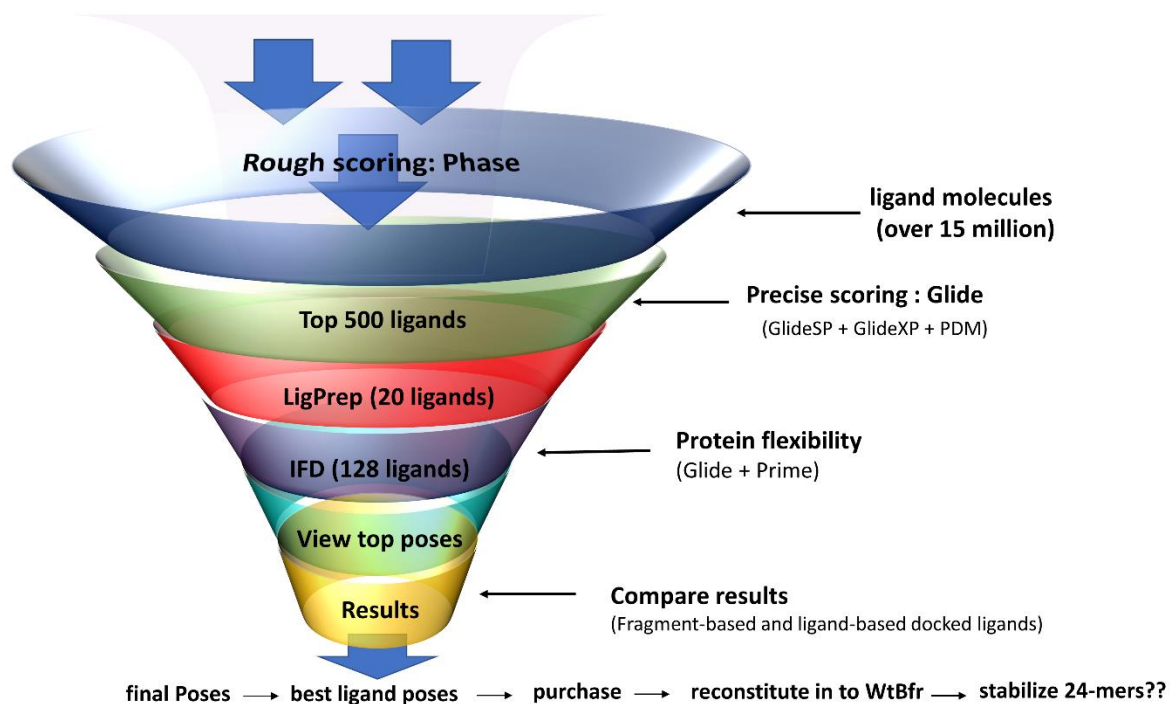
Figure 5. Docking funnel. Summarizes the process involved in the docking process. The Docking funnel was designed in Microsoft PowerPoint.

Glide SP and XP dockings were performed on the hits to obtain a set of molecules which optimized their fit into the Bfr receptor site. The 500 phase ligands were selected for a quick Standard Precision glide (Glide SP) docking, followed by an Extra Precision glide. A receptor grid was generated using the procedure outlined in the Schrödinger Academy Maestro 11 Tutorials entitled, "Structure-based Virtual Screening using Glide." [189] The receptor grid was made by loading the ligand (protoporphyrin IX) and receptor molecule (Bfr dimer-nonheme) into the workspace and specifying which molecules should be treated as ligand and receptor in the Glide application. This process creates green and purple boxes around the ligand and receptor, respectively (Figure 5.1). The receptor box defines the outermost regions that a docked molecule can occupy to pass the initial docking stage, whereas the green ligand box defines the centroid of

the docked molecule (where it must be positioned to pass the initial docking stages. The X, Y and Z coordinates of the ligand box were adjusted to 10 Å, 8 Å and 6 Å respectively, to ensure that the size is not dependent on the size of the docked ligand, but well positioned in the center of the grid. The van der Waal (vdW) radii of the receptor were scaled by a factor of 0.8 to partially account for suboptimal fits that could be associated with minor receptor movements. Also, highly flexible receptor sidechains were frozen in this process to allow many ligands to dock within a reasonable time. A constraint was set on the receptor to enable free rotation of rotatable groups such as serines and tyrosines. Several ligand conformations were docked into a fixed receptor conformation. However, only one conformation was retained after the post-docking minimization step. The same receptor was used to dock all the molecules from the two pharmacophore hypotheses. Each molecule was docked within 10.8 seconds in Glide SP. The SP docked ligands were further used for Glide XP (same receptor conformation) which took about 6.8 minutes per ligand on 7 processors. The output files were added to the project table and the Entry list in order of decreasing glide scores. The ligand poses were visualized in the workspace by loading the ligands into the workspace with a fixed receptor molecule.
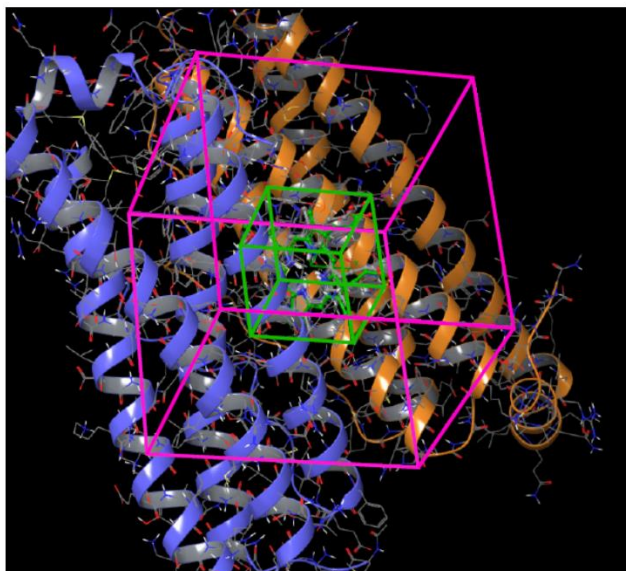
Figure 5.1 The Bfr grid. The green area defines the centroid of the docked ligand while the purple dox defines the region where a ligand must occupy to pass the initial docking stage. Screen captured from Maestro 11.

4.9. Induced Fit Docking (IFD)

The Induced FIT docking was run selecting the top 15 tight binders, 3 moderate binders, and 2 weak binders from Glide XP, using the Market select database (Schrodinger, LLC). The 20 ligands were prepared employing LigPrep to generate each ligand's corresponding low energy 3D structure; overall 128 structures were obtained and used for IFD using the flow chart outlined by Sherman *et al*. (Figure 5.2)[173]
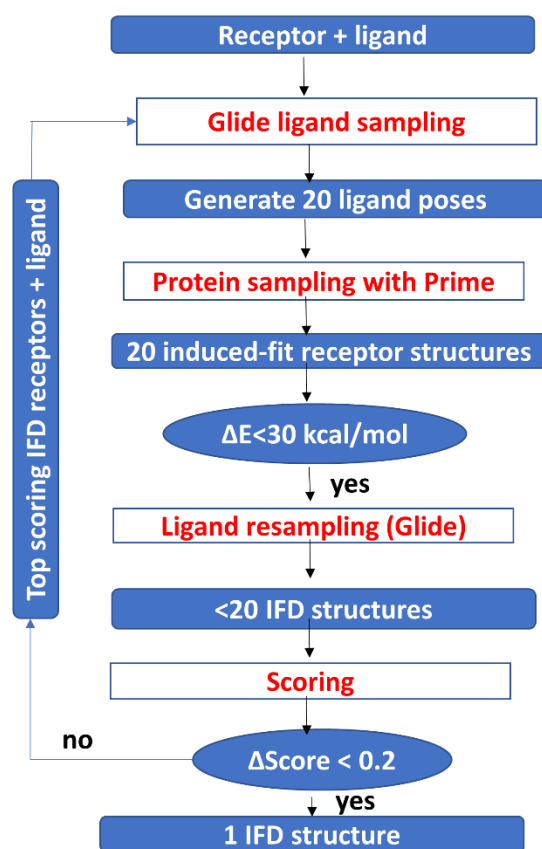
Figure 5.2. The induced fit docking protocol. A flow chart showing how both Glide and Prime are used to account for protein flexibility, as demonstrated by Sherman *et al*.[173]

5. RESULTS AND DISCUSSION

5.1. Protein Preparation and Optimization

The starting point of any computational calculation is to obtain an optimized structure that is suitable for the application since the simulated result is dependent on the initial structure. The ability to fill in missing parts of the structure is important for protein X-ray crystal structures (which focuses more on heavy atoms than smaller ones). Several protein crystal structures have missing atoms in flexible regions that tend to fragment the structure, making it difficult to work with computationally. However, the Protein Preparation wizard has proven to be an indispensable tool for this task. The entire Bfr structure and its environment were modified with this tool. The Bfr subunit was extracted from the native 24-mer as well as unwanted solvent molecules were removed from the crystal structure. The structure was successfully optimized and minimized with the Protein Preparation Wizard. The extracted ligand and receptor were further used in modeling processes outlined above. The Figures below show the structure of 1Bfr going from an intact 24-mer to a dimer (heme-bound) and the ligand. These applications show how powerful the OPLS3 force field sufficiently covers large molecules (over 31 thousand atoms in Bfr) without any errors or software issues, as seen in Figure 4.9.

5.1.1. Ligand Preparation

Similarly, the 2D ligand structures obtained from the Glide rigid receptor docking were successfully optimized to generate the correct ionization states, tautomeric states, and the corresponding low energy structures were used for IFD-based docking. Each ligand generated a maximum of 32 ligand conformers/isomers, but not all ligands had multiple conformers and isomers. A total of 128 ligands (3D) were generated from the 20 original Glide ligands (2D).

5.2. Ligand-based Pharmacophore Modeling (Phase Screen)

The ligand-based pharmacophore approach generated a possible spatial arrangement of the most energetically favorable sites necessary for ligand recognition. The arrangement was based on the structures of the ligand and the receptor and their interaction. Even though this approach of pharmacophore modeling is thought of as biased (as an already known ligand is used as the important mapping component), it provides some meaningful insight into the structure and

170

binding properties. The generated pharmacophore had 7 interactive sites (Figure 5.3). Four of them are from the aromatic groups associated with the 4 pyrrole rings (R13-R16) with equal theoretical binding energies (-2.84 kcal/mol each). Two hydrophobic interactions related to the ethylene groups on R14 and R16 (-0.6 kcal/mol and -0.43 kcal/ mol, respectively). Lastly, an ionic interaction (N12, -0.36 kcal/mol) associated with one of the propionate groups that is observed to interact with Lysine 53 in the receptor (Figure 5.5b). The analysis of these results showed that the 4 aromatic groups appear to be the most important interacting groups, followed by the two hydrophobic groups and the single negative charge interaction.

The 5 top-scoring hits obtained from the Phase screen were structurally similar to protoporphyrin IX but had different substituents at sites H7 and H8. The top 4 hits obtained (Figure 5.3.1) validate the generated hypothesis since the virtual screen was able to find molecules that were structurally similar to the reference ligand. Overall, Phase screen is a rapid approach that limits the number of molecules to work with, but there is the need for an extensive docking to eliminate false ligand candidates.
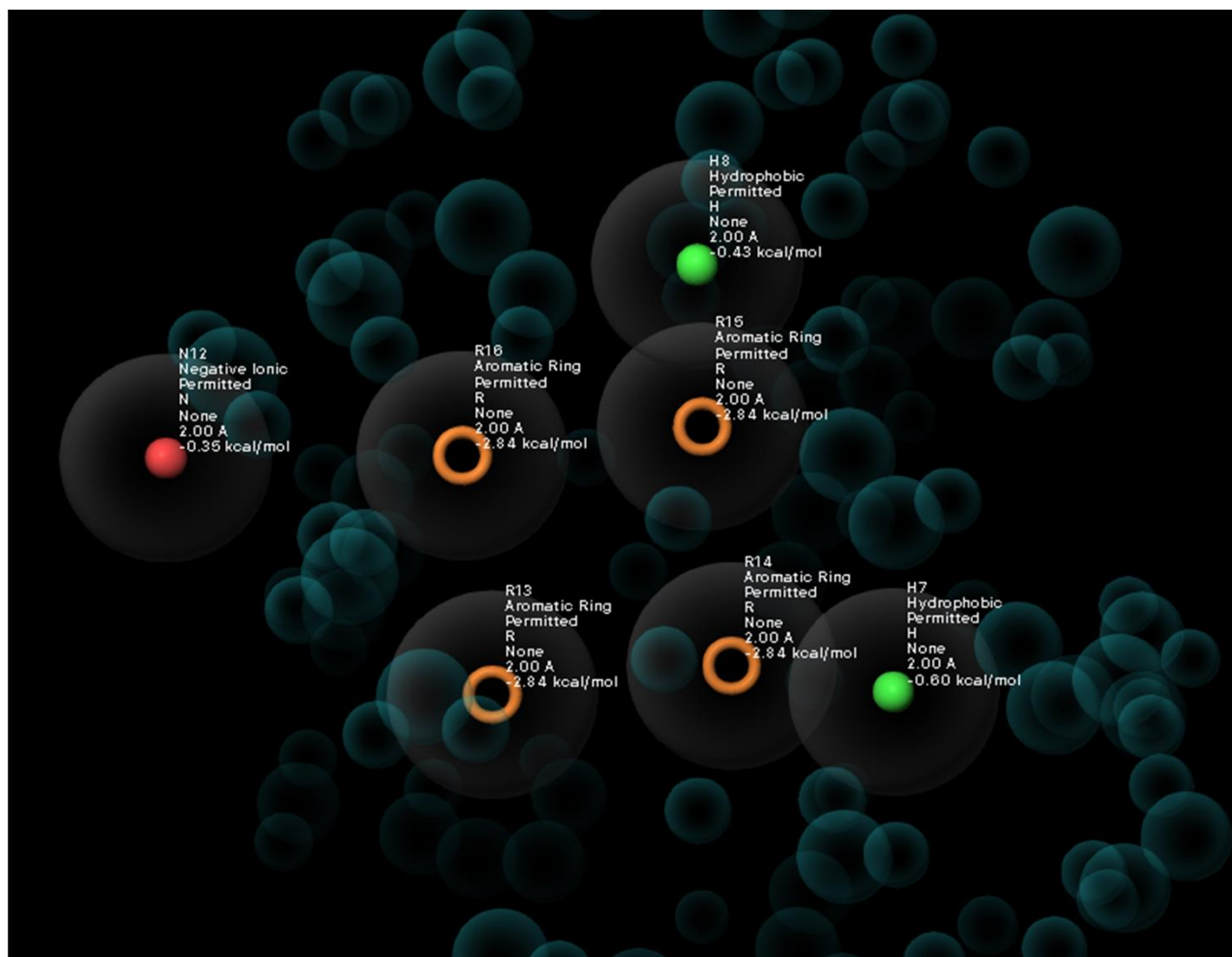
Figure 5.3. The Bfr Pharmacophore (Ligand-based). The teal spheres show the excluded volumes, indicating that a ligand could experience steric clashes (from receptor molecules) within those regions. The Orange color represents aromatic groups, red for ionic groups and green for hydrophobic groups. Image was exported from Maestro 11.
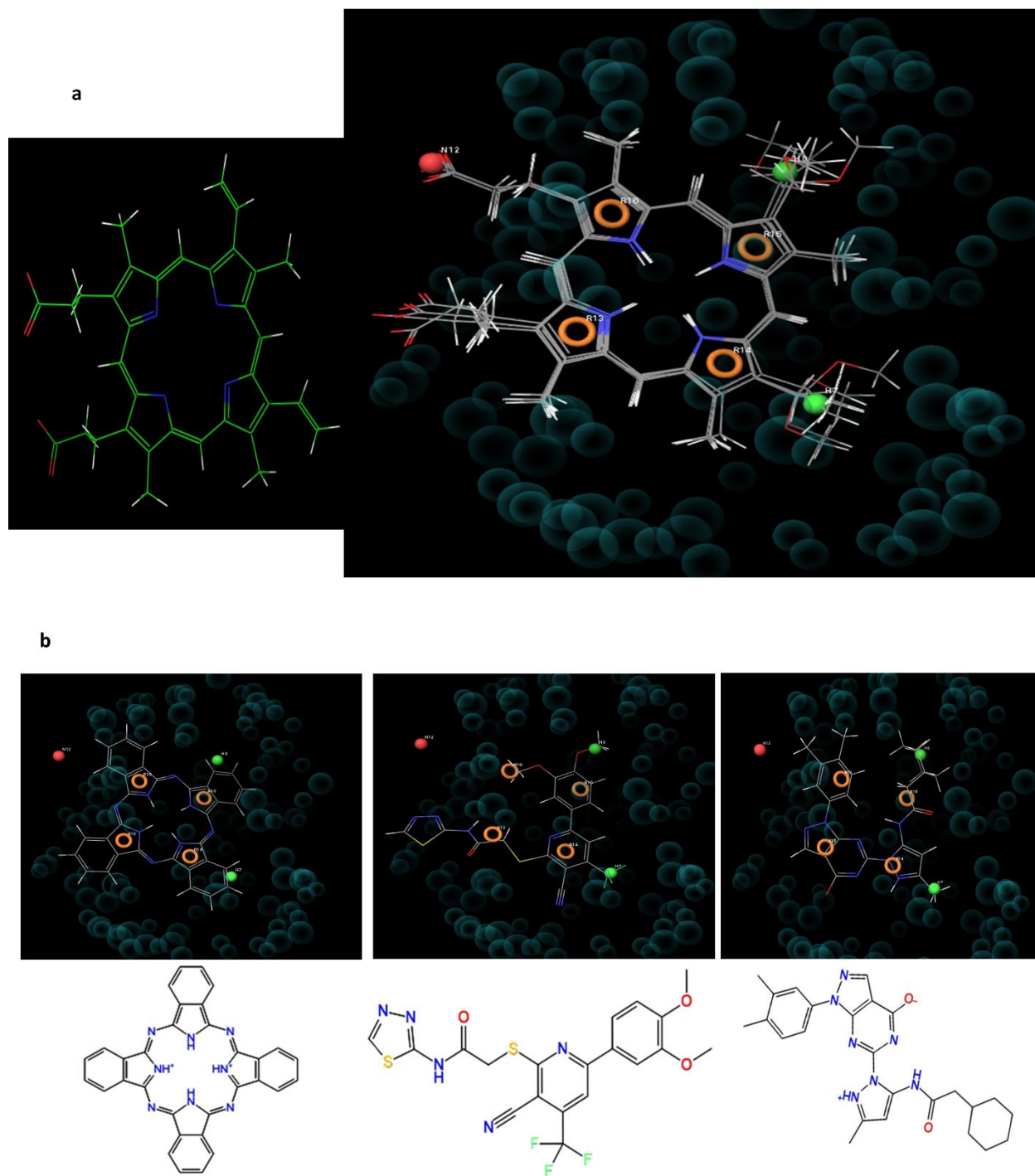
Figure 5.3.1. Phase Screen hits. (a) An overlay of the top 4 hits from Phase Screen with the heme ligand on the left. (b) Other interesting hits obtained and their respective structures. All images were obtained from the Maestro 11 application.

5.3. Fragment-based Pharmacophore Modeling (Phase)

The fragment-based approach employed 667 small molecules which eventually resulted in the identification of seven small fragments (Figure 4.9.1. c) of importance to map out the defined binding site (Figure 4.9.1 b), which resulted in 7 essential interactions (Figure 5.3.2). Three out of seven were aromatic regions (R 19, R 21, R 22), two hydrophobic sites (H 13, H14) and two Donor sites. All the sites allow interaction within 2 Å distance. The fragment-based approach ranks by distances. The top 4 Phase screen molecules had a score of about 1.7 (Figure 5.3.3), which is similar to the scores obtained for the non-protoporphyrin molecules from the ligand-based approach indicating a valid hypothesis.
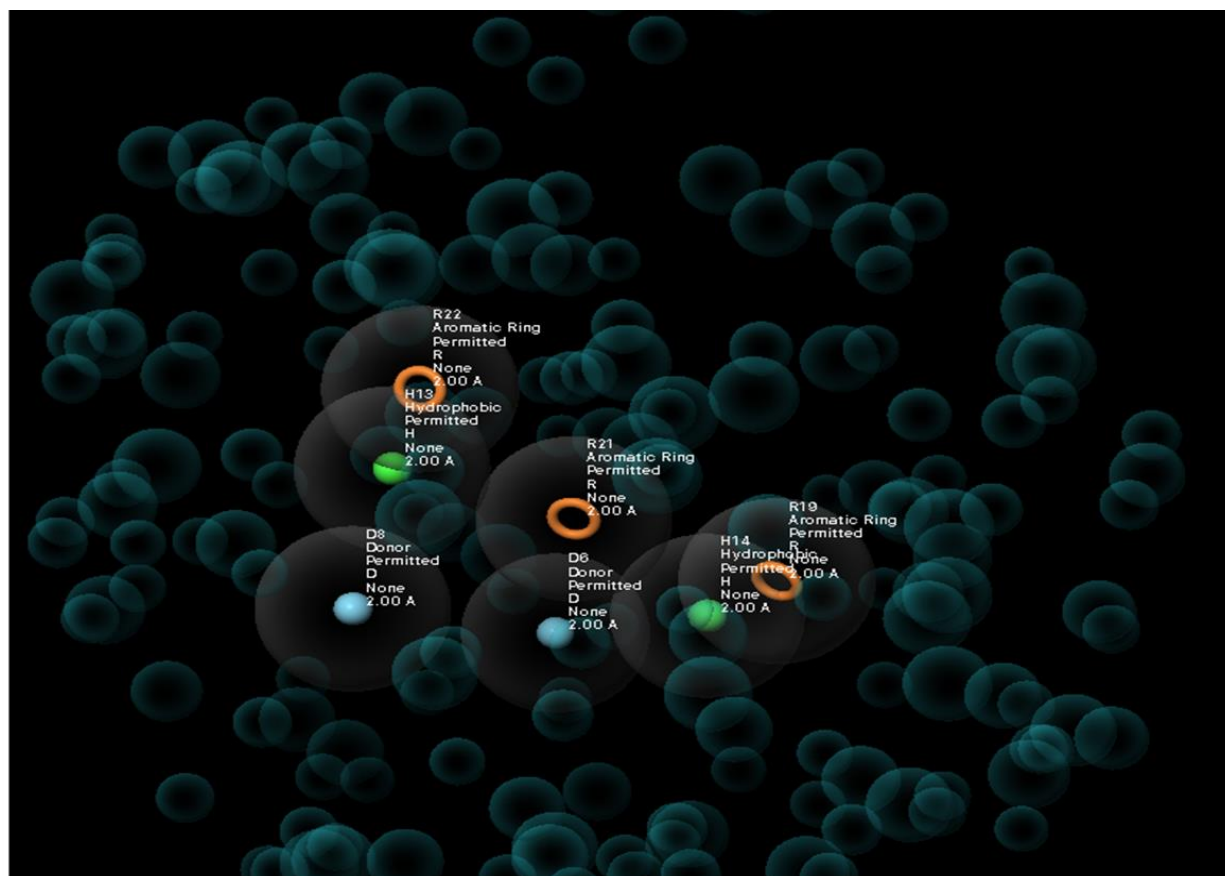
Figure 5.3.2. Fragment-based hypothesis. The fragment-based ePharmacophore showing 7 sites: 3 aromatics (orange), 2 donor sites (light), and 2 hydrophobic sites (green). This image was exported from Maestro 11.
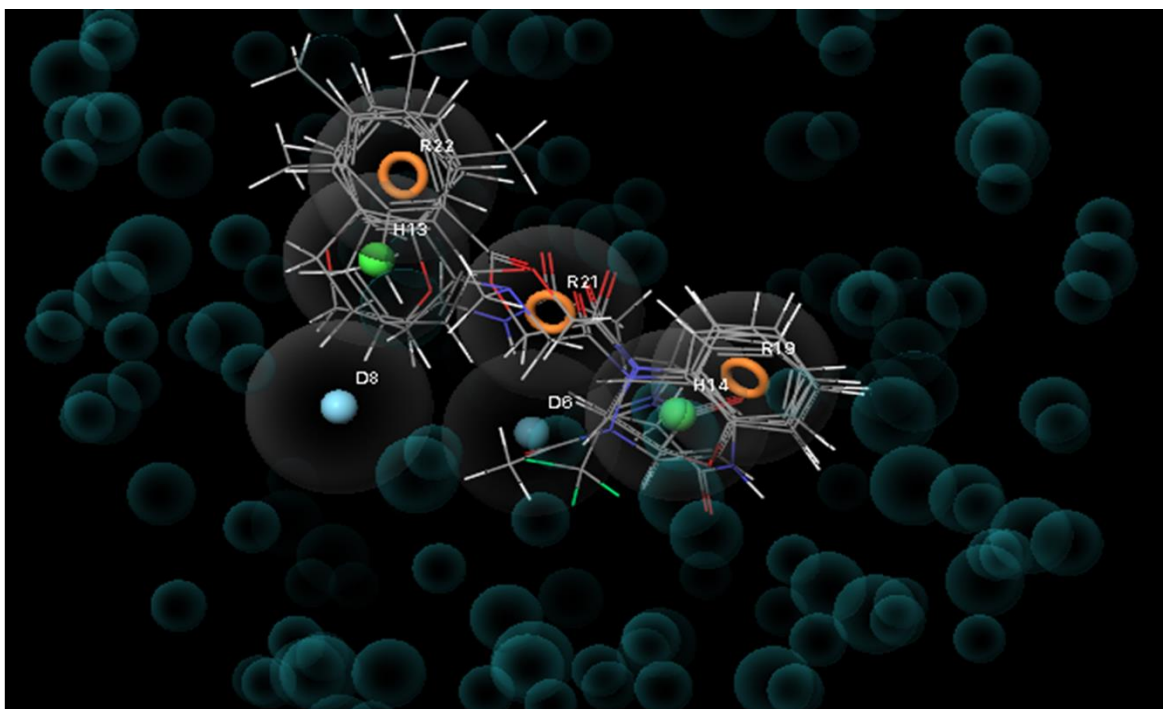
Figure 5.3.3. An overlay of the top 4 fragment-based ligand molecules. Image was exported from the Maestro 11 application.

5.4. Glide XP Docking

Glide docking results obtained from the top 500 fragment-based hits were similar to the ligand-based results, regarding the score and the pose viewer interactions between the receptor molecules and the docked ligand. Met52 was observed to contribute unfavorably to the interaction energies in most cases. Phe26 was quite favorable in both ligand and fragment-based results despite its size it was involved in either pi-pi or pi-cation interactions with most of the ligands. This shows that not all bulky amino acids at the active sites are sterically hindering, this observation agrees with the work reported by Nam and coworkers where they reported Trp 26 at the heme binding site of RsBFr (bacterioferritin from Rhodobacter sphaeroides) to be an essential residue for stabilizing the heme molecule.[63a] The most favorable residue at the binding site is Asn 23 which is involved in hydrogen bonding (donor/acceptor) with the docked molecules.

Glide SP and XP docking applications were used as a more precise approach to sort the Phase screen ligands. The docking is used to eliminate false ligands which may have successfully passed the initial Phase screen process. The ligands were individually docked to the Bfr receptor grid in rigid receptor approach within a reasonable amount of time. The docking process begins by generating several conformers of the Phase screen ligand and roughly scoring them to obtain promising ligand poses (initial docking stage). The promising ligand poses are further refined in torsional space in the field of the receptor molecule with a distance-dependent dielectric model to a fewer number of poses which are finally minimized with full ligand flexibility (Post Docking minimization, PDM).[190] The 0.8 vdW scaling factor was used to soften the active site potential and compensate for some receptor flexibility, as well as to generate a diverse ensemble of poses. A constraint was set in the grid to allow rotation of Ser and Tyr to mimic aqueous conditions. Also, the docking process generated tautomeric forms of the neutral ligand molecules using the Epik (Empirical pKa prediction) application.[191] Highly flexible side chains with poor interactions were frozen to minimize computation time.

5.5. Glide Ligand Poses

A ligand pose shows how the ligand interacts with the receptor molecules in 3D. Several ligands were docked into the receptor, and the top pose (from Glide XP using the CAP2011) is seen in Figure 5.5a. A typical pose shows how favorable or unfavorable a ligand-receptor interaction is. The residues with favorable interactions are colored blue, whereas unfavorable interactions are red. This is very useful in that the preferred ligand structure could be fine-tuned to reduce steric clashes with the Bfr receptor molecule. Met52 was observed as unfavorable for all the non-protoporphyrin IX ligands (Figure 5.3.4, molecules 1, 6-10). The unfavorability may arise from the sterics associated with the side chains. A previous study in the Honek laboratory reported the formation of a 24-mer in Bfr constructs with Met52 mutation (bFTM52L).[122] Even though it been reported that the heme-Met52 ligation is essential for Bfr oligomerization, [101] the presence of intersubunit salt bridges, hydrophobic interactions, and other tertiary interactions could stabilize the Bfr 24-mer similar to other non-heme ferritins. Lysine 53 was also observed as another unfavorable residue in most cases even when it interacted with the propionate groups in the protoporphyrin-like ligands. This supports the low interaction energy obtained from the negative site N12 (Figure 5.3.5, molecules 2-5). This shows that the formation of a salt bridge

between the receptor and the ligand molecule may not be very critical to binding affinity and may also explain why ligand 1 (Figure 5.3.4. a) is the best ligand despite the lack of that particular interaction.

The protoporphyrin-like ligands were expected to be the top ligands since they are structurally similar to the reference ligand. This was obvious in the Phase screen results as the top 5 molecules were all identical to that structure (average 2.5 kcal/mol PhaseScore). However, the docking results indicate that resemblance to the reference ligand does not always correspond to a true ligand. Phase screen molecule 5 (Figure 5.3.1. b, the first molecule on the left) was eliminated from the Glide docking due to lack of rotatable groups (zero rotatable groups), this results in high ligand strain penalty (E_{ligand_strain}, eq (3)) on the overall Glide score. Overall the top 9 molecules did not have many unfavorable interactions (red dashed lines) and they have comparable binding affinities with no Epik state penalties (table 2). Epik state penalties for the various tautomer states of a particular ligand is applied to the Glide score. Ligand 10 had a different docking score from the Glide score due to its 0.0106 penalty.

The Glide model score is useful for ranking conformers of the same molecule since it accounts for the effects of the orientation of a particular conformer in the forcefield in terms of its electrostatic and vdW energies.
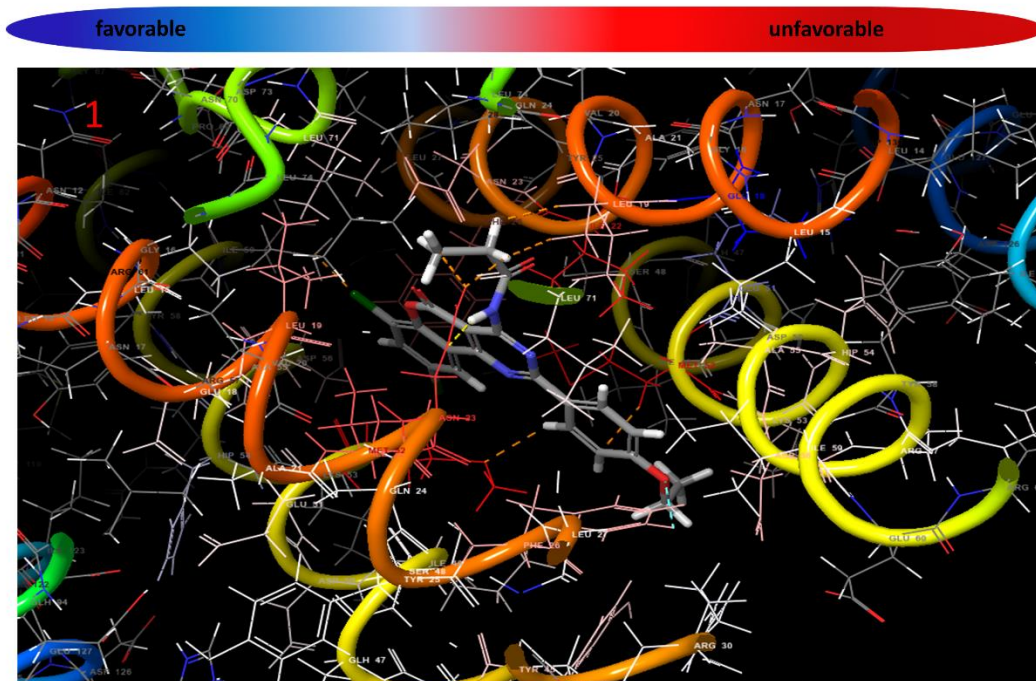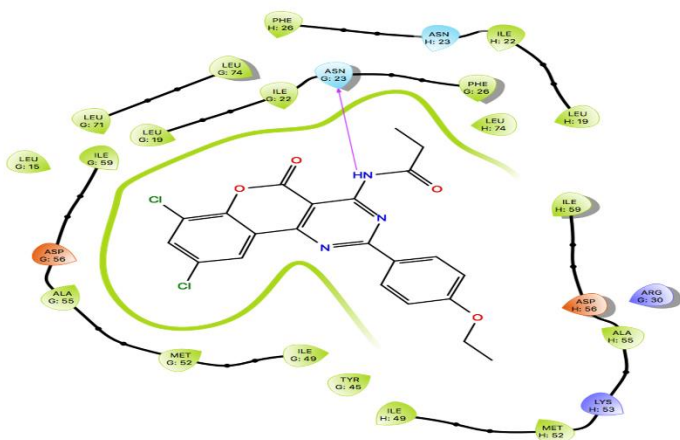
d
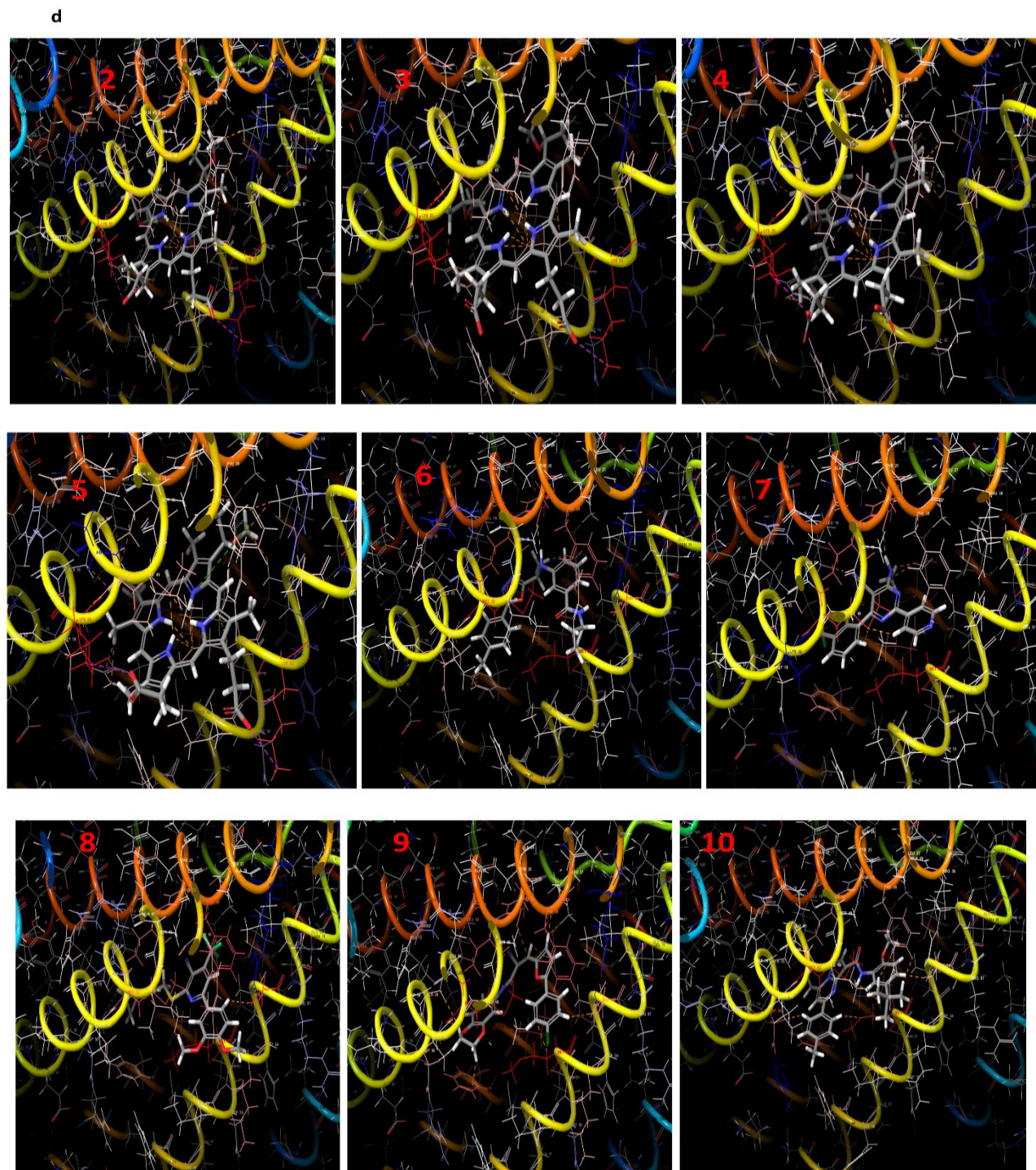
Figure 5.3.4. Glide ligand poses (from the ligand-based screen). (a) The most active ligand pose oriented in the receptor. The receptor backbone in thin tubes; the side-chains are in sticks while the ligand is in the skeletal model. (b) Ligand interaction diagram for the ligand and receptor. (c) Screenshot of the reaction legend. (d) Glide ligand poses of the top 10 higher affinity ligands. All images were exported from the Maestro Glide docking application.
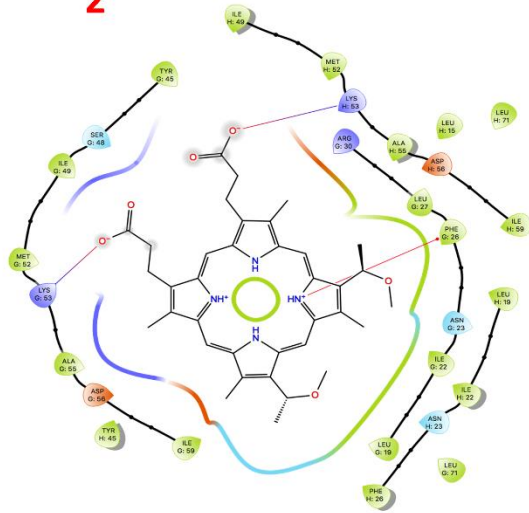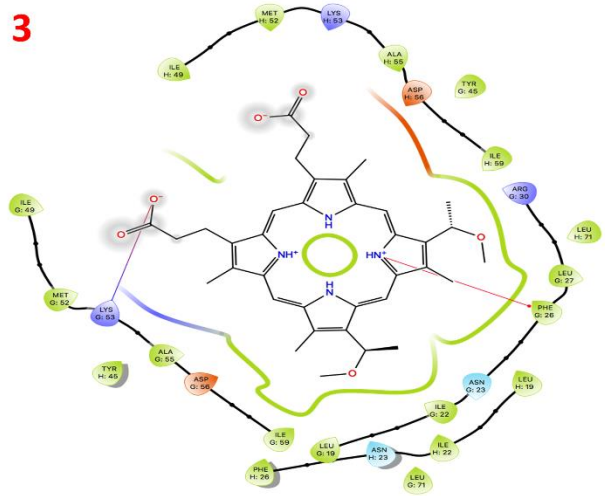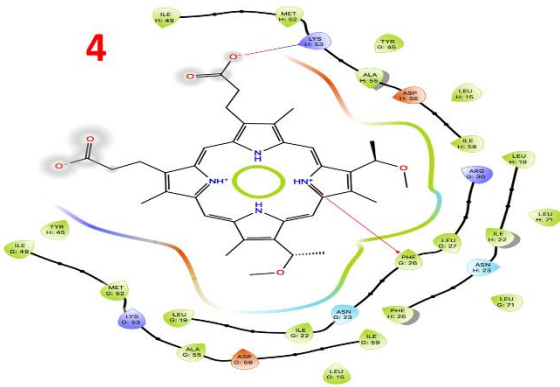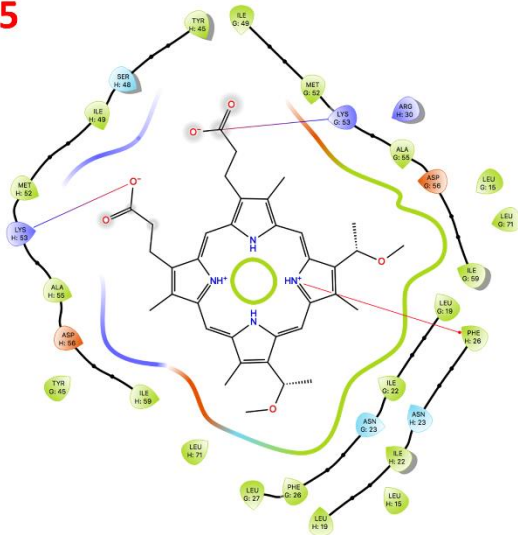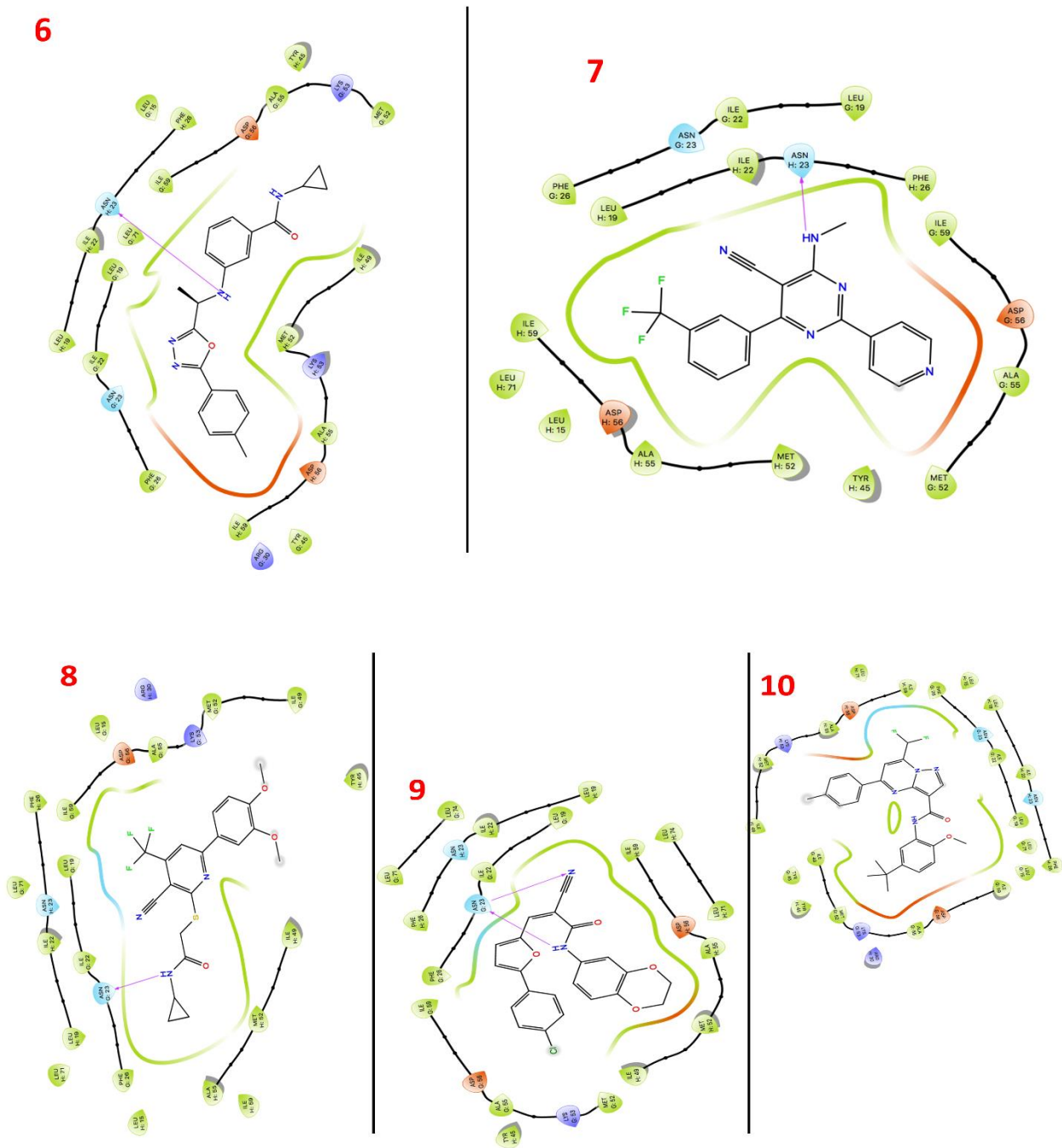
Figure 5.3.5. Glide Ligand interaction diagrams. Shows 2D representation of the interaction between the docked ligand (from ligand-based) and Bfr receptor. All images were exported from the Ligand interaction application.

Title(and rank)	glide rotatable bonds	docking score (kcal/mol)	XP GScore (kcal/mol)	glide gscore (kcal/mol)	Epik State Penalty	Glide Emoel (kcal/mol)	Phase Screen Score	Num Sites Matched	Phase Screen rank	MW
CACPD2011a-0000589618 (1)	6	-14.6154	-14.6154	-14.6154	0	-94.9481	1.831493	4	432	458.2941
CACPD2011a-0001370096 (2)	10	-14.3959	-14.3959	-14.3959	0	-117.346	2.496549	7	2	626.7419
CACPD2011a-0001370096 (3)	10	-14.3679	-14.3679	-14.3679	0	-113.792	2.467321	7	3	626.7419
CACPD2011a-0001370096 (4)	10	-14.2135	-14.2135	-14.2135	0	-111.59	2.707366	7	1	626.7419
CACPD2011a-0001370096 (5)	10	-14.0664	-14.0664	-14.0664	0	-101.966	2.429149	7	4	626.7419
CACPD2011a-0000044676 (6)	7	-13.6725	-13.6725	-13.6725	0	-86.8348	1.840587	5	289	362.425
CACPD2011a-0001151394 (7)	5	-13.6573	-13.6573	-13.6573	0	-76.0154	1.845117	5	222	355.3166
CACPD2011a-0000775268 (8)	10	-13.6566	-13.6566	-13.6566	0	-80.0337	1.841849	4	270	437.4354
CACPD2011a-0001327395 (9)	6	-13.3699	-13.3805	-13.380	0.0106	-86.1236	1.844164	4	240	406.8185
CACPD2011a-0000330615 (10)	7	-13.273	-13.273	-13.273	0	-97.2185	1.868433	4	75	464.507

Table 2. Cumulative results from Phase Screen and Glide using the ligand-based molecules.
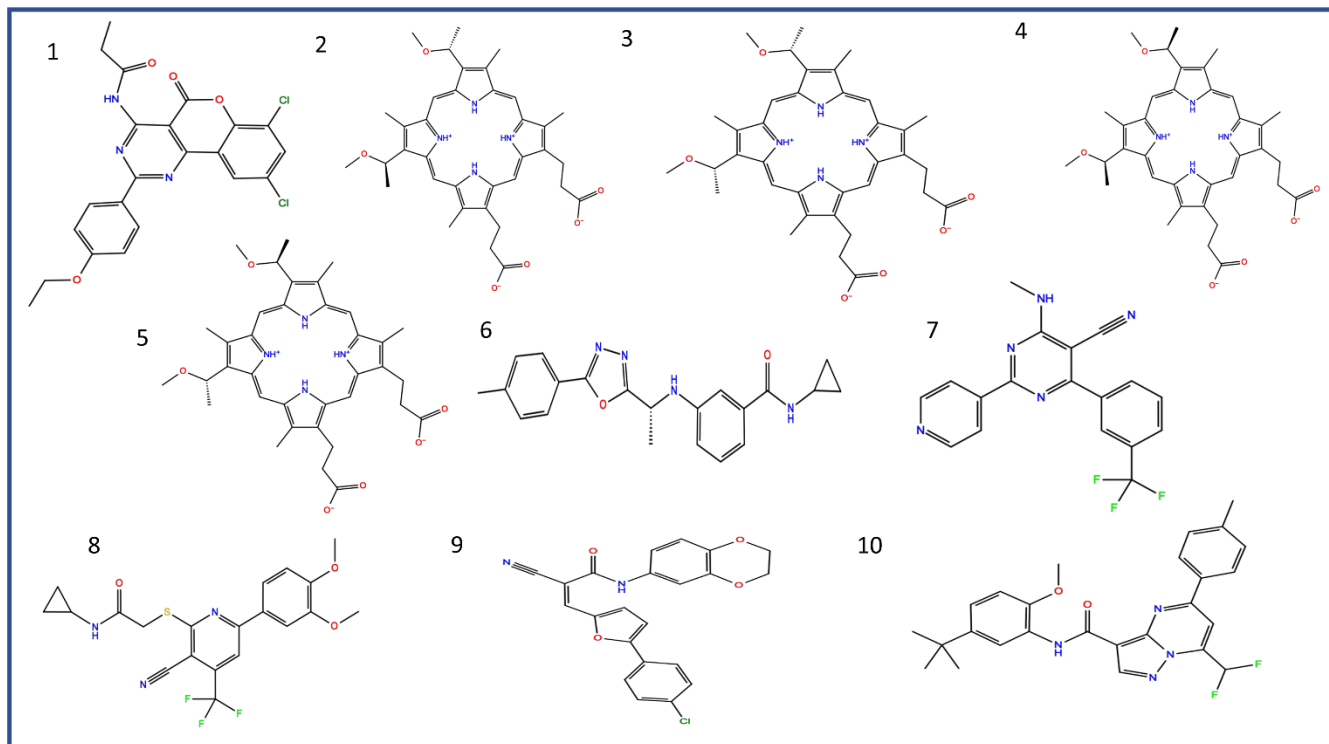
Figure 5.3.6. Top 10 Glide ligand-based structures. The structure which was originally ranked 432 from Phase screen emerged as the most active ligand in Glide.

5.5.1. Glide Ligand Poses from Fragment-based Screen

Similar to the ligand-based molecules, the fragment-based molecules were ranked on the Phase Score, and the top 500 were used in Glide docking to obtain the highest and lowest affinity binding molecules. The results are summarized in Figures 5.3.7. and 5.3.8. and the table 3 below. The ligands obtained here are not structurally similar to the native heme molecule but have similar planarities to heme. The lack of reference ligand produced an active site mold without any similarity to the ligand-based pharmacophore. Hence, this can account for the structural differences between the structures from the two approaches.

An advantage of this approach is that it diversifies the number of ligand molecules which may not resemble heme, but have other functional groups capable of binding to the Bfr receptor site. These molecules could be useful in defining the essential interactions that stabilize the Bfr-24-mer conformation.

Analysis of the Phase Screen results indicate that the molecules were high affinity ligands since most of the top 500 molecules scored above 1.5 (the top 10 are shown in table 3, column 8). However, most of the Phase molecules were reranked during the Glide docking process as seen in column 8 of Table 3 and apart from the protoporphyrin-like molecules from the ligand-based screen which had almost all the 7 sites matched (Table 2, column 8) the number of matched interaction sites in both approaches were similar (mostly 3-5 sites). To compute the theoretical binding energies of the top 500 ligand-based structures, they were docked into the same Bfr receptor in Figure 5.1. Moreover, the results are shown in Table 3, as well as the ligand poses in Figure 5.3.7. and Figure 5.3.85. Again, the docking scores were very high and comparable to the ligand-based approach.
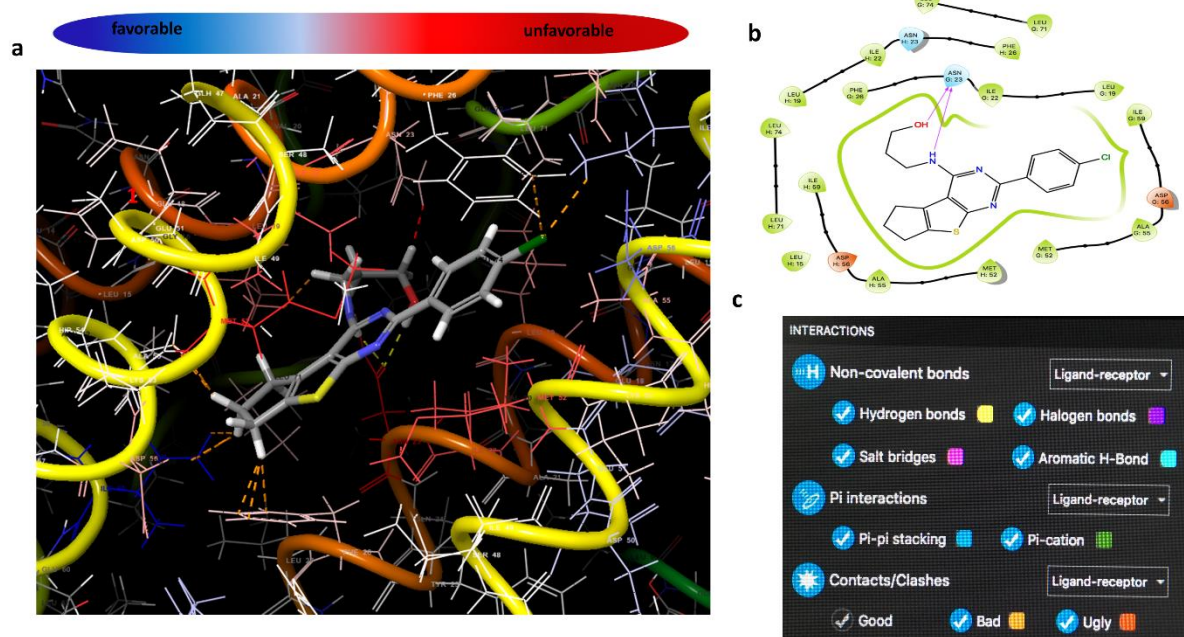
d
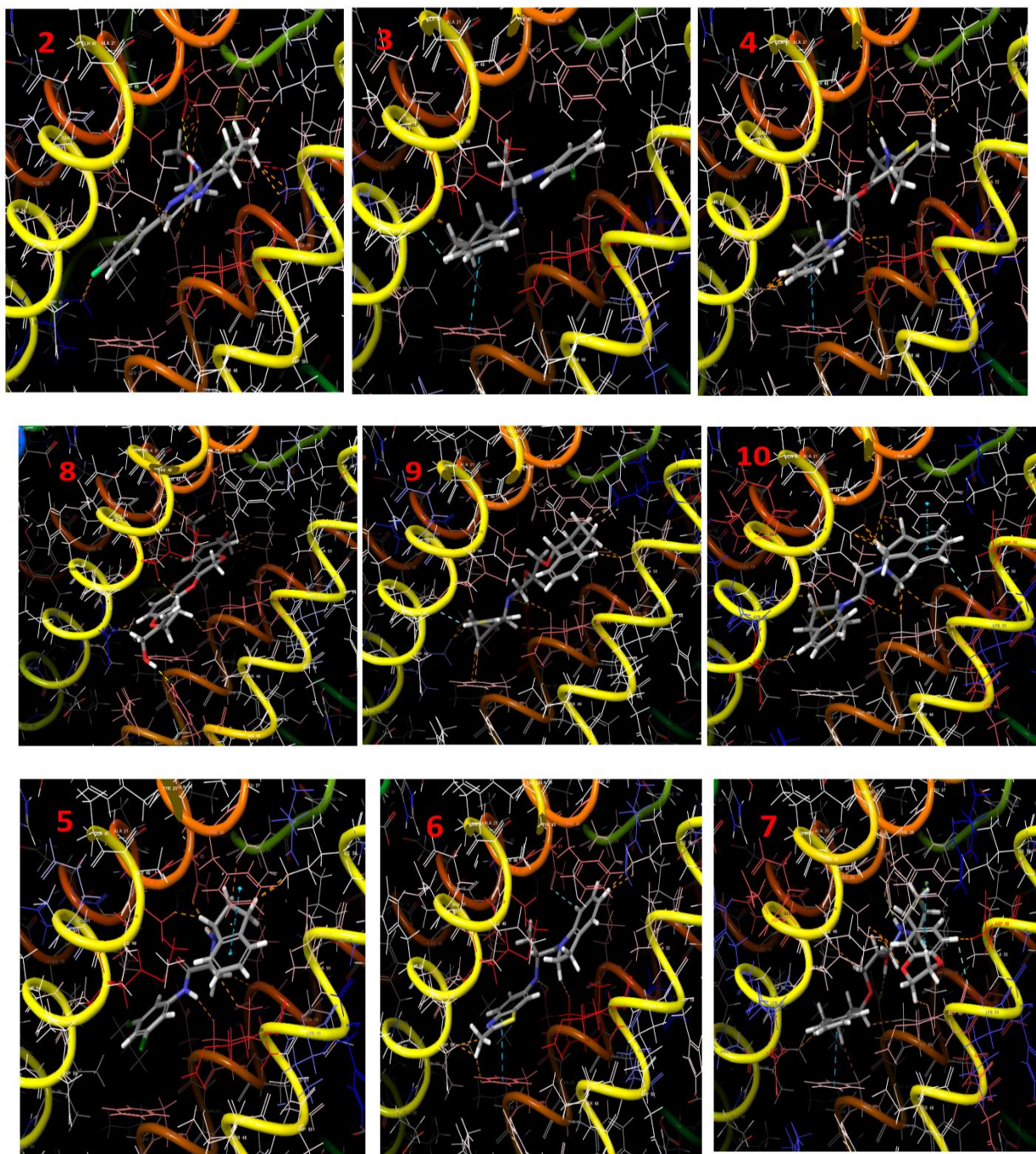
Figure 5.3.7. Glide ligand poses (from the fragment-based screen). (a) Shows the most active ligand pose oriented in the receptor. The receptor was represented in a thin tube while the ligand is in a skeletal model. (b) Ligand interaction diagram for the ligand and receptor. (c) Screenshot of the reaction legend. (d) the ligand poses of the top 10 active ligands. All images were exported from the Maestro Glide docking application.

Figure 5.3.8. Ligand interaction diagrams for top fragment-based ligands.

All images were exported from the Maestro Ligand interaction tool.
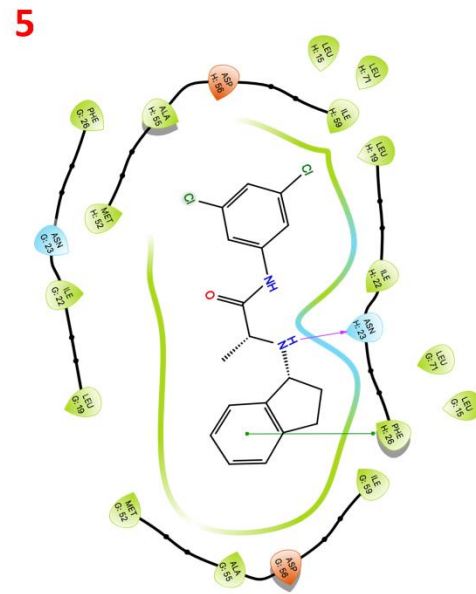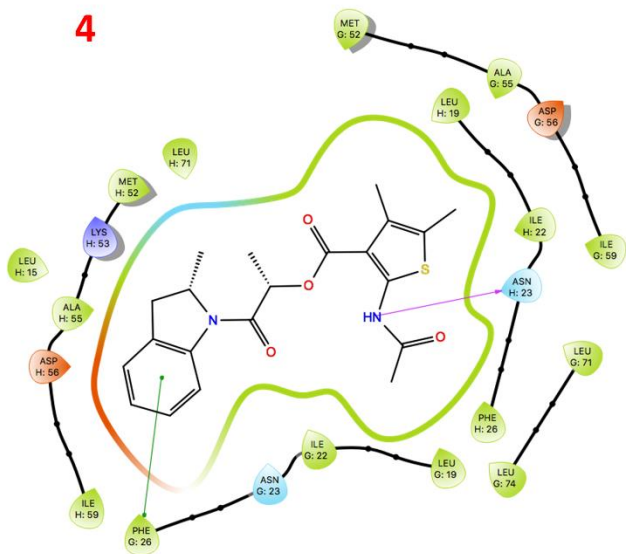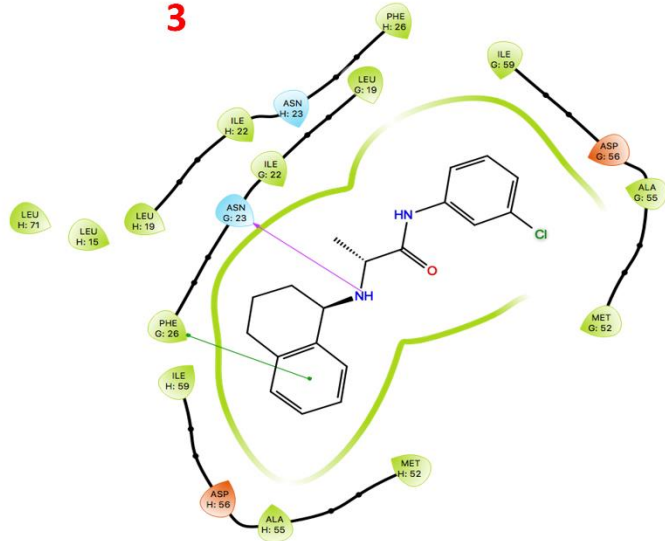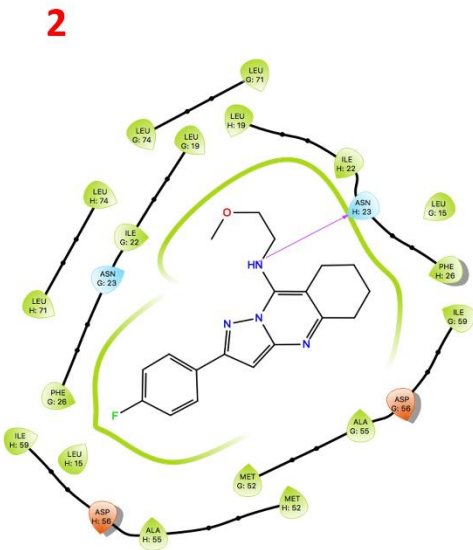
Title	glide rotatable bonds	docking score (kcal/mol)	XP GScore (kcal/mol)	glide gscore (kcal/mol)	Epik State Penalty	PhaseScreenScore	Num Sites Matched	PhaseScreen rank	MW
CACPD2011a-0000683356 (1)	6	-13.4496	-13.4496	-13.4496	0	1.628496	4	482	368.4527
CACPD2011a-0001437887 (2)	6	-13.3764	-13.3764	-13.3764	0	1.699115	5	9	359.873
CACPD2011a-0000172092 (3)	5	-13.291	-13.291	-13.291	0	1.633196	4	385	340.3946
CACPD2011a-0001803558 (4)	5	-13.2249	-13.6633	-13.6633	0.4384	1.656209	4	127	328.8358
CACPD2011a-0002004337 (5)	7	-13.1922	-13.1922	-13.1922	0	1.635785	4	347	400.4913
CACPD2011a-0001494242 (6)	5	-13.118	-13.5473	-13.5473	0.4293	1.640145	4	287	349.2543
CACPD2011a-0002008716 (7)	4	-13.0227	-13.0227	-13.0227	0	1.658965	4	112	337.4387
CACPD2011a-0001998234 (8)	7	-12.7963	-12.7963	-12.7963	0	1.631319	5	432	410.5259
CACPD2011a-0001435668 (9)	5	-12.7911	-12.7911	-12.7911	0	1.65746	5	115	408.4007
CACPD2011a-0002575133 (10)	9	-12.6746	-14.0088	-14.0088	1.3342	1.62793	5	496	357.5096

Table 3. Docking results (fragment-based). shows how the Phase screen molecules perform in the Glide docking
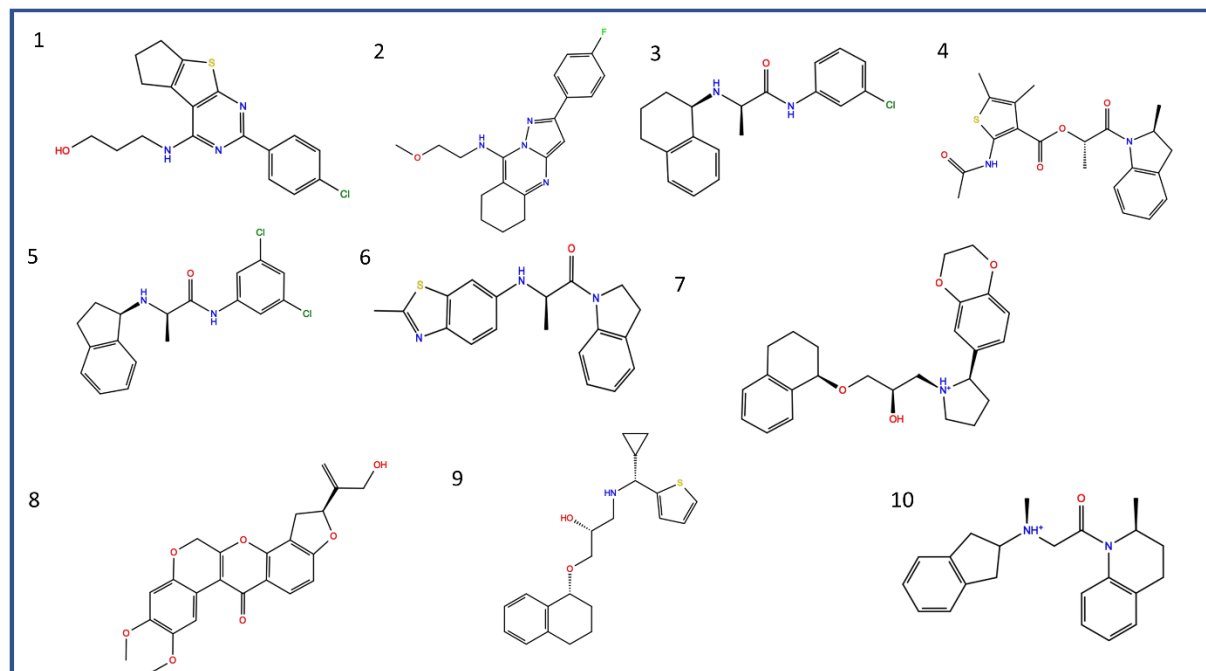
Figure 5.3.9. Top 10 ligand structures (fragment-based).

188

5.6. Induced Fit Docking

In order to account for the Bfr receptor flexibility, the IFD docking strategy described earlier was employed. The molecules from the Aldrich Market Direct® database were used since it contained virtually all the molecules in the CAP2011 database as well as those in the current database and information was present that would allow for purchasing of various molecules from commercial suppliers. Hence, the database was screened against the two hypotheses as described for the CAP2011 database to obtain ligand-based and fragment-based molecules/ligands. The two sets of ligands were separately docked into the same Bfr rigid receptor to obtain their Glide scores.

The top 15, middle 3, and bottom 2 structures from each set of data were prepared by employing the LigPrep program and further used in IFD docking, employing the same receptor, but this time using Prime and Glide together to obtain a more flexible Bfr receptor as illustrated in the flowchart (Figure 5.2). The IFD score was then used to rank the ligands into tight binders, moderate binders, and weak binders. Figures 5.4.and 5.4.1 below show the rank and the identification of the structures from the ligand-based and fragment-based, respectively.
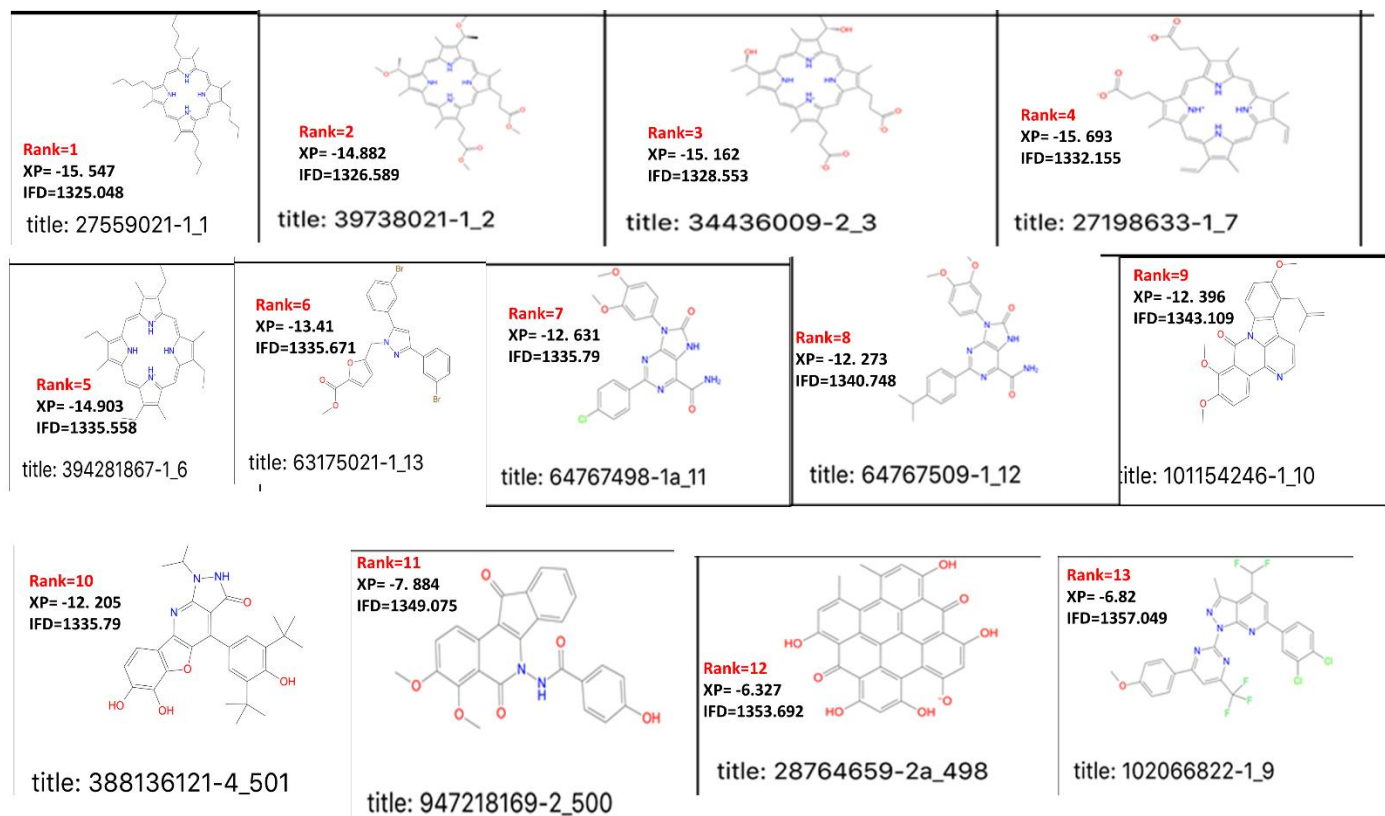
Figure 5.4. IFD rank of the top 13 molecules obtained from the Ligand-based screen. The ranks are in red, followed by the Glide XP score in kcal/mol also, the IFD score in kcal/mol. The title is explained as (ex. 947218169-2_500) the database ID (in red) and Glide rank (in blue).
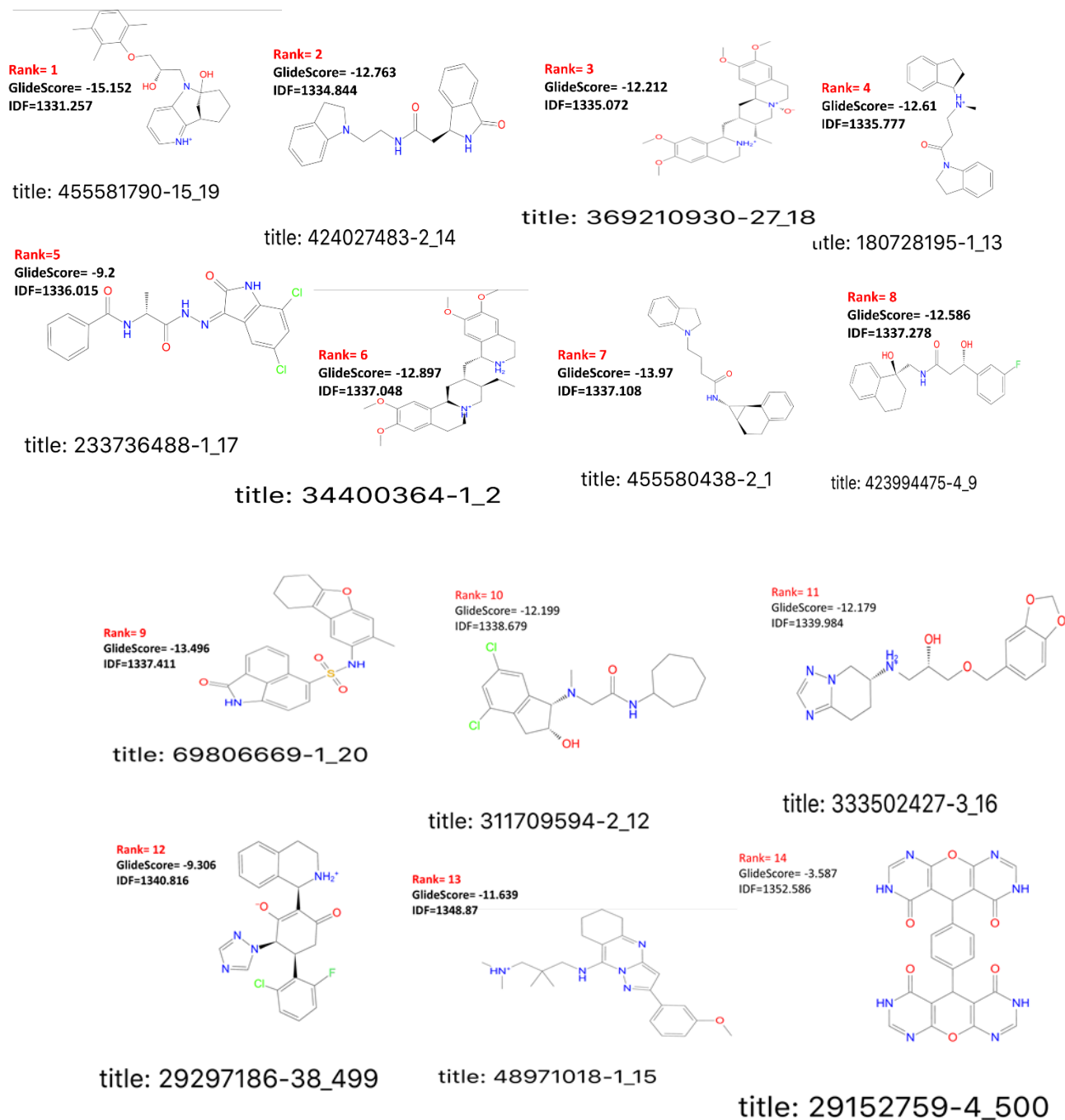
Figure 5.4.1. IFD rank of the top 14 molecules obtained from the Fragment-based screen.
The ranks are in red, followed by the Glide XP score in kcal/mol and IFD also score in kcal/mol.
The title is explained as (ex. 947218169-**2_500**) the database ID (in red) and Glide rank (in blue).

5.7. Reconstitution of the Docked Molecules into WTBfr

The results from IFD docking, using the Market select® database, were combined (both fragment-based and ligand-based) and compared to select a mixture of final ligands to purchase from suppliers. The selection contained most active ligands, moderately active ligands, and least active ligands. Overall, 11 ligands were studied: 4 protoporphyrin analogs (Protoporphyrin IX Fe (II) (heme or FePP), Protoporphyrin IX Zn (II) (ZnPP) Protoporphyrin IX molecule (PP), and Protoporphyrin IX dimethyl ester (DMEPP)) seen Figure 5.5 and 7 others from the 2 screens (Table 4) The selection of ligands with different binding energies will be used to study how the IFD score correlates to experimental results.
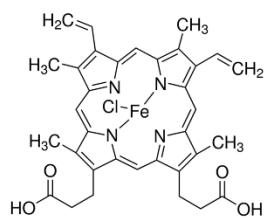
Chloroprotoporphyrin IX iron(III), Ferriprotoporphyrin IX chloride, Hemin(chloride) (FePP)
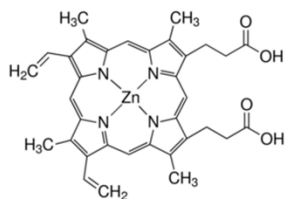
Protoporphyrin IX zinc(II) (ZnPP)
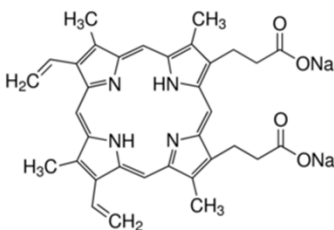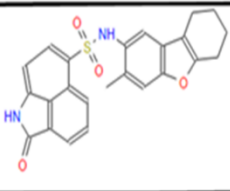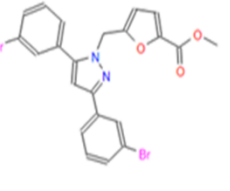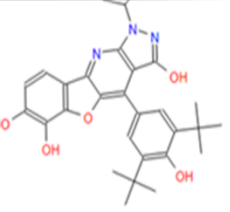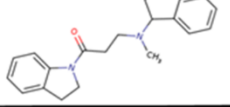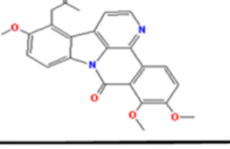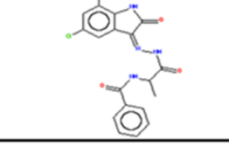
Protoporphyrin IX disodium salt (PP)
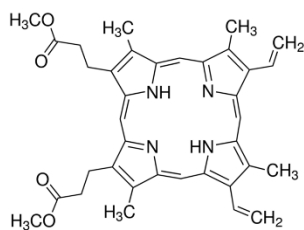
Protoporphyrin IX dimethyl ester (DMEPP)

Figure 5.5. The structures of heme analogs used for the initial encapsulation trials. All of the compounds were purchased from Sigma-Aldrich, Saint Louis, MO, USA

192

Chemical Structure	Chemical Name	MW/Empirical formula	Screen Type, rank	Supplier, ID
	N- (3- methyl- 6,7,8,9- tetrahydrodibenzo[b,d] furan- 2- yl)- 2- oxo- 1,2- dihydrobenzo[cd]indole- 6- sulfonamide	432.5/ C24H20N2O4S	FB, 9	Vitas-M Laboratory, Ltd (ID: STL229904)
	methyl5- {[3,5- bis (3- bromophenyl)- 1H- pyrazol- 1- yl]methyl}furan- 2- carboxylate	516.19/ C22H16Br2N2O3	LB, 6	Vitas-M Laboratory, Ltd (ID: STL391594)
	4- (3,5- di- tert- butyl- 4- hydroxyphenyl)- 1- (propan- 2- yl)- 1H- [1]benzofuro[3,2- b]pyrazolo [4,3- e] pyridine- 3,6,7- triol	503.6/ C29H33N3O5	LB, 10	Vitas-M Laboratory, Ltd (ID: STL314126)
	--	382.5/C23H30N2O3	FB, 1	Enamine (ID:Z2034808557)
	N-[3-(2,3-dihydro-1H-indol-1-Yl)-3-oxopropyl]-N-methyl-1-indanamine	340.2/C21H24N2O	FB, 2	Enamine (ID:Z354487900)
	--	414.5/C25H22N2O4	LB, 9	Pharmeks (ID:PHAR271869)
	--	405.2/C25H22N2O4	FB, 5	ID:Innovopharm Ltd (ID: STT-00066601)

Table 4. Structures of ligand molecules to be used for future encapsulation. FB and LB represent fragment-based and ligand-based, respectively. These molecules were purchased from Sigma-Aldrich Market Select.

Preliminary reconstitution experiments were set up with the four protoporphyrin analogs by combining 1 molar equivalent of WTBfr with 2 molar equivalents of the ligands at 80 °C for 10 minutes each and allowing the mixtures to cool down to room temperature. The mixtures were then buffer-exchanged into 150 mM NaCl, 50 mM MOPS or HEPES-pH 7.5 and filtered through a 0.45 um nylon syringe filter.

Dynamic Light Scattering (DLS) analysis was used to analyze the sample for quarternary structures in the solution after reconstitution. The preliminary data are shown below, Figure 5.6, indicates that all the analogs shift the equilibrium (Figure 4) towards a 24-mer with an average size of 16 ± 2 nm.

Finally, calibration curves (of size to mole equivalents of heme analogs) will be used to determine the minimum mole equivalent of heme required to form a 24-mer. This knowledge will then be used to study the possible binding of the remaining seven compounds and study their effects on their ability to reconstitute the 24-mer quaternary structure of Bfr. If observed, a correlation with the experimental ranking of each compound in the group with the computed ranking would be studied.
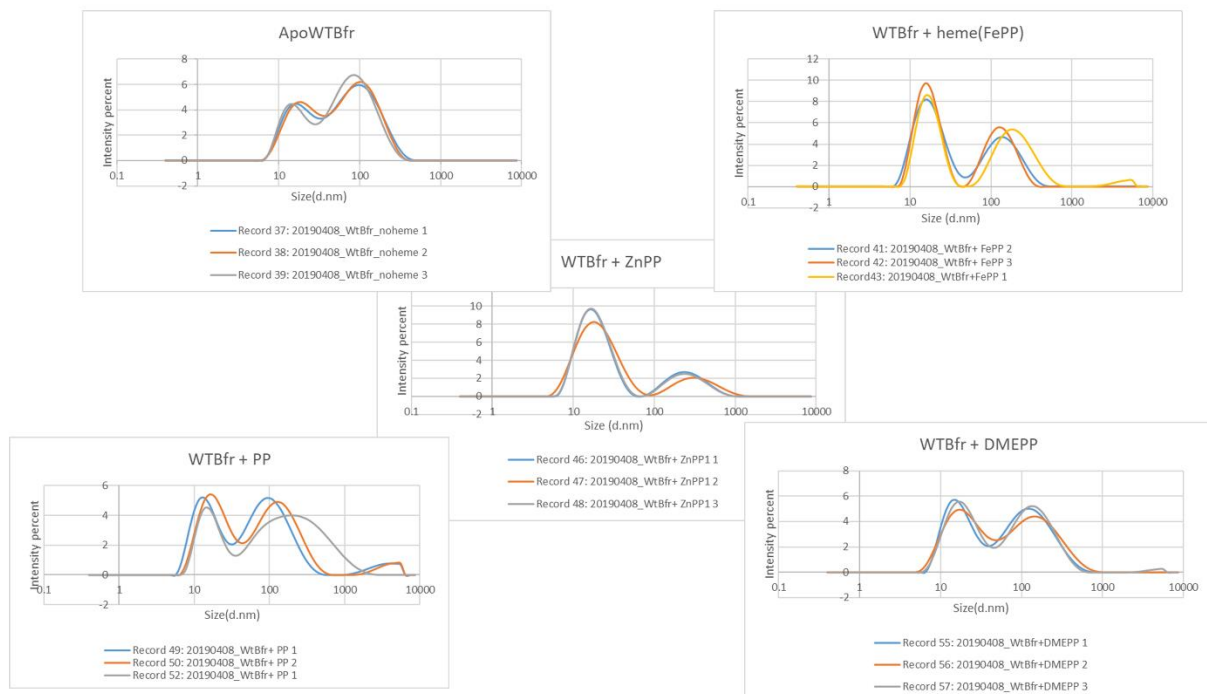
Figure 5.6. Preliminary DLS analysis on ligand reconstitution. WTBfr: Wild-type Bfr, FePP: hemin, ZnPP: protoporphyrin IX Zn (II), PP: protoporphyrin IX, DMEPP: protoporphyrin IX dimethyl ester

5.8. Summary and Future Directions for Bfr Pharcophore Modeling

This project has demonstrated the application of computational methods to a complex system, Bfr. Significant progress has been made regarding initial protein preparation and minimization. Also, a combination of Phase and Glide applications has been used to identify novel compounds that might bind to the Bfr-heme binding pocket. SP docking is sufficient in identifying "actives" for Bfr binding sites since it computes similar binding energies as the XP method, but in a fraction of the time. However, XP becomes essential when dealing with ligands with different conformations. Moreover, the rapid Phase screen results obtained v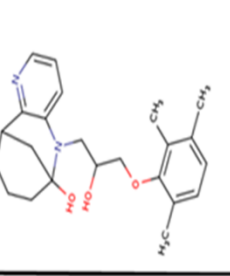alidate the initial hypothesis, but a higher Phase score does not guarantee tighter binding, docking is required to obtain the binding energies, and FID has proven to be useful for estimating the binding energies between Bfr and the stimulated ligand molecules.

SUMMARY AND FUTURE DIRECTIONS

6.1. Encapsulation of Inorganic Nanoparticles

The work in this thesis has explored how two microbial ferritins nanoplatforms, *Archaeoglobus fulgidus* ferritin (AfFtn) and *E. coli* Bacterioferritin (Bfr), can be engineered to make them more versatile. Expanding the versatilities of these nanoplatforms could repurpose them for nonnative applications such as drug delivery vehicles,[4b,38] hydrogen gas production (energy),[5-6] new materials with diverse functionalities,[2] water purification,[40] and many more.

Firstly, the Bfr nanoplatform was engineered both experimentally and computationally (Chapter 4). Previous investigations in the Honek laboratory and the research outlined herein involved the redesigning of the Bfr gene to encode for internal (C-terminal) polyhistidine tags which were used for entrapping NTA functionalized guest molecules like NTA-functionalized 5 nm gold nanoparticles (AuNP) (Figure 6). The AuNP contained about 4000 atoms of Au coated with a proprietary linker to which the NTA is attached. The AuNP was used as a proof of concept for encapsulating inorganic materials for applications in imaging. For example, gadolinium-based contrast agents can be functionalized with NTA groups and internalized into Bfr via the His-Ni^{2+}-NTA interaction. Encapsulated contrast agents can be utilized in Magnetic Resonance Imaging (MRI). [192]

In line with this, encapsulation of NTA-functionalized upconversion nanoparticles into HisBfr is under study in collaboration with Yang Hu (graduate student, Honek (Chemistry) and Lu (Physics) laboratories). The NaGdF4: Yb, Er (20%, 2%) upconversion nanoparticles were synthesized in oleic acid/ 1-octadecene (3:7) by a combined synthesis (at 150 °C) and nucleation process (at 300 °C). The synthesized nanoparticles with an oleic acid corona have 5.6 nm diameter, within the dimensions of the internal cavity of Bfr. Encapsulation is mediated by stripping the hydrophobic corona with 0.1 M HCl and reacting the bare nanoparticles with high-affinity N-(phosphonomethyl)iminodiacetic acid. Lastly, the nanoparticles, with phosphate corona are then modified with NTA functional groups to form NaGdF4: YbEr-NTA for encapsulation into HisBfr. Future work will explore the encapsulation of these types of nanomaterials within Bfr.
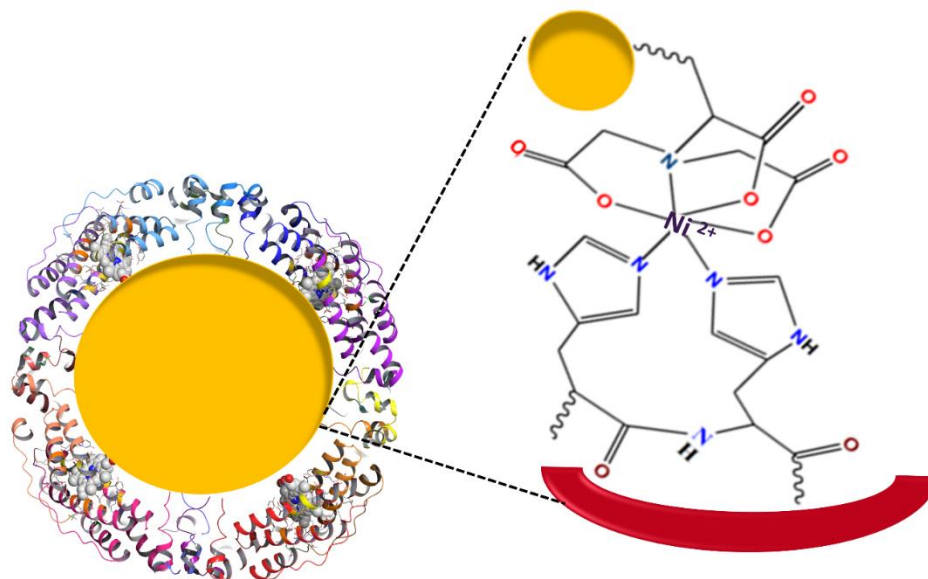
Figure 6. Inorganic nanoparticle encapsulation.

6.1. Hydrophobic Engineering

Additionally, a health-related application was explored by the hydrophobic engineering of Bfr. Manipulation of Bfr to encapsulate hydrophobic guest molecules such as hydrophobic drugs was investigated in a proof-of-concept by using the His-tagged Bfr (HisBfr) construct to encapsulate aggregates (micelles) of NTA functionalized lipid molecules. Encapsulating variable chain lengths of these lipids into either homopolymeric Bfr (100 % HisBfr) or heteropolymeric Bfr (50% HisBfr: 50 % WTBfr) allowed for control over both the capsule dimensions and the extent of hydrophobicity of Bfr. The hydrophobic cavity may find applications in transporting hydrophobic drugs such as curcumin. Curcumin is a polyphenolic compound derived from the turmeric spice. This compound has anti-inflammatory and anti-oxidant as well as an antitumor agent [193]. Luckily, it can be administered at a very high dose, up to 12 g per day,[194] yet suffers from bioavailability at this dose as a result of its limited intestinal solubility.[188,193-194] Most of the research on curcumin has focused on breaking this physiological barrier and reaping the medicinal benefits of curcumin. For example, Berginc *et al*. reported the addition of piperine, a curcumin metabolism inhibitor, into curcumin-containing formulations to inhibit curcumin glucuronidation (Figure 6.1). Also, another problem is the nonspecific binding of curcumin to the

mucous lining, which could be overcome by compromising (reversibly) the mucus integrity to minimize non-specific binding of curcumin.[193]

Hence, "hydrophobic" Bfr could overcome this limitation by serving as a medium to solubilize the curcumin molecule and transport it to the target. The ability of Bfr to encapsulate hydrophobic drugs within the hydrophobically engineered cavity was previously modeled with pyrene molecules by Omaima Ben Krayem (Honek laboratory). The HisBfr containing C19-NTA micelles internalized the pyrene molecules successfully. The process was monitored by pyrene emission, and it was found that the pyrene molecules within HisBfr-C19-NTA were shielded from the solution by the HisBfr-micelle cluster. Maturation of pyrene emission was observed upon declustering the HisBfr subunits with increasing concentrations of GdnHCl (up to 8 M).[121]

Regularly sized C19-NTA- Bfr capsules obtained from this work (with a mixture of WTBfr and HisBfr) could enhance its EPR effects, and warrants a further look into solubilizing the pyrene molecules within these regularly sized capsules. The knowledge gained from pyrene solubilization could be transferred to other hydrophobic drugs not limited to: curcumin (Figure 6.1), carmustine,[195] bevacizumab (Avastin™),[195] and haloperidol[196] (Figure 6.2). Furthermore, AUC and pyrene fluorescence will be used to determine the aggregation numbers of the lipid molecules within the Bfr capsules.
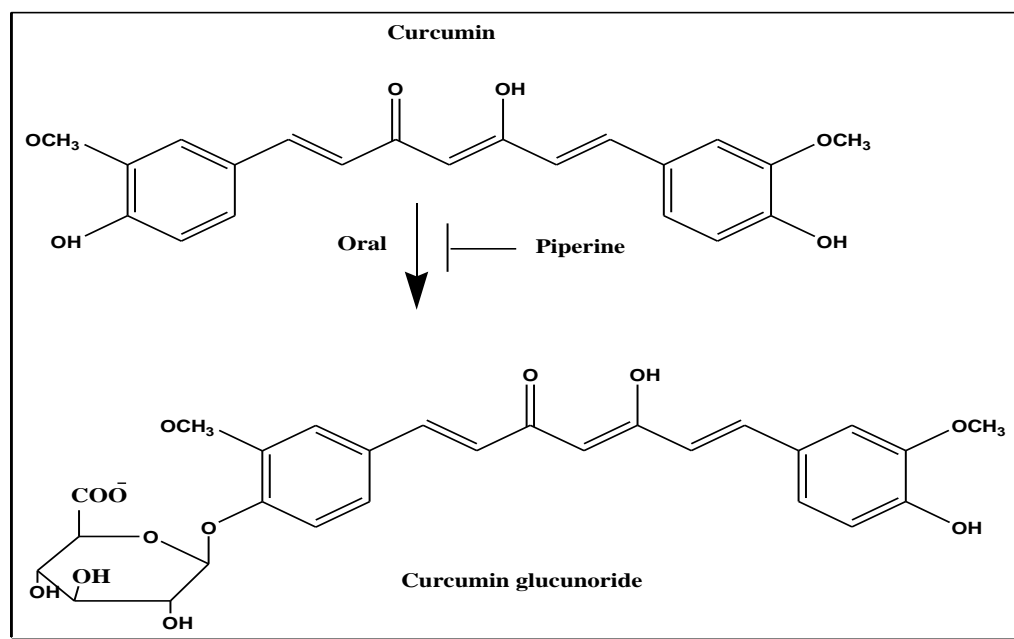
Figure 6.1. Piperine inhibition of curcumin glucuronidation
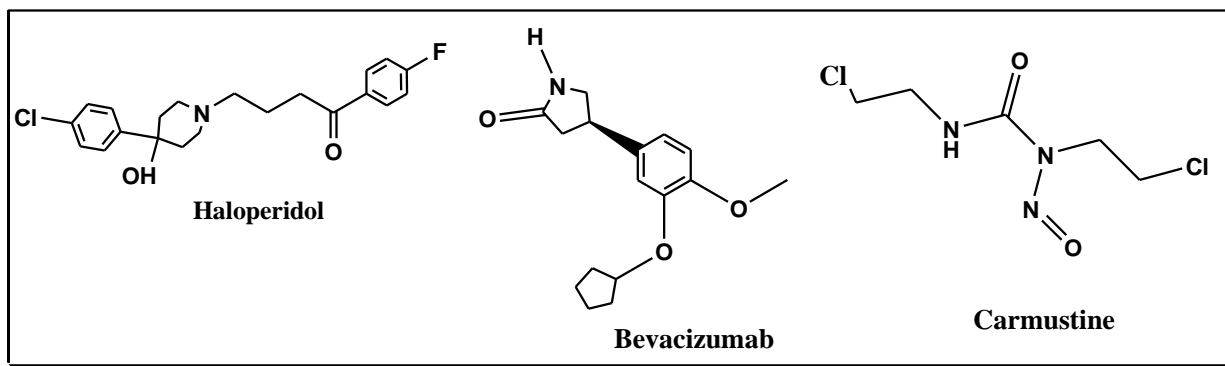
Figure 6.2. Structures of several hydrophobic drugs.

6.2. Surface Modification

Moreover, one of the most challenging issues in drug delivery is that of targeting the drug to a specific receptor. [29] This can be achieved by modifying the formulation with cell-targeting moieties which specifically target cells expressing surface antigens particular to an infection or a disease.[4b,36a,197], For example, the surface of human apoferritin nanocages were modified with an RGD recognition sequence (Cys-Asp-Cys-Arg-Gly-Asp-Cys-Phe-Cys) which was successfully used to deliver doxorubicin to U87MG subcutaneous tumor models. The drug-loaded apoferritin was reported to have very high efficacy (up to 73 wt %) and less cardiotoxicity than free doxorubicin.[197]

Additionally, the surface modification can be useful in multi-target drug (MTD) design. This design principle is becoming popular in drug discovery as a result of varied causative factors in diseases such as cancer, Alzheimer's (AD), and Parkinson's (PD) diseases.[198] The traditional "on-target" drugs specifically target a receptor on the diseased cell to prevent unwanted side effects. However, it not sufficient to treat the "current incurable diseases" (listed above). [199] [200] Hence, the emergence of the single drug molecule, multi-target drug, capable of achieving therapeutic effects in complex diseases.[199,200,198] A single Bfr capsule could be engineered to display variable tags/ peptide sequences/ antibodies that could recognize different receptor molecules in these diseases.[3,201] This way, it increases the probability of binding to at least one receptor and allows the efficient delivery of the encapsulate MTD.

Here, the Bfr nanoplatform has been engineered with multiple N-terminal surface tags that enable Bfr to undergo transglutaminase-catalyzed reactions (QTag-Bfr), Sortase mediated modification (Sort tag-Bfr) and also the tendency to reconstitute RNase activity (Qtag1-Bfr and Qtag2-Bfr). The Qtag2-Bfr has been used to demonstrate the ability of Bfr to be used in target cargo delivery vehicles similar to apoferritin above. The model cargo used was AuNP in the interior of Qtag2-Bfr while the Qtag on the surface was covalently modified with dansylcadaverine (DC). The success with DC, as primary amine, could be extended to incorporate diverse biological molecules such as antibodies and intact proteins. Incorporation of intact proteins could be beneficial for developing binary complexes of biomolecules.[2]

Lastly, the amino acid residues, such as exposed lysines on Bfr, can be modified via NHS activation to introduce diverse molecules. Current work in the Honek laboratory (by Fatima Merza, graduate student) involves the building of 3D ferritin frameworks (networks) through ferritin-polyethylene glycol (PEG)-biotin-avidin networks. The four biotin sites on avidin engage biotinylated ferritins in a 3D network, as seen in Figure 6.3 below. Also, the other exposed amino acid functional groups in the network, on ferritin surfaces, could possibly be chemoselectively modified with enzymes, sugars, and other groups.**.**
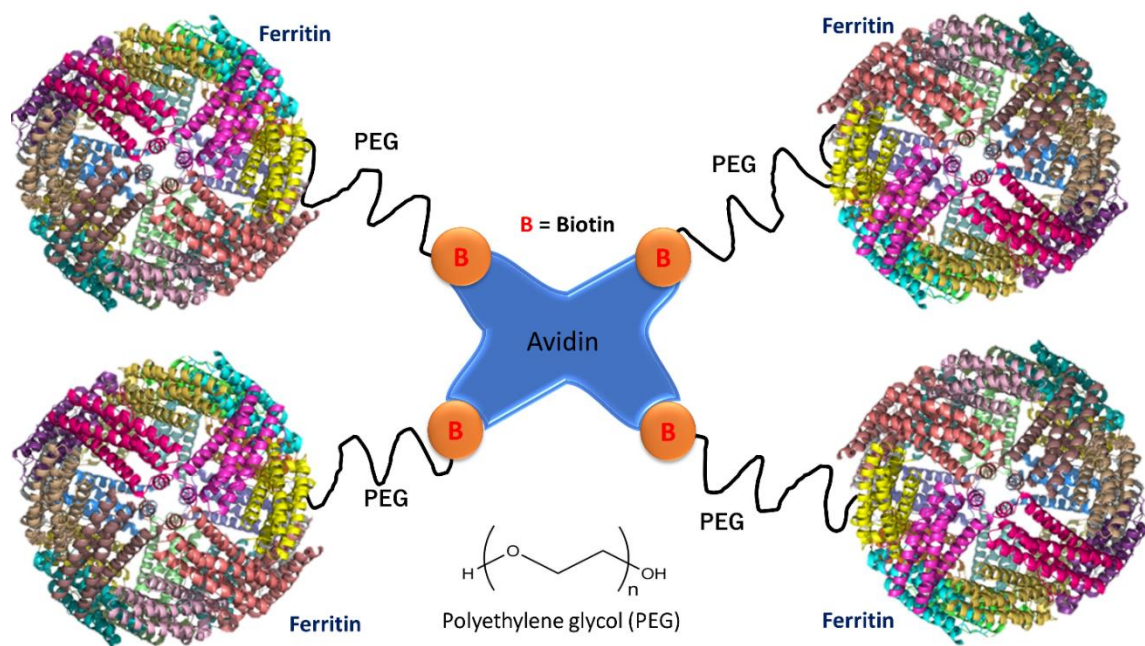
Figure 6.3. Representation of a 3D ferritin-based protein network.

6.3. Encapsulation of His-tagged Guest Molecules into Bfr

The reverse polarity encapsulation strategy discussed in a previous chapter investigated the process of encapsulating His-tagged guest molecules into the Bfr cavity. In reverse polarity, the heme cassette is used as the biotemplating tool due to the ability to chemically modify it and reinsert it into the Bfr dimer interface. The modified bisheme-NTA interacts with the His-tags of the guest molecule (in the presence of Ni^{2+}) to internalize the guest into the Bfr cavity (Figure 6.4). This approach of modifying Bfr extends the host encapsulation capabilities to include His-tagged enzymes and proteins as most enzymes are now genetically engineered with polyhistidine tags for their ease of purification. Hence to use Bfr as an enzyme container, it is necessary to use this reverse polarity to accommodate the enzymes.

His-tagged enhanced GFP (HisGFP) was used as a model His-tagged protein for this study. Successful encapsulation was monitored by HisGFP absorbance (at 490 nm). Also, the extent of HisGFP fluorophore protection was briefly studied with DTNB, and the results showed that the Bfr shell protects the HisGFP molecules from DTNB modification. However, a detailed titration with varying DTNB concentrations should be performed to obtain the exact concentration of DTNB at which the fluorescence begins to decrease.
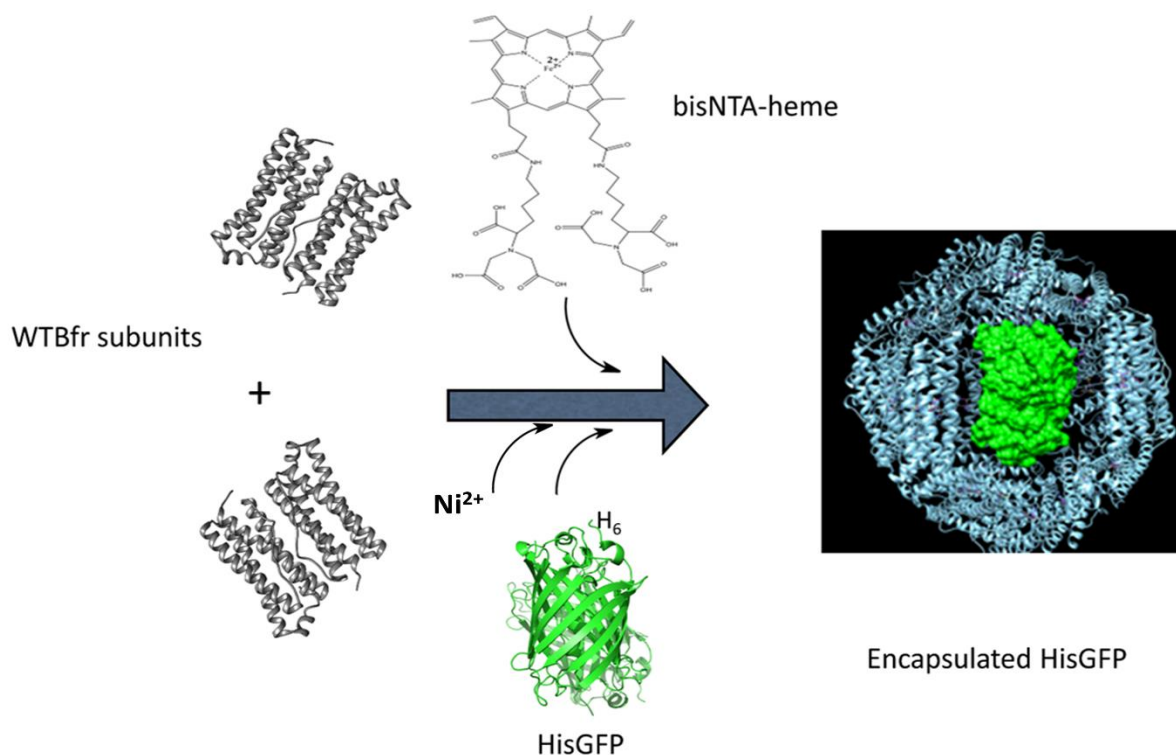
Figure 6.4. The reverse polarity encapsulation of HisGFP into BisNTA-heme-Bfr.

6.4. Characterization and Modification of *Archaeoglobus Fulgidus* Ferritin (AfFtn)

This work explored the ferritin from the hyperthermophilic organism *Archaeglobus fulgidus* to study how the presence of the AfFtn triangular pores will affect the chemistry of encapsulated guest molecules compared to bacterioferritin. The AfFtn was successfully engineered into three different constructs (WTBfr, HisAfFtn, and AfFtnC54) for various encapsulation applications. Similar to Bfr, the protein yields were about 30 mg/L, and all the subunit masses were validated by +ESI-MS.

The NTA-functionalized gold nanoparticle (5 nm in diameter) was encapsulated successfully into HisAfFtn. The encapsulation was mediated by decreasing the salt concentration to 10 mM to allow subunit dissociation and increasing the salt concentration to 150 mM to cause subunit association and guest entrapment. However, the AfFtnC54 construct did not associate in 150 mM salt, but in 20 mM $MgCl_2$. The AfFtnC54 construct was designed for encapsulating a

His-tagged glucohydrolase (and other His-tagged guests) since the AfFtn capsule is a non-heme containing ferritin.

Again, HisGFP was used as a model for encapsulating a His-tagged protein since it is easy to monitor its encapsulation as well as there is an existing knowledge on it from the Bfr experiments. The AfFtnC54 was modified via maleimide chemistry to introduce NTA functionality to the C54 residue (toward the interior of the cavity, Figure 6.5a). The resultant AfFtnC-54-NTA was then utilized to entrap HisGFP in a 20 mM $MgCl_2$, 25 mM HEPES-pH 7.5 buffer. The results from this encapsulation showed that HisGFP was encapsulated under these conditions and will be applied to the encapsulation of a glucohydrolase enzyme to study the effects of the cage porosity on substrate specificity.

In the future, the thermostable His-tagged glucohydrolase, alpha-D-glucoside glucohydrolase, from Megazyme (through Cederlane, Burlington, ON) will be encapsulated into AfFtnC54-NTA. The encapsulation will be similar to the HisGFP above, mediated by low salt and $MgCl_2$. After the encapsulation, the entrapped alpha glucohydrolase will be examined for its ability to hydrolyze various oligomers of the sugar substrate. Some of the sugars will include maltose, p-nitrophenyl α-ᴅ-galactopyranoside, and maltotetraose up to maltooctaose. The masses of the substrates (sugars) will be determined by electrospray ionization mass spectrometry before incubation with the encapsulated enzyme and post-incubation to determine the structure of the hydrolyzed products. Also, some of the sugars have fluorescent labels which can be used to monitor the hydrolysis.

Of interest is how the triangular pore of AfFtn is capable of screening the substrate molecules from the entrapped enzyme (as seen in Figure 6.5b). This platform may have applications in stabilizing the entrapped enzymes under high temperature as well. Thermostability can be studied by NMR spectroscopy where the changes in chemical shifts can be used to study the denaturation of the enzyme. It might be possible that the entrapped glucohydrolase, when heated, will be protected against heat denaturation and have similar NMR chemical shifts to the intact enzyme.

Overall, significant progress has been made towards AfFtn and Bfr biotemplating. Both capsules may have similar chemistries in terms of guest encapsulation and protection. More importantly, encapsulation of non-native guest molecules into AfFtn will provide beneficial

insights to the scaffold since it was recently reported,[8] and there is limited knowledge about the structure and its unusual pore arrangement.
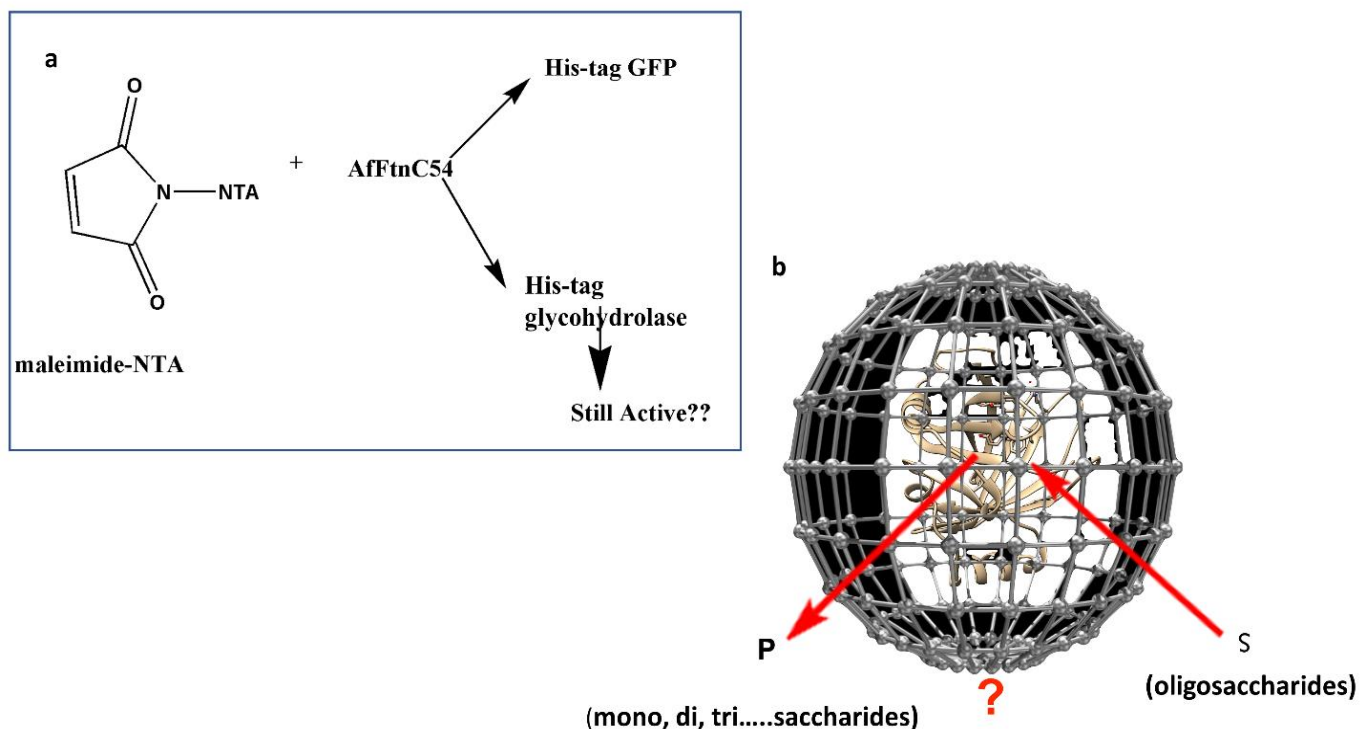
Figure 6.5. The AfFtn catalytic container. (a) Illustrates the modification of AfFtnC54 to AfFtn C54-NTA for HisGFP and His-alpha glucosidase encapsulation. (b) This figure is a representation of our hypothesis; the encapsulated enzyme may or may not be able to hydrolyze the substrates, S, into products, P, depending on how large the substrate is.

REFERENCES

1. Uchida, M.; Klem, M. T.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L. O.; Young, M.; Douglas, T., Biological Containers: Protein Cages as Multifunctional Nanoplatforms. *Advanced Materials* **2007,** *19* (8), 1025-1042.

2. Liljeström, V.; Mikkilä, J.; Kostiainen, M. A., Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. *Nature Communications* **2014,** *5*.

3. Rachel, N. M.; Toulouse, J. L.; Pelletier, J. N., Transglutaminase-Catalyzed Bioconjugation Using One-Pot Metal-Free Bioorthogonal Chemistry. *Bioconjugate Chemistry* **2017,** *28* (10), 2518-2523.

4. (a) Lei, Y.; Hamada, Y.; Li, J.; Cong, L.; Wang, N.; Li, Y.; Zheng, W.; Jiang, X., Targeted tumor delivery and controlled release of neuronal drugs with ferritin nanoparticles to regulate pancreatic cancer progression. *Journal of Controlled Release : Official Journal of the Controlled Release Society* **2016,** *232*, 131-42;

   (b) Falvo, E.; Tremante, E.; Fraioli, R.; Leonetti, C.; Zamparelli, C.; Boffi, A.; Morea, V.; Ceci, P.; Giacomini, P., Antibody-drug conjugates: targeting melanoma with cisplatin encapsulated in protein-cage nanoparticles based on human ferritin. *Nanoscale* **2013,** *5* (24), 12278-12285.

5. Varpness, Z.; Peters, J. W.; Young, M.; Douglas, T., Biomimetic Synthesis of a H2 Catalyst Using a Protein Cage Architecture. *Nano Letters* **2005,** *5* (11), 2306-2309.

6. Clark, E. R.; Kurtz, D. M., Jr., Photosensitized H2 generation from "one-pot" and "two-pot" assemblies of a zinc-porphyrin/platinum nanoparticle/protein scaffold. *Dalton Transactions (Cambridge, England : 2003)* **2016,** *45* (2), 630-8.

7. Jeong, C. K.; Kim, I.; Park, K. I.; Oh, M. H.; Paik, H.; Hwang, G. T.; No, K.; Nam, Y. S.; Lee, K. J., Virus-directed design of a flexible BaTiO3 nanogenerator. *ACS Nano* **2013,** *7* (12), 11016-25.

8. Johnson, E.; Cascio, D.; Sawaya, M. R.; Gingery, M.; Schroder, I., Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic archaeon Archaeoglobus fulgidus. *Structure (London, England : 1993)* **2005,** *13* (4), 637-48.

9. van Eerde, A.; Wolterink-van Loo, S.; van der Oost, J.; Dijkstra, B. W., Fortuitous structure determination ofas-isolated'Escherichia coli bacterioferritin in a novel crystal form. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **2006,** *62* (11), 1061-1066.

10. Marsh, J. A.; Teichmann, S. A., Structure, dynamics, assembly, and evolution of protein complexes. *Annual Review of Biochemistry* **2015,** *84*, 551-75.

11.	Alonso, J. M.; Gorzny, M. L.; Bittner, A. M., The physics of tobacco mosaic virus and virus-based devices in biotechnology. *Trends in Biotechnology* **2013,** *31* (9), 530-8.

12.	Tyagi, R.; Duquerroy, S.; Navaza, J.; Guddat, L. W.; Duggleby, R. G., The crystal structure of a bacterial class II ketol-acid reductoisomerase: domain conservation and evolution. *Protein science : a publication of the Protein Society* **2005,** *14* (12), 3089-100.

13.	Virnau, P.; Mallam, A.; Jackson, S., Structures and folding pathways of topologically knotted proteins. *Journal of physics. Condensed Matter : An Institute of Physics Journal* **2011,** *23* (3), 033101.

14.	(a) Cardarelli, L.; Lam, R.; Tuite, A.; Baker, L. A.; Sadowski, P. D.; Radford, D. R.; Rubinstein, J. L.; Battaile, K. P.; Chirgadze, N.; Maxwell, K. L.; Davidson, A. R., The crystal structure of bacteriophage HK97 gp6: defining a large family of head-tail connector proteins. *Journal of Molecular Biology* **2010,** *395* (4), 754-68;

(b)	Wikoff, W. R.; Liljas, L.; Duda, R. L.; Tsuruta, H.; Hendrix, R. W.; Johnson, J. E., Topologically linked protein rings in the bacteriophage HK97 capsid. *Science (New York, N.Y.)* **2000,** *289* (5487), 2129-33.

15.	Wang, C. Y.; Zhang, Q. F.; Gao, Y. Z.; Zhou, X. P.; Ji, G.; Huang, X. J.; Hong, J.; Zhang, C. X., Insight into the three-dimensional structure of maize chlorotic mottle virus revealed by Cryo-EM single particle analysis. *Virology* **2015,** *485*, 171-8.

16.	Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; et al., Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **1991,** *349* (6309), 541-4.

17.	Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.

18.	Kim, K. K.; Kim, R.; Kim, S. H., Crystal structure of a small heat-shock protein. *Nature* **1998,** *394* (6693), 595-9.

19.	Wynne, S. A.; Crowther, R. A.; Leslie, A. G., The crystal structure of the human hepatitis B virus capsid. *Molecular cell* **1999,** *3* (6), 771-80.

20.	 Pieters, B. J.; van Eldijk, M. B.; Nolte, R. J.; Mecinović, J., Natural supramolecular protein assemblies. *Chemical Society Reviews* **2016,** *45*, 24-39.

21.	Rothemund, P. W., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297-302.

22.   Shih, W. M.; Quispe, J. D.; Joyce, G. F., A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **2004,** *427* (6975), 618-21.

23.   Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410* (6828), 541-8.

24.   Sinnott SB, A. R., Carbon nanotubes: synthesis, properties, and applications. *Critical Reviews in Solid State Material Sciences* **2001,** *26*, 145–249.

25.   (a) Teunissen, E. A.; de Raad, M.; Mastrobattista, E., Production and biomedical applications of virus-like particles derived from polyomaviruses. *Journal of Controlled Release : Official Journal of the Controlled Release Society* **2013,** *172* (1), 305-321;

 (b)   Wu, Y.; Yang, H.; Shin, H.-J., Viruses as self-assembled nanocontainers for encapsulation of functional cargoes. *Korean Journal of Chemical Engineering* **2013,** *30* (7), 1359-1367.

26.   Roldao, A.; Mellado, M. C.; Castilho, L. R.; Carrondo, M. J.; Alves, P. M., Virus-like particles in vaccine development. *Expert Review of Vaccines* **2010,** *9* (10), 1149-76.

27.   Touze, A.; Coursaget, P., In vitro gene transfer using human papillomavirus-like particles. *Nucleic Acids Research* **1998,** *26* (5), 1317-23.

28.   Andrews, S. C.; Robinson, A. K.; Rodriguez-Quinones, F., Bacterial iron homeostasis. *FEMS Microbiology Reviews* **2003,** *27* (2-3), 215-37.

29.   Jahangirian, H.; Lemraski, E. G.; Webster, T. J.; Rafiee-Moghaddam, R.; Abdollahi, Y., A review of drug delivery systems based on nanotechnology and green chemistry: green nanomedicine. *International Journal of Nanomedicine* **2017,** *12*, 2957-2978.

30.   Iwahori, K.; Yoshizawa, K.; Muraoka, M.; Yamashita, I., Fabrication of ZnSe nanoparticles in the apoferritin cavity by designing a slow chemical reaction system. *Inorganic Chemistry* **2005,** *44* (18), 6393-400.

31.   Young, T. D. M., Host–guest encapsulation of materials by assembled virus protein cages. *Nature* **1998,** *393*, 152-155.

32.   Zheng, B.; Yamashita, I.; Uenuma, M.; Iwahori, K.; Kobayashi, M.; Uraoka, Y., Site-directed delivery of ferritin-encapsulated gold nanoparticles. *Nanotechnology* **2010,** *21* (4), 045305.

33.   Dixit, S. K.; Goicochea, N. L.; Daniel, M. C.; Murali, A.; Bronstein, L.; De, M.; Stein, B.; Rotello, V. M.; Kao, C. C.; Dragnea, B., Quantum dot encapsulation in viral capsids. *Nano Letters* **2006,** *6* (9), 1993-9.

34. O. Tagit, M. V. d. R., M. Brasch, Y. Ma and J. J. L. M. Cornelissen *, Quantum dot encapsulation in virus-like particles with tuneable structural properties and low toxicity. *RSC Advances* **2017,** *7* (38110).

35. Rother, M.; Nussbaumer, M. G.; Renggli, K.; Bruns, N., Protein cages and synthetic polymers: a fruitful symbiosis for drug delivery applications, bionanotechnology and materials science. *Chemical Society Reviews* **2016,** *45* (22), 6213-6249.

36. (a) Bao, G.; Mitragotri, S.; Tong, S., Multifunctional nanoparticles for drug delivery and molecular imaging. *Annual Review of Biomedical Engineering* **2013,** *15*, 253-82.

   (b) Wang, W.; Knovich, M. A.; Coffman, L. G.; Torti, F. M.; Torti, S. V., Serum ferritin: Past, present and future. *Biochimica et Biophysica Acta* **2010,** *1800* (8), 760-9.

37. Uchida, M.; Kosuge, H.; Terashima, M.; Willits, D. A.; Liepold, L. O.; Young, M. J.; McConnell, M. V.; Douglas, T., Protein cage nanoparticles bearing the LyP-1 peptide for enhanced imaging of macrophage-rich vascular lesions. *ACS Nano* **2011,** *5* (4), 2493-502.

38. Papaefthymiou, G. C., The Mossbauer and magnetic properties of ferritin cores. *Biochimica et Biophysica Aacta* **2010,** *1800* (8), 886-97.

39. Genove, G.; DeMarco, U.; Xu, H.; Goins, W. F.; Ahrens, E. T., A new transgene reporter for in vivo magnetic resonance imaging. *Nature Medicine* **2005,** *11* (4), 450-4.

40. Morse, G. K.; Brett, S. W.; Guy, J. A.; Lester, J. N., Review: Phosphorus Removal and Recovery Technologies. *Science of The Total Environment* **1998,** *212* (1), 69-81.

41. Aitken-Rogers, H.; Singleton, C.; Lewin, A.; Taylor-Gee, A.; Moore, G. R.; Le Brun, N. E., Effect of phosphate on bacterioferritin-catalysed iron(II) oxidation. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* **2004,** *9* (2), 161-70.

42. Jacobs, J. F.; Hasan, M. N.; Paik, K. H.; Hagen, W. R.; van Loosdrecht, M. C., Development of a bionanotechnological phosphate removal system with thermostable ferritin. *Biotechnology and Bioengineering* **2010,** *105* (5), 918-23.

43. Maria, T.; Jethro, M.; Wei, X.; Xiao H., Protein-Based Bioelectronics. *ACS Biomaterials Science & Engineering* **2016,** *2*, 1211−1223.

44. Wu, H.; Gao, W.; Yin, Z., Materials, Devices and Systems of Soft Bioelectronics for Precision Therapy. *Advanced Healthcare Materials* **2017,** *6* (10).

45. Bostick, C. D.; Mukhopadhyay, S.; Pecht, I.; Sheves, M.; Cahen, D.; Lederman, D., Protein bioelectronics: a review of what we do and do not know. *Reports on progress in physics. Physical Society (Great Britain)* **2018,** *81* (2), 026601.

46.   Masato, Tominaga, Kazuki, S.; Manabu, M.; Isao T., Electrostatic modification of ferritin onto polypeptide-functionalized indium oxide electrode surfaces: Electrochemical and AFM studies. *Journal of Electroanalytical Chemistry* **2005,** *579* (1), 51-58.

47.   Biswas, S.; Kinbara, K.; Oya, N.; Ishii, N.; Taguchi, H.; Aida, T., A tubular biocontainer: metal ion-induced 1D assembly of a molecularly engineered chaperonin. *Journal of the American Chemical Society* **2009,** *131* (22), 7556-7.

48.   Kim, E. Y.; Tullman-Ercek, D., Engineering nanoscale protein compartments for synthetic organelles. *Current Opinion in Biotechnology* **2013,** *24* (4), 627-32.

49.   Conrado, R. J.; Mansell, T. J.; Varner, J. D.; DeLisa, M. P., Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metabolic Engineering* **2007,** *9* (4), 355-63.

50.   (a) Minten, I. J.; Nolte, R. J.; Cornelissen, J. J., Complex assembly behavior during the encapsulation of green fluorescent protein analogs in virus derived protein capsules. *Macromolecular Bioscience* **2010,** *10* (5), 539-45;

(b)   Dustin P. Patterson, B. S., Kheireddine El-Boubbou, John van der Oost, Peter E. Preveligec; Douglas, a. T., Virus-like particle nanoreactors: programmed encapsulation of the thermostable CelB glycosidase inside the P22 capsid. *Soft Matter* **2012,** *8*, 10158–10166.

51.   Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N. A.; Rowan, A. E.; Christianen, P. C.; Maan, J. C.; Verduin, B. J.; Cornelissen, J. J.; Nolte, R. J., A virus-based single-enzyme nanoreactor. *Nature Nanotechnology* **2007,** *2* (10), 635-9.

52.   Xiao, C.; Kuznetsov, Y. G.; Sun, S.; Hafenstein, S. L.; Kostyuchenko, V. A.; Chipman, P. R.; Suzan-Monti, M.; Raoult, D.; McPherson, A.; Rossmann, M. G., Structural studies of the giant mimivirus. *PLoS biology* **2009,** *7* (4), e92.

53.   Astruc, D.; Lu, F.; Aranzaes, J. R., Nanoparticles as recyclable catalysts: the frontier between homogeneous and heterogeneous catalysis. *Angewandte Chemie (International ed. in English)* **2005,** *44* (48), 7852-72.

54.   Theil, E. C.; Behera, R. K.; Tosha, T., Ferritins for chemistry and for life. *Coordination Chemistry Reviews* **2013,** *257* (2), 579-586.

55.   Carrondo, M. A., Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *The EMBO Journal* **2003,** *22* (9), 1959-68.

56.   Andrews, S. C., The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica et Biophysica Acta* **2010,** *1800* (8), 691-705.

57. Boukhalfa, H.; Crumbliss, A. L., Chemical aspects of siderophore mediated iron transport. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and Medicine* **2002,** *15* (4), 325-39.

58. Gerl, M.; Jaenicke, R.; Smith, J. M.; Harrison, P. M., Self-assembly of apoferritin from horse spleen after reversible chemical modification with 2,3-dimethylmaleic anhydride. *Biochemistry* **1988,** *27* (11), 4089-96.

59. Andrews S.C.; Smith J. M.; Guest J. R.; Harrison P. M., Amino acid sequence of the bacterioferritin (cytochrome b1) of Escherichia coli-K12. *Biochemical and Biophysical Research Communications* **1989,** *158,* (2), 489-96.

60. Tatur, J.; Hagen, W. R.; Matias, P. M., Crystal structure of the ferritin from the hyperthermophilic archaeal anaerobe Pyrococcus furiosus. *Journal of Biological Inorganic Chemistry* **2007,** *12* (5), 615-630.

61. Moore G. R.; Kadir F. H.; al-Massad F. K.; Le Brun N. E.; Thomson A. J.; Greenwood C.; Keen J. N.; Findlay J. B., Structural heterogeneity of Pseudomonas aeruginosa bacterioferritin. *Biochemical Journal* **1994,** *384*, 493-497.

62. Ha, Y.; Shi, D.; Small, G. W.; Theil, E. C.; Allewell, N. M., Crystal structure of bullfrog M ferritin at 2.8 A resolution: analysis of subunit interactions and the binuclear metal center. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* **1999,** *4* (3), 243-56.

63. (a) Nam, K. H.; Xu, Y.; Piao, S.; Priyadarshi, A.; Lee, E. H.; Kim, H.-Y.; Jeon, Y. H.; Ha, N.-C.; Hwang, K. Y., Crystal structure of bacterioferritin from Rhodobacter sphaeroides. *Biochemical and Biophysical Research Communications* **2010,** *391* (1), 990-994;

(b) Cobessi, D.; Huang, L. S.; Ban, M.; Pon, N. G.; Daldal, F.; Berry, E. A., The 2.6 A resolution structure of Rhodobacter capsulatus bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic 'special position'. *Acta crystallographica. Section D, Biological Crystallography* **2002,** *58* (Pt 1), 29-38;

(c) Wahlgren, W. Y.; Omran, H.; von Stetten, D.; Royant, A.; van der Post, S.; Katona, G., Structural characterization of bacterioferritin from Blastochloris viridis. *PloS one* **2012,** *7* (10), e46992.

64. Bernacchioni, C.; Ghini, V.; Pozzi, C.; Di Pisa, F.; Theil, E. C.; Turano, P., Loop electrostatics modulates the intersubunit interactions in ferritin. *ACS Chemical Biology* **2014,** *9* (11), 2517-25.

65. Zhao, G.; Ceci, P.; Ilari, A.; Giangiacomo, L.; Laue, T. M.; Chiancone, E.; Chasteen, N. D., Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli. *The Journal of Biological Chemistry* **2002,** *277* (31), 27689-96.

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

67. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W.; Fan, R.; Sun, X.; Orner, B. P. In *Alanine-shaving mutagenesis to determine key interfacial residues governing the assembly of a nano-cage maxi-ferritin*, *The Journal of Biological Chemistry*, Apr 16; 2010; pp 12078-86.

68. Huard, D. J.; Kane, K. M.; Tezcan, F. A., Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nature Chemical Biology* **2013,** *9* (3), 169-76.

69. Sana, B.; Johnson, E.; Le Magueres, P.; Criswell, A.; Cascio, D.; Lim, S., The Role of Nonconserved Residues of Archaeoglobus fulgidus Ferritin on Its Unique Structure and Biophysical Properties. *The Journal of Biological Chemistry* **2013,** *288* (45), 32663-72.

70. Ahmed, U.; Latham, P. S.; Oates, P. S., Interactions between hepatic iron and lipid metabolism with possible relevance to steatohepatitis. *World journal of Gastroenterology* **2012,** *18* (34), 4651-8.

71. Gammella, E.; Maccarinelli, F.; Buratti, P.; Recalcati, S.; Cairo, G., The role of iron in anthracycline cardiotoxicity. *Frontiers in Pharmacology* **2014,** *5*, 25.

72. Weinberg, E. D., The Lactobacillus anomaly: total iron abstinence. *Perspectives in Biology and Medicine* **1997,** *40* (4), 578-83.

73. Posey, J. E.; Gherardini, F. C., Lack of a role for iron in the Lyme disease pathogen. *Science (New York, N.Y.)* **2000,** *288* (5471), 1651-3.

74. Ferreira, C.; Bucchini, D.; Martin, M. E.; Levi, S.; Arosio, P.; Grandchamp, B.; Beaumont, C., Early embryonic lethality of H ferritin gene deletion in mice. *The Journal of Biological Chemistry* **2000,** *275* (5), 3021-4.

75. Acosta-Cabronero, J.; Cardenas-Blanco, A.; Betts, M. J.; Butryn, M.; Valdes-Herrera, J. P.; Galazky, I.; Nestor, P. J., The whole-brain pattern of magnetic susceptibility perturbations in Parkinson's disease. *Brain : A Journal of Neurology* **2017,** *140* (1), 118-131.

76. Mylonas, C.; Kouretas, D., Lipid peroxidation and tissue damage. *In vivo (Athens, Greece)* **1999,** *13* (3), 295-309.

77. Levi, S.; Salfeld, J.; Franceschinelli, F.; Cozzi, A.; Dorner, M. H.; Arosio, P., Expression and structural and functional properties of human ferritin L-chain from Escherichia coli. *Biochemistry* **1989,** *28* (12), 5179-84.

78.   Le Brun, N. E.; Crow, A.; Murphy, M. E.; Mauk, A. G.; Moore, G. R., Iron core mineralisation in prokaryotic ferritins. *Biochimica et Bbiophysica Acta* **2010,** *1800* (8), 732-44.

79.   Toussaint, L.; Bertrand, L.; Hue, L.; Crichton, R. R.; Declercq, J. P., High-resolution X-ray structures of human apoferritin H-chain mutants correlated with their activity and metal-binding sites. *Journal of Molecular Biology* **2007,** *365* (2), 440-52.

80.   Masuda, T.; Goto, F.; Yoshihara, T.; Mikami, B., Crystal structure of plant ferritin reveals a novel metal binding site that functions as a transit site for metal transfer in ferritin. *The Journal of Biological Chemistry* **2010,** *285* (6), 4049-59.

81.   Arenas-Salinas, M.; Townsend, P. D.; Brito, C.; Marquez, V.; Marabolli, V.; Gonzalez-Nilo, F.; Matias, C.; Watt, R. K.; Lopez-Castro, J. D.; Dominguez-Vera, J.; Pohl, E.; Yevenes, A., The crystal structure of ferritin from Chlorobium tepidum reveals a new conformation of the 4-fold channel for this protein family. *Biochimie* **2014,** *106*, 39-47.

82.   Garcia, P.; Bruix, M.; Rico, M.; Ciofi-Baffoni, S.; Banci, L.; Ramachandra Shastry, M. C.; Roder, H.; de Lumley Woodyear, T.; Johnson, C. M.; Fersht, A. R.; Barker, P. D., Effects of heme on the structure of the denatured state and folding kinetics of cytochrome b562. *Journal of Molecular Biology* **2005,** *346* (1), 331-44.

83.   Nordlund, P.; Sjoberg, B. M.; Eklund, H., Three-dimensional structure of the free radical protein of ribonucleotide reductase. *Nature* **1990,** *345* (6276), 593-8.

84.   Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal from of cyctochrome b1 (Bacterioferritin) from E. coli. *Acta crystallographica. Section D, Biological Crystallography* **1998,** *54* (Pt 1), 16-24.

85.   Crow, A.; Lawson, T. L.; Lewin, A.; Moore, G. R.; Le Brun, N. E., Structural basis for iron mineralization by bacterioferritin. *Journal of the American Chemical Society* **2009,** *131* (19), 6808-13.

86.   Crichton, R. R.; Declercq, J. P., X-ray structures of ferritins and related proteins. *Biochimica et Biophysica Acta* **2010,** *1800* (8), 706-18.

87.   Laghaei, R.; Evans, D. G.; Coalson, R. D., Metal binding sites of human H-chain ferritin and iron transport mechanism to the ferroxidase sites: a molecular dynamics simulation study. *Proteins* **2013,** *81* (6), 1042-50.

88.   Honarmand Ebrahimi, K.; Hagedoorn, P.-L.; Hagen, W. R., Unity in the Biochemistry of the Iron-Storage Proteins Ferritin and Bacterioferritin. *Chemical Reviews* **2014,** *115* (1), 295-326.

89.   Tosha, T.; Ng, H.-L.; Bhattasali, O.; Alber, T.; Theil, E. C., Moving Metal Ions through Ferritin−Protein Nanocages from Three-Fold Pores to Catalytic Sites. *Journal of the American Chemical Society* **2010,** *132* (41), 14562-14569.

90.   Honarmand Ebrahimi, K.; Hagedoorn, P. L.; Hagen, W. R., Phosphate accelerates displacement of Fe(III) by Fe(II) in the ferroxidase center of Pyrococcus furiosus ferritin. *FEBS Letters* **2013,** *587* (2), 220-5.

91.   Frolow, F.; Kalb, A. J.; Yariv, J., Structure of a unique twofold symmetric haem-binding site. *Nature Structural Biology* **1994,** *1* (7), 453-60.

92.   Le Brun, N. E.; Andrews, S. C.; Guest, J. R.; Harrison, P. M.; Moore, G. R.; Thomson, A. J., Identification of the ferroxidase centre of Escherichia coli bacterioferritin. *The Biochemical Journal* **1995,** *312 (Pt 2)*, 385-92.

93.   Quail, M. A.; Jordan, P.; Grogan, J. M.; Butt, J. N.; Lutz, M.; Thomson, A. J.; Andrews, S. C.; Guest, J. R., Spectroscopic and voltammetric characterisation of the bacterioferritin-associated ferredoxin of Escherichia coli. *Biochemical and Biophysical Research Communications* **1996,** *229* (2), 635-42.

94.   Rivera, M., Bacterioferritin: Structure, Dynamics, and Protein-Protein Interactions at Play in Iron Storage and Mobilization. *Accounts of Chemical Research* **2017,** *50* (2), 331-340.

95.   Vieille, C.; Zeikus, G. J., Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews : MMBR* **2001,** *65* (1), 1-43.

96.   Tama, F.; Brooks, C. L., 3rd, The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *Journal of Molecular Biology* **2002,** *318* (3), 733-47.

97.   Teschke, C. M.; McGough, A.; Thuman-Commike, P. A., Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophys Journal* **2003,** *84* (4), 2585-92.

98.   Lewin, A.; Moore, G. R.; Le Brun, N. E., Formation of protein-coated iron minerals. *Dalton Transactions (Cambridge, England : 2003)* **2005,** (22), 3597-610.

99.   Rana, S.; Yeh, Y. C.; Rotello, V. M., Engineering the nanoparticle-protein interface: applications and possibilities. *Current Opinion in Chemical Biology* **2010,** *14* (6), 828-34.

100.   Andrews, S. C.; Le Brun, N. E.; Barynin, V.; Thomson, A. J.; Moore, G. R.; Guest, J. R.; Harrison, P. M., Site-directed replacement of the coaxial heme ligands of bacterioferritin generates heme-free variants. *The Journal of Biological Chemistry* **1995,** *270* (40), 23268-74.

101. Wong, S. G.; Abdulqadir, R.; Le Brun, N. E.; Moore, G. R.; Mauk, A. G., Fe-haem bound to Escherichia coli bacterioferritin accelerates iron core formation by an electron transfer mechanism. *The Biochemical Journal* **2012,** *444* (3), 553-60.

102. Yasmin, S.; Andrews, S. C.; Moore, G. R.; Le Brun, N. E., A new role for heme, facilitating release of iron from the bacterioferritin iron biomineral. *The Journal of Biological Chemistry* **2011,** *286* (5), 3473-83.

103. Lee, J. H.; Song, C.; Kim, D. H.; Park, I. H.; Lee, S. G.; Lee, Y. S.; Kim, B. G., Glutamine (Q)-peptide screening for transglutaminase reaction using mRNA display. *Biotechnology and Bioengineering* **2013,** *110* (2), 353-62.

104. (a) Hallborn, J.; Carlsson, R., Automated screening procedure for high-throughput generation of antibody fragments. *BioTechniques* **2002,** *Suppl*, 30-7;

      (b) Bazan, J.; Całkosiński, I.; Gamian, A., Phage display--a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Human Vaccines & Immunotherapeutics* **2012,** *8* (12), 1817-28;

      (c) Adey, N. B.; Mataragnon, A. H.; Rider, J. E.; Carter, J. M.; Kay, B. K., Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene* **1995,** *156* (1), 27-31;

      (d) Frenzel, A.; Schirrmann, T.; Hust, M., Phage display-derived human antibodies in clinical development and therapy. *mAbs* **2016,** *8* (7), 1177-1194;

      (e) Loset, G. A.; Bogen, B.; Sandlie, I., Expanding the versatility of phage display I: efficient display of peptide-tags on protein VII of the filamentous phage. *PloS One* **2011,** *6* (2), e14702.

105. Rabuka, D., Chemoenzymatic methods for site-specific protein modification. *Current Opinion in Chemical Biology* **2010,** *14* (6), 790-6.

106. Genz, M.; Strater, N., Posttranslational incorporation of noncanonical amino acids in the RNase S system by semisynthetic protein assembly. *Methods in Molecular Biology (Clifton, N.J.)* **2014,** *1216*, 71-87.

107. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B., N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angewandte Chemie International Edition* **2006,** *45* (32), 5307-5311.

108. Ducharme, J.; Auclair, K., Use of bioconjugation with cytochrome P450 enzymes. *Biochimica et biophysica acta. Proteins and Proteomics* **2018,** *1866* (1), 32-51.

109. Hermanson, G. T., Chapter-1 Functional Targets; Chapter-2 the Chemistry of Reactive Groups. In *Bioconjugate Techniques,* Academic Press, **1996**, 1-137.

110. van Vught, R.; Pieters, R. J.; Breukink, E., Site-specific functionalization of proteins and their applications to therapeutic antibodies. *Computational and Structural Biotechnology Journal* **2014,** *9*, e201402001.

111. Soundrarajan, N.; Sokalingam, S.; Raghunathan, G.; Budisa, N.; Paik, H. J.; Yoo, T. H.; Lee, S. G., Conjugation of proteins by installing BIO-orthogonally reactive groups at their N-termini. *PloS One* **2012,** *7* (10), e46741.

112. Khoshnejad, M.; Greineder, C. F.; Pulsipher, K. W.; Villa, C. H.; Altun, B.; Pan, D. C.; Tsourkas, A.; Dmochowski, I. J.; Muzykantov, V. R., Ferritin Nanocages with Biologically Orthogonal Conjugation for Vascular Targeting and Imaging. *Bioconjugate Chemistry* **2018,** *29* (4), 1209-1218.

113. Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V., Antibody conjugates with unnatural amino acids. *Molecular Pharmaceutics* **2015,** *12* (6), 1848-62.

114. Laulhere, J.P.; Laboure, A.M.; Wuytswinkel, O.V.; Gagnon, J.; Briat, J.F., Purification, characterization and function of bacterioferritin from the cyanobacterium Synechocystis PCC 6803. *Biochemical Journal* **1992,** *281*, 785-793.

115. Hudson, A. J.; Andrews, S. C.; Hawkins, C.; Williams, J. M.; Izuhara, M.; Meldrum, F. C.; Mann, S.; Harrison, P. M.; Guest, J. R., Overproduction, purification and characterization of the Escherichia coli ferritin. *European Journal of Biochemistry* **1993,** *218* (3), 985-95.

116. Jung, Y.; Kwak, J.; Lee, Y., High-level production of heme-containing holoproteins in Escherichia coli. *Applied Microbiology and Biotechnology* **2001,** *55* (2), 187-91.

117. Truffi, M.; Fiandra, L.; Sorrentino, L.; Monieri, M.; Corsi, F.; Mazzucchelli, S., Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacological Research* **2016,** *107*, 57-65.

118. Mènard, N.; Tsapis, N.; Poirier, C.; Arnauld, T.; Moine, L.; Lefoulon, F.; Pèan, J. M.; Fattal, E., Drug solubilization and in vitro toxicity evaluation of lipoamino acid surfactants. *International Journal of Pharmaceutics* **2012,** *423* (2), 312-20.

119. Hu, Y., Samanta, D., Parelkar, S., Hong, SW., Wang, Q., Russell, T., Emrick, T. , Ferritin-polymer conjugates: Grafting chemistry and integration into nanoscale assemblies. *Advanced Function Materials* **2010,** *20*, 3603-3612.

120. van der Ven, A., Protein Engineering of Bacterioferritin: Applications to Bionanotechnology. M.Sc. Thesis, University of Waterloo, Canada, ON, **2015**.

121. Krayem, O. B. Hydrophobic Engineering of a Bacterial Nanodimensional Capsule Protein, M.Sc. Thesis, University of Waterloo, Canada, ON, **2017**.

122. Suttisansanee, U., Biochemistry in Bacterioferritin, M.Sc. Thesis, University of Waterloo, Canada, ON, **2006**.

123. Schrödinger *Schrödinger Release 2017-4: Maestro, Schrödinger, LLC, New York, NY, 2017.*, New York, NY, **2017**.

124. Neumann, U.; Hofsteenge, J., Interaction of semisynthetic variants of RNase A with ribonuclease inhibitor. *Protein Science : A Publication of the Protein Society* **1994,** *3* (2), 248-56.

125. Raines, R. T.; McCormick, M.; Van Oosbree, T. R.; Mierendorf, R. C., The S.Tag fusion system for protein purification. *Methods in Enzymology* **2000,** *326*, 362-76.

126. Chen, Q.; Sun, Q.; Molino, N. M.; Wang, S. W.; Boder, E. T.; Chen, W., Sortase A-mediated multi-functionalization of protein nanoparticles. *Chemical Communications (Cambridge, England)* **2015,** *51* (60), 12107-10.

127. Skala, W.; Goettig, P.; Brandstetter, H., Do-it-yourself histidine-tagged bovine enterokinase: a handy member of the protein engineer's toolbox. *Journal of biotechnology* **2013,** *168* (4), 421-5.

128. Ohtsuka, T.; Ota, M.; Nio, N.; Motoki, M., Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors. *Bioscience, Biotechnology, and Biochemistry* **2000,** *64* (12), 2608-13.

129. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A., ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **2003,** *31* (13), 3784-8.

130. Calisti, L.; Benni, I.; Cardoso Trabuco, M.; Baiocco, P.; Ruzicka, B.; Boffi, A.; Falvo, E.; Malatesta, F.; Bonamore, A., Probing bulky ligand entry in engineered archaeal ferritins. *Biochimica et Biophysica Acta* **2017,** *1861* (2), 450-456.

131. Tatur, J.; Hagedoorn, P. L.; Overeijnder, M. L.; Hagen, W. R., A highly thermostable ferritin from the hyperthermophilic archaeal anaerobe Pyrococcus furiosus. *Extremophiles : Life under Extreme Conditions* **2006,** *10* (2), 139-48.

132. Antonowicz, I., The role of enteropeptidase in the digestion of protein and its development in human fetal small intestine. *Ciba Foundation Symposium* **1979,** (70), 169-87.

133. Light, A.; Janska, H., Enterokinase (enteropeptidase): comparative aspects. *Trends in Biochemical Sciences* **1989,** *14* (3), 110-2.

134. Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **2003,** *60* (5), 523-33.

135. Shahravan, S. H.; Qu, X.; Chan, I. S.; Shin, J. A., Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. *Protein Expression and Purification* **2008,** *59* (2), 314-9.

136. Lu, D.; Futterer, K.; Korolev, S.; Zheng, X.; Tan, K.; Waksman, G.; Sadler, J. E., Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. *Journal of Molecular Biology* **1999,** *292* (2), 361-73.

137. Biolabs, N. E. Protocol for use with Enterokinase, light chain (P8070) https://international.neb.com/protocols/2018/04/05/protocol-for-use-with-enterokinase-light-chain-p8070.

138. Light, A.; Savithri, H. S.; Liepnieks, J. J., Specificity of bovine enterokinase toward protein substrates. *Analytical Biochemistry* **1980,** *106* (1), 199-206.

139. Stellwagen, E.; Rysavy, R.; Babul, G., The conformation of horse heart apocytochrome c. *The Journal of Biological Chemistry* **1972,** *247* (24), 8074-7.

140. Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W., Site-specific protein labelling and immobilization mediated by microbial transglutaminase. *Chemical Communications (Cambridge, England)* **2014,** *50* (50), 6604-6.

141. Zhu, Y.; Tramper, J., Novel applications for microbial transglutaminase beyond food processing. *Trends in Biotechnology* **2008,** *26* (10), 559-65.

142. Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature's biological glues. *The Biochemical Journal* **2002,** *368* (Pt 2), 377-96.

143. GRoss, E., Sequencing of proteins and peptides. *Methods in Enzymology* **1967,** *11*, 238-255.

144. Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F., Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.)* **1999,** *112*, 531-52.

145. Zhengding Su, S. Z., Apraku D. Donkor, Costas Tzoganakis, and John F Honek, Controllable Delivery of Small-Molecule Compounds to Targeted Cells Utilizing Carbon Nanotubes. *Journal of the American Chemical Society* **2011,** *133* (18), 6874-6877.

146. Richards, F. M.; Vithayathil, P. J., The preparation of subtilisn-modified ribonuclease and the separation of the peptide and protein components. *The Journal of Biological Chemistry* **1959,** *234* (6), 1459-65.

147. Heim, R.; Cubitt, A. B.; Tsien, R. Y., Improved green fluorescence. *Nature* **1995,** *373* (6516), 663-4.

148.   Schneider, C. A., Rasband, W.S., Eliceiri, K.W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **2012,** *9*, 671-675.

149.   Inouye, S.; Tsuji, F. I., Evidence for redox forms of the Aequorea green fluorescent protein. *FEBS Letters* **1994,** *351* (2), 211-4.

150.   Riddles, P. W.; Blakeley, R. L.; Zerner, B., Ellman's reagent: 5,5'-dithiobis(2-nitrobenzoic acid)-A Reexamination. *Analytical Biochemistry* **1979,** *94* (1), 75-81.

151.   Yang, X.; Arosio, P.; Chasteen, N. D., Molecular diffusion into ferritin: pathways, temperature dependence, incubation time, and concentration effects. *Biophys Journal* **2000,** *78* (4), 2049-59.

152.   Douglas, T.; Young, M., Viruses: making friends with old foes. *Science (New York, N.Y.)* **2006,** *312* (5775), 873-5.

153.   Su, Z.; Wang, Q., A hierarchical assembly process to engineer a hydrophobic core for virus-like particles. *Angewandte Chemie (International ed. in English)* **2010,** *49* (52), 10048-50.

154.   Kwak, M.; Minten, I. J.; Anaya, D. M.; Musser, A. J.; Brasch, M.; Nolte, R. J.; Mullen, K.; Cornelissen, J. J.; Herrmann, A., Virus-like particles templated by DNA micelles: a general method for loading virus nanocarriers. *Journal of the American Chemical Society* **2010,** *132* (23), 7834-5.

155.   Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R., A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Analytical Biochemistry* **1999,** *273* (1), 73-80.

156.   (a) Han, J. C.; Han, G. Y., A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Analytical Biochemistry* **1994,** *220* (1), 5-10;

(b)   Krezel, A.; Latajka, R.; Bujacz, G. D.; Bal, W., Coordination properties of tris(2-carboxyethyl)phosphine, a newly introduced thiol reductant, and its oxide. *Inorganic Chemistry* **2003,** *42* (6), 1994-2003;

(c)   Shen, W.; Zhong, H.; Neff, D.; Norton, M. L., NTA directed protein nanopatterning on DNA Origami nanoconstructs. *Journal of the American Chemical Society* **2009,** *131* (19), 6660-1.

157.   Kantner, T.; Watts, A. G., Characterization of Reactions between Water-Soluble Trialkylphosphines and Thiol Alkylating Reagents: Implications for Protein-Conjugation Reactions. *Bioconjugate Chemistry* **2016,** *27* (10), 2400-2406.

158. Barbara, M.; Thomas, R.; Jean-Christophe, R.; Laurent, G.; Laurent, V.; Robert, P., Reduction with tris(2-carboxyethyl)phosphine (TCEP) enables the use of an S-sulphonate protecting group for thiol-mediated bioconjugation. *Royal Society of Chemistry Advances,* **2014,** (4,), 7725-7728

159. Sanchez, A.; Pedroso, E.; Grandas, A., Oligonucleotide cyclization: the thiol-maleimide reaction revisited. *Chemical communications (Cambridge, England)* **2013,** *49* (3), 309-11.

160. E. Hedaya, S. T., The preparation and reactions of stable phosphorus ylides derived from maleic anhydrides, maleimides or isomaleimides *Tetrahedron* **1968,** *24*, 2241-2254.

161. Holmberg, A.; Blomstergren, A.; Nord, O.; Lukacs, M.; Lundeberg, J.; Uhlen, M., The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **2005,** *26* (3), 501-10.

162. Meir Wilchek, E. A. B., [2] Introduction to avidin-biotin technology. *Methods in Enzymology* **1990,** *184*, 5-13.

163. Schrodinger https://www.schrodinger.com/sites/default/files/maestro_11_qrc.pdf.

164. RCSB http://www.rcsb.org. http://www.rcsb.org.

165. Wermuth, C.G.; Ganellin C. R.; Lindberg P.; L.A. Mitscher, L. A., Glossary Of Terms Used In Medicinal Chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry* **2009,** *70(5)*, 1129-1143.

166. Loving, K.; Salam, N. K.; Sherman, W., Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *Journal of Computer-aided Molecular Design* **2009,** *23* (8), 541-54.

167. Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A., PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *Journal of Computer-aided Molecular Design* **2006,** *20* (10-11), 647-71.

168. Schrödinger *Schrödinger Release 2017-4: Phase, Schrödinger, LLC, New York, NY, **2017***.

169. Schrödinger *Schrödinger Release 2017-4: Glide, Schrödinger, LLC, New York, NY, **2017***.

170. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1750-9.

171. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry* **2006,** *49* (21), 6177-96.

172.	Schrödinger Accounting for Protein Flexibility. https://www.schrodinger.com/system/files/protein_flexibility_2019-1.pdf.

173.	Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry* **2006,** *49* (2), 534-53.

174.	Jiang, F.; Kim, S. H., "Soft docking": matching of molecular surface cubes. *Journal of Molecular Biology* **1991,** *219* (1), 79-102.

175.	Carlson, H. A.; McCammon, J. A., Accommodating protein flexibility in computational drug design. *Molecular Pharmacology* **2000,** *57* (2), 213-8.

176.	Kua, J.; Zhang, Y.; McCammon, J. A., Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *Journal of the American Chemical Society* **2002,** *124* (28), 8260-7.

177.	Antunes, D. A.; Devaurs, D.; Kavraki, L. E., Understanding the challenges of protein flexibility in drug design. *Expert Opinion on Drug Discovery* **2015,** *10* (12), 1301-13.

178.	Schrödinger *Schrödinger Release **2017-4**: Prime, Schrödinger, LLC, New York, NY, 2017*, New York, 2017.

179.	Sherman, W.; Beard, H. S.; Farid, R., Use of an induced fit receptor structure in virtual screening. *Chemical Biology & Drug Design* **2006,** *67* (1), 83-4.

180.	Jorgensen, W. L.; Tirado-Rives, J., The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **1988,** *110* (6), 1657-66.

181.	Scott J. Weiner, P. A. K., David A. Case, U. Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **1984,** *106* (3), 765-784.

182.	MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, K.; *et al*., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry* **1998,** *102* (18), 3586-3616.

183.	William L. Jorgensen, David S. Maxwell, and, and Julian Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996,** *118* (45), 11225-11236.

184. George A. Kaminski and, R. A. F., Julian Tirado-Rives and, and William L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

185. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry* **2005,** *26* (16), 1752-80.

186. Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W., Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *Journal of Chemical Theory and Computation* **2012,** *8* (8), 2553-8.

187. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory andComputation* **2016,** *12* (1), 281-96.

188. Patsahan, T.; Ilnytskyi, J. M.; Pizio, O., On the properties of a single OPLS-UA model curcumin molecule in water, methanol and dimethyl sulfoxide. Molecular dynamics computer simulation results. *Condensed Matter Physics,* **2017,** *20* (23003), 1-20.

189. Schrödinger *http://content.schrodinger.com/Training+Material/Glide/Structure-Based_Virtual_Screening_with_Glide_2016-3.pdf*, 2016.

190. Schrödinger Rigid Receptor Docking with Glide. (accessed 20/02/2018).

191. Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C., Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of Computer-aided Molecular Design* **2010,** *24* (6-7), 591-604.

192. Mitsumori, L. M.; Bhargava, P.; Essig, M.; Maki, J. H., Magnetic resonance imaging using gadolinium-based contrast agents. *Topics in MagneticResonance Imaging : TMRI* **2014,** *23* (1), 51-69.

193. Berginc, K.; Trontelj, J.; Basnet, N. S.; Kristl, A., Physiological barriers to the oral delivery of curcumin. *Die Pharmazie* **2012,** *67* (6), 518-24.

194. Anand, P.; Kunnumakkara, A. B.; Newman, R. A.; Aggarwal, B. B., Bioavailability of curcumin: problems and promises. *Molecular Pharmaceutics* **2007,** *4* (6), 807-18.

195. Cardona, A. F.; Rojas, L.; Wills, B.; Ruiz-Patino, A.; Abril, L.; Hakim, F.; Jimenez, E.; Useche, N.; Bermudez, S.; Mejia, J. A.; Ramon, J. F.; Carranza, H.; Vargas, C.; Otero, J.; Archila, P.; Rodriguez, J.; Rodriguez, J.; Behaine, J.; Gonzalez, D.; Jacobo, J.; Cifuentes, H.; Feo, O.; Penagos, P.; Pineda, D.; Ricaurte, L.; Pino, L. E.; Vargas, C.; Marquez, J. C.; Mantilla, M. I.; Ortiz, L. D.; Balana, C.; Rosell, R.; Zatarain-Barron, Z. L.; Arrieta, O., A comprehensive analysis of factors related to carmustine/bevacizumab response in recurrent glioblastoma. *Clinical & Translational Oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **2019**.

196. Tyler, M. W.; Zaldivar-Diez, J.; Haggarty, S. J., Classics in Chemical Neuroscience: Haloperidol. *ACS Chemical Neuroscience* **2017,** *8* (3), 444-453.

197. Zhen, Z.; Tang, W.; Chen, H.; Lin, X.; Todd, T.; Wang, G.; Cowger, T.; Chen, X.; Xie, J., RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors. *ACS Nano* **2013,** *7* (6), 4830-7.

198. Ramsay, R. R.; Popovic-Nikolic, M. R.; Nikolic, K.; Uliassi, E.; Bolognesi, M. L., A perspective on multi-target drug discovery and design for complex diseases. *Clinical and Translational Medicine* **2018,** *7* (1), 3.

199. Bolognesi, M. L., Polypharmacology in a single drug: multitarget drugs. *Current Medicinal Chemistry* **2013,** *20* (13), 1639-45.

200. Rusnati, M.; Oreste, P.; Zoppetti, G.; Presta, M., Biotechnological engineering of heparin/heparan sulphate: a novel area of multi-target drug discovery. *Current Pharmaceutical Design* **2005,** *11* (19), 2489-99.

201. Lin, X.; Xie, J.; Niu, G.; Zhang, F.; Gao, H.; Yang, M.; Quan, Q.; Aronova, M. A.; Zhang, G.; Lee, S.; Leapman, R.; Chen, X., Chimeric ferritin nanocages for multiple function loading and multimodal imaging. *Nano Letters* **2011,** *11* (2), 814-9.

TABLE OF CONTEXT-GLOSSARY

AfFtn: *Archaeoglobus fulgidus* ferritin

Ala: Alanine

Arg: Arginine

Asn: Asparagine

Asp: Aspartate

Amu: Atomic mass unit

Au: Absorbance units

AUC: Analytical Ultracentrifugation

Bfr: Bacterioferritin

Biotin-XNTA: Biotin nitrilotriacetic acid

BfMF: Bullfrog middle chain ferritin

C14-NTA: Myristol-Nitrilotriacetic acid

C16-NTA: Palmitoyl-Nitrilotriacetic acid

C19-NTA: Nanodecanoyl-Nitrilotriacetic acid

CCMV: Cowpea chlorotic mottle virus

CtFtn: *Chlorobium tepidum* ferritin

CMC: Critical Micelle Concentration

Cys: Cysteine

DLS: Dynamic light scattering

DGS-NTA: 1, 2-dioleoyl-*sn*-glycero-3- [(N-(5-amino-1 carboxypentyl) iminodiaceticacid) succinyl]

DMF: Dimethylformamide

DMSO: Dimethyl Sulfoxide

Dps: DNA-binding proteins from starved cells

EDC: N-(3-dimethylaminopropyl-N'-ethyl)-carbodiimide

EPR: Enhanced permeability and retention

ESI: Electrospray ionization

Equiv: Equivalents

FRET: Förster resonance energy transfer

FITC: Fluorescein isothiocyanate

Ftn: Ferritin

GFP: Green Fluorescent Protein

Gln: Glutamine

Glu: Glutamate

Gly: Glycine

Glide SP: Glide Standard Precision

Glide XP: Glide Extra Precision

GdnHCl: Guanidine HCl

His: Histidine

His$_6$-tag: Hexahistidine affinity tag

HisGFP: His-tagged enhanced GFP

HoSF: Horse spleen ferritin

HPV: Human Papillomavirus

HRP: Horseradish peroxidase

Hsp: Heat-shock protein

HuHF: Human heavy chain ferritin

HuLF: Human light chain ferritin

Ile: Isoleucine

IFD: Induced Fit Docking

IMAC: immobilized metal affinity chromatography

kDa: kiloDalton

keV: kiloelectron volt

Leu: Leucine

LigPrep: Ligand Preparation

Lys: Lysine

Lys-NTA: Nα, Nα-Bis(carboxymethyl)-L-lysine

MRI: Magnetic Resonance Imaging

Met: Methionine

MetAP: Bacterial methionyl aminopeptidase

µL: microliters

MTD: Multi target drug

mM: millimolar

MW: Molecular Weight

NBA: N-Butylamine

Native PAGE: Native polyacrylamide gel electrophoresis

nM: nanomolar

nm: nanometer

NCL: Native chemical ligation

NHS: N-Hydroxysuccinimide

NP: Nanoparticle

NTA: Nitrilotriacetic acid

PEG: Polyethylene glycol

Phe: Phenylalanine

Psi: Pounds per square inch

Pro: Proline

Prep Wizard: Preparation Wizard

PfFtn: *Pyrococcus furiosus* ferritin

QD: Quantum dots

QTag: Glutamine tag

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC: Size exclusion chromatography

Ser: Serine

TEM: Transmission electron microscopy

Thr: Threonine

Trp: Tryptophan

Tyr: Tyrosine

Val: Valine

VLP: Virus-like protein

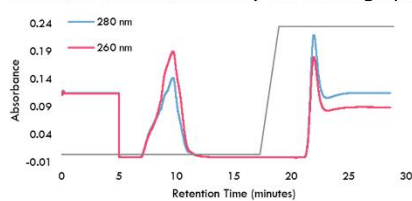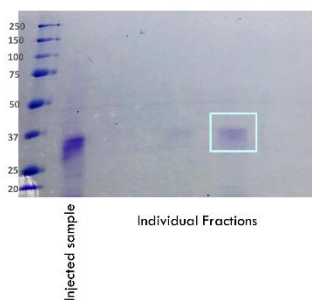Monitoring Purification by SDS- PAGE
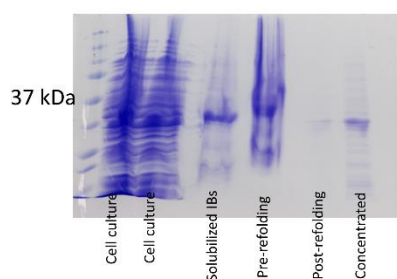
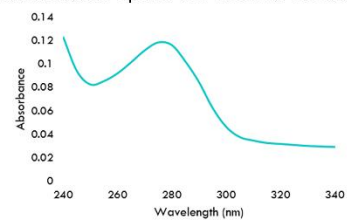Negative Ion Exchange Chromatography
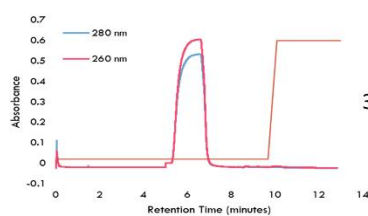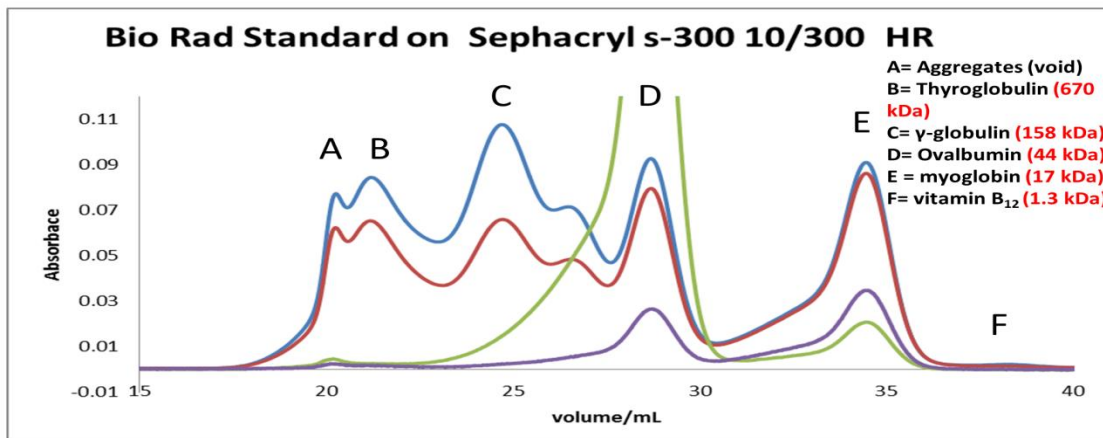
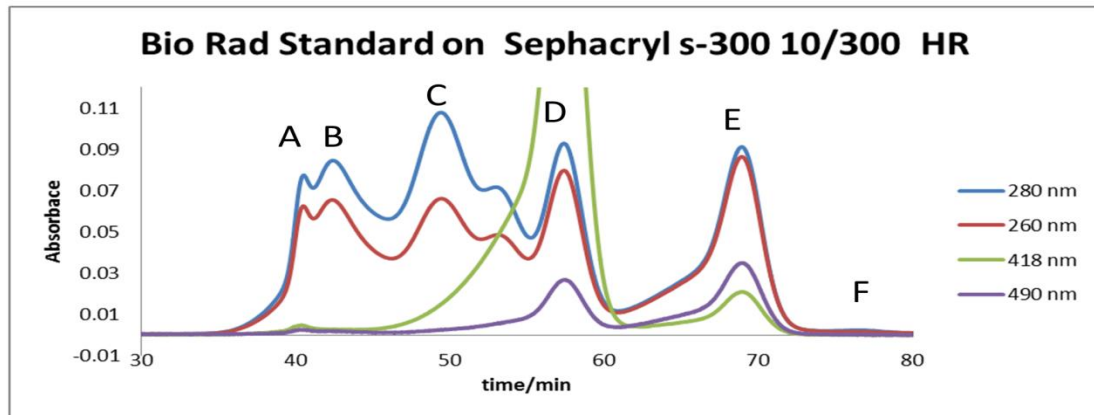Testing IEC Fractions
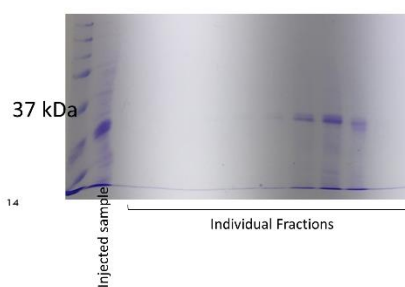
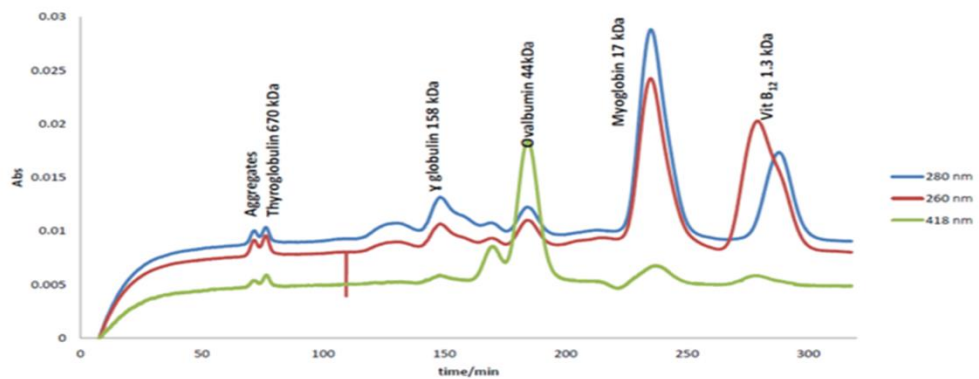Immobilized Metal Affinity Chromatography

Testing IMAC Fractions

Absorbance Spectrum of Final Product

Appendix 1. Purification and Characterization of EK_C122S_His5. Source: Evan Shepherdson

Appendix 2.SEC calibration profiles, using Sephacryl™ 300 10/300 HR and Sephacryl™ 300 26/60HR. calibration plots was prepared by running BioRad Gel filtration standards (#1511109 containing: 5 mg thyroglobulin- 670,000 Da, 5 mg g-globulin-158,000, 5 mg ovalbumin-44,000 , 2.5 mg myoglobin-17,000 Da, and 0.5 vitamin B12-1,350 Da)

227
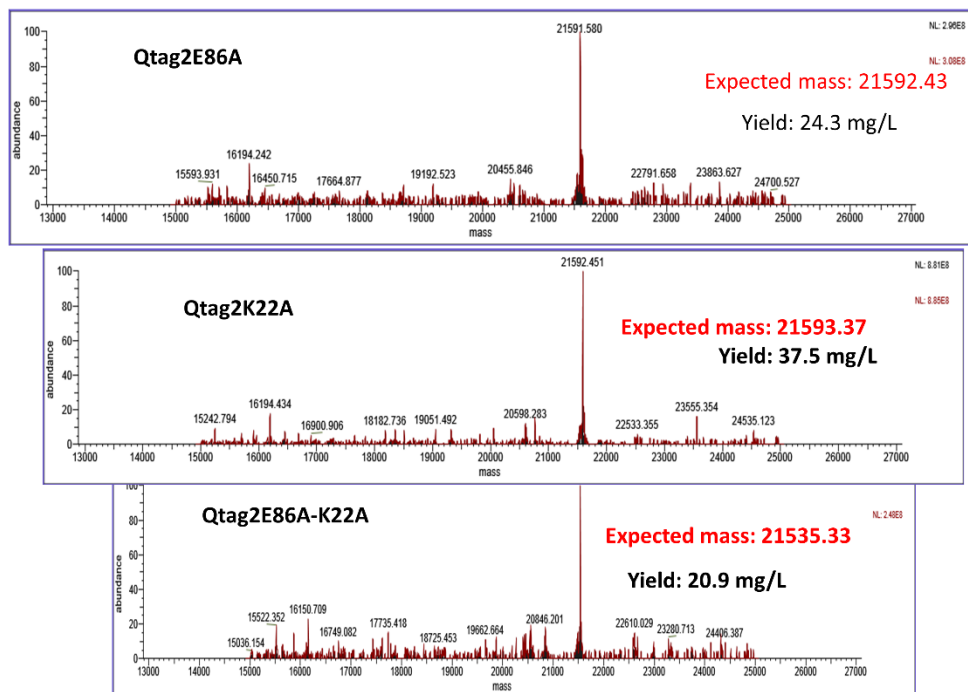
Appendix 3. +ESI data of Qtag2E86A, Qtag2K22A, and Qtag2E86A-K22.
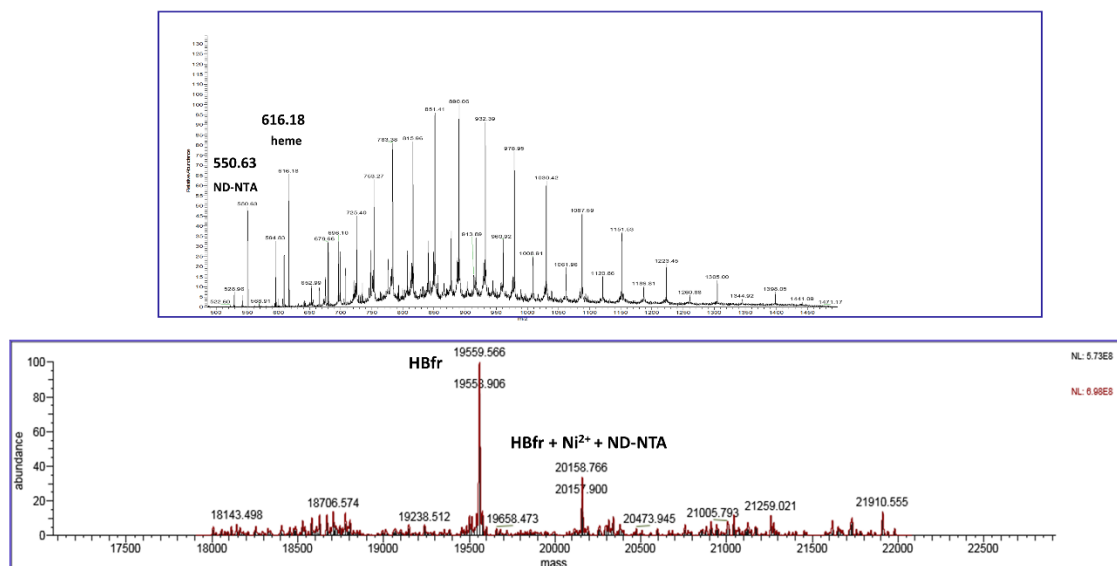
Appendix 4. +ESI data for HisBfr-Ni^{2+} - C19-NTA using 1 mg/mL of C19-NTA.
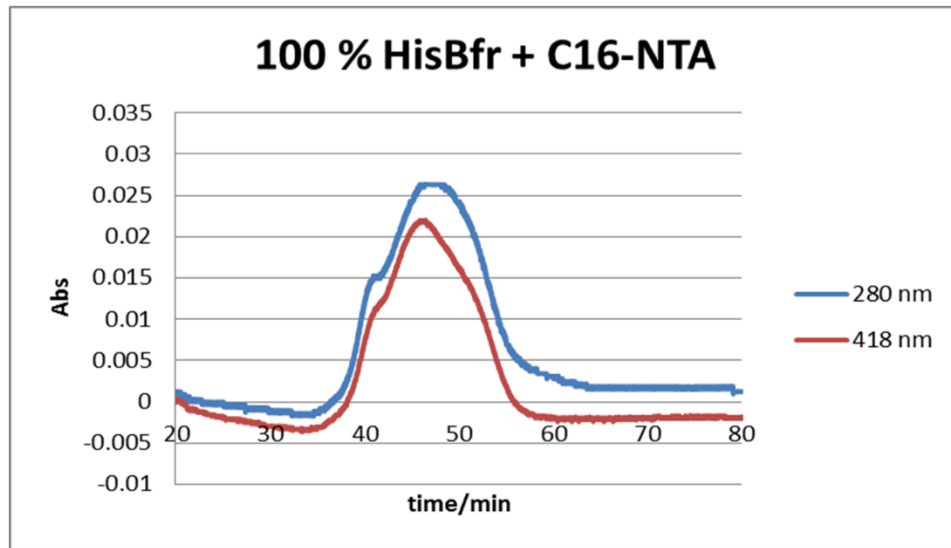
Appendix 5. SEC profiles of C14-NTA and C16-NTA encapsulated into 100 % HisBfr. The C14-NTA encapsulation resulted in identical SEC profile.
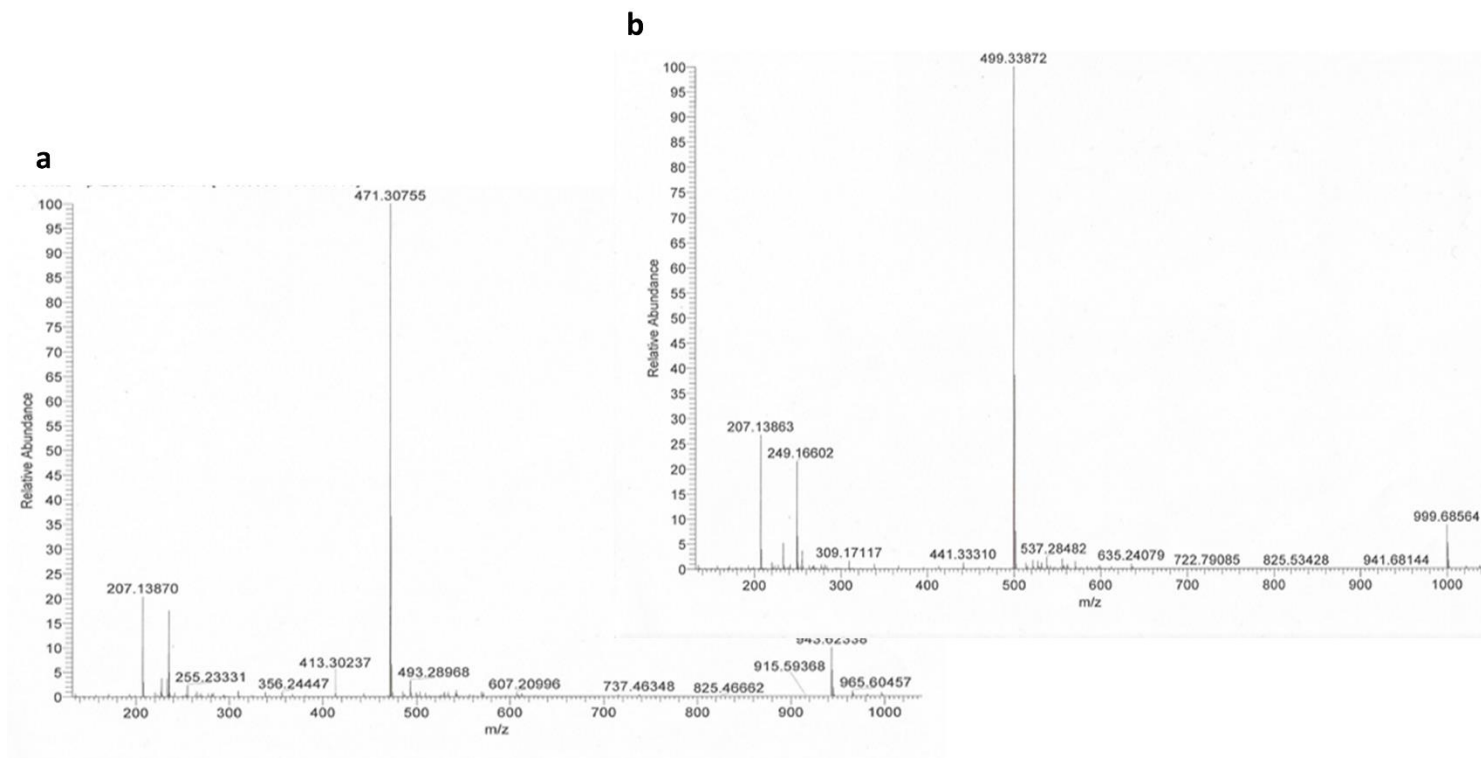
Appendix 6. Mass Spectrometry results for synthesized NTA-fatty acids. (a) Myristoyl-NTA showing the expected 471 Da peak. Accurate mass data m/z: calculated for C24H43N2O7: 471.3075; Found: 471.3076. (b) Palmitoyl-NTA with the expected peak at 499 Da. Accurate mass data m/z: calculated for C26H47N2O7: 499.3387; Found: 499.3389. Source: April Marple

"C14_NTA Feb5" 1 1 "C:\Users\marp\OneDrive\Documents\494\494\NMR\NMR_Data\C14_NTA Feb5"
Myristoyl-NTA in DMSO (Made Jan 16)

a

"C16_NTA (C) Feb5" 1 1 "C:\Users\marp\OneDrive\Documents\494\494\NMR\Tuesday_C13\C16_NTA (C) Feb5"
Palmitoyl-NTA in DMSO (Made Jan 11)

b
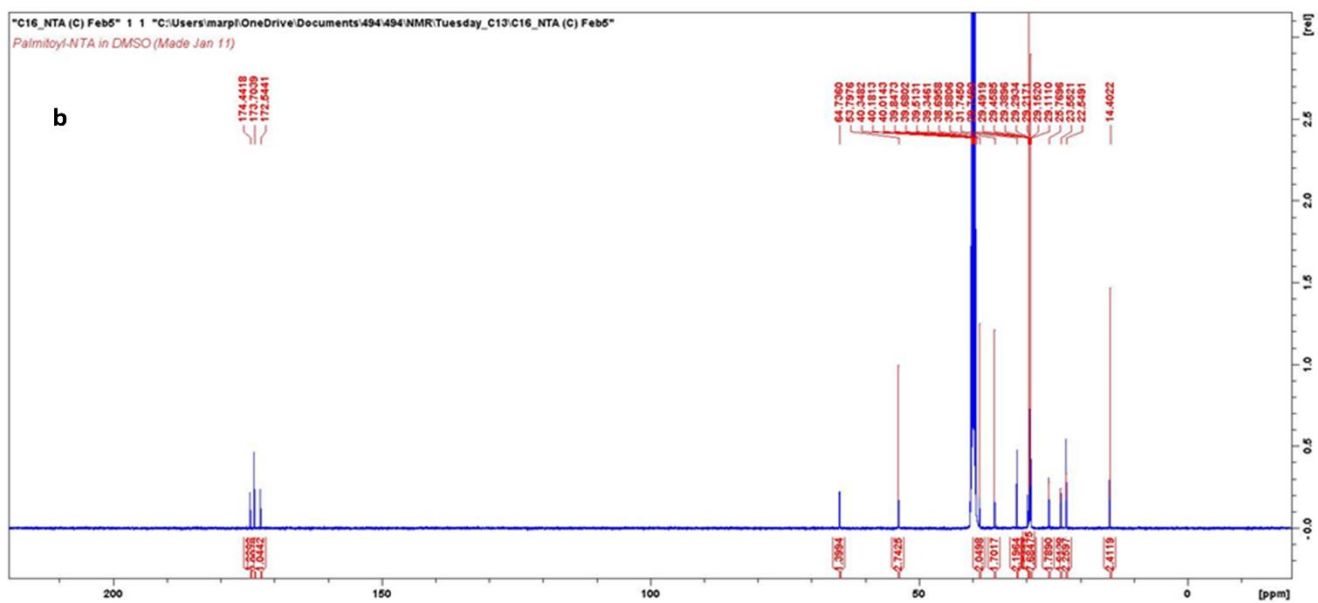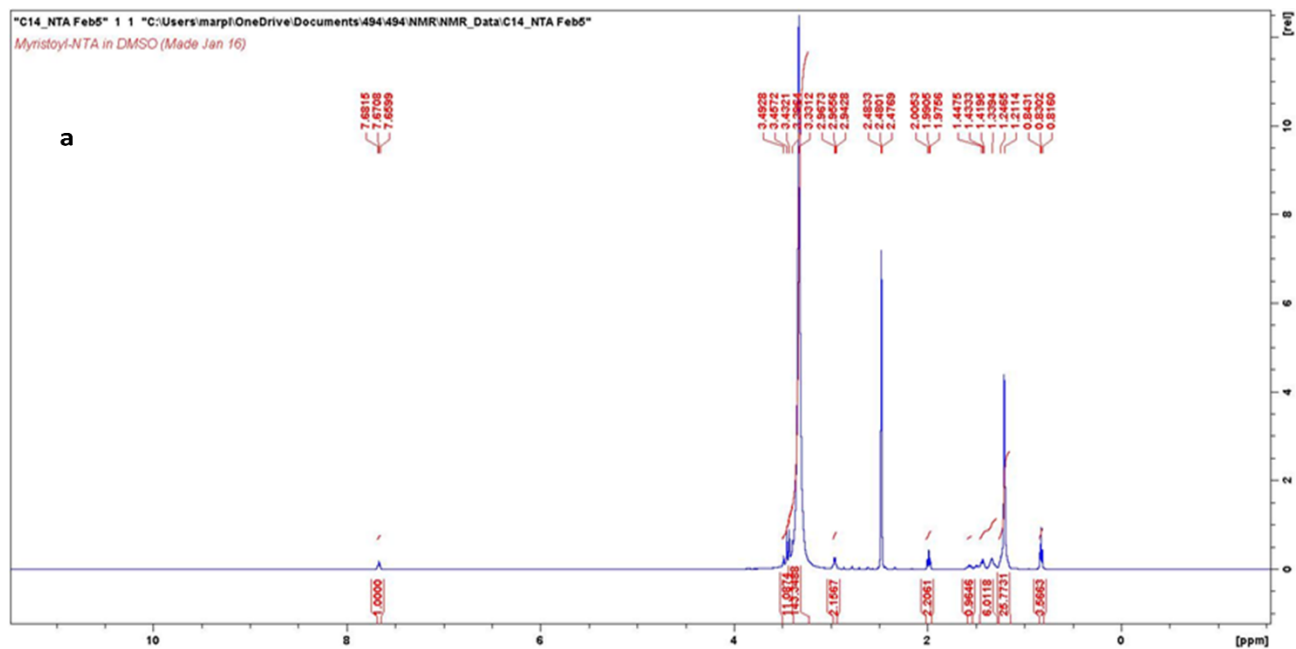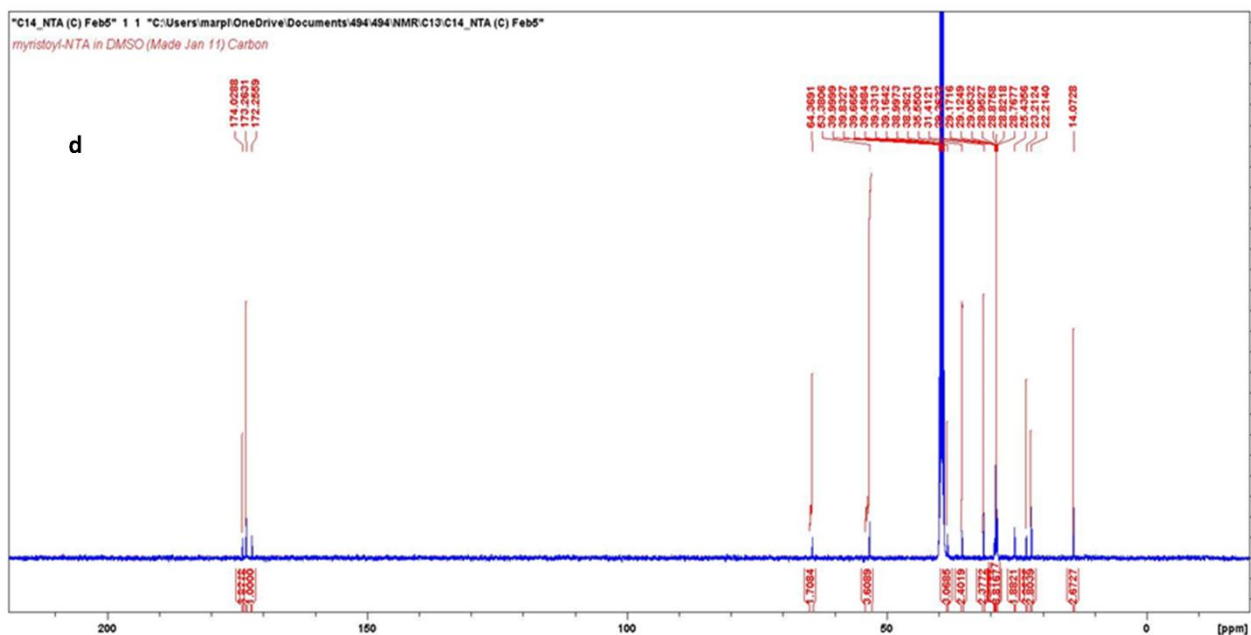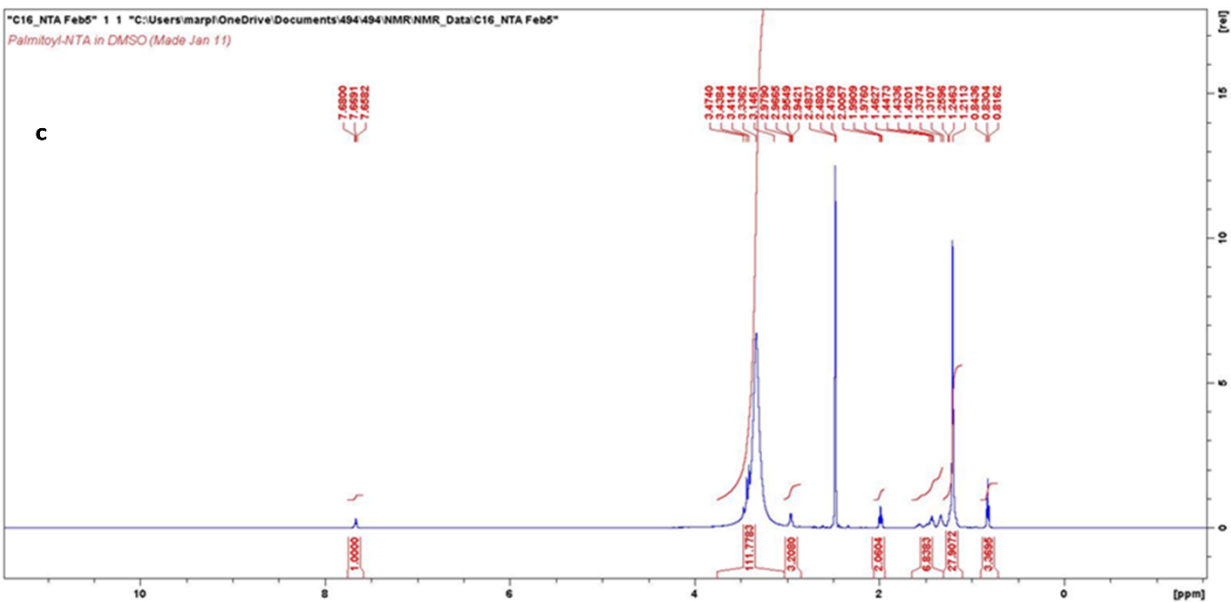
Appendix 7. NMR Results. (a) ^{1}H-NMR of synthesized C14-NTA characterized as follows: (500 MHz, DMSO) δ 7.67 (t, 1H, J= 5.35 Hz), 3.49-3.39 (m, 5H), 2.95 (dd, 2H, J = 6.4 Hz, J = 12.25 Hz), 1.99 (t, 2H, J = 7.45), 1.56 (q, 2H, J = 7.15), 1.43 (m, 4H), 1.24 – 1.21 (m, 22H), 0.83 (t, 3H, J = 6.45).

(b) ^{1}H-NMR of C16-NTA results were (500 MHz, DMSO) δ 7.66 (t, 1H, J = 5.45), 3.47 (m, 5H), 2.95 (dd, 2H, J = 6.4, J = 12.05), 1.99 (t, 2H, J = 7.45), 1.58 (q, 2H, J = 7.7), 1.43 (m, 4H), 1.33 (t, 6H, J = 7.9), 1.24 – 1.21 (m, 20H), 0.83 (t, 3H, J = 7.1). As well as

(c) ^{13}C-NMR for C14-NTA results were (125 MHz, DMSO) δ 174.0 (s), 173.2 (s), 172.2 (s), 64.3 (d), 53.3 (t), 38.3 (t), 35.5 (t), 31.4 (t), 29.3 (t), 29.1 – 28.7 (t), 25.4 (t), 23.2 (t), 22.2 (t), 14.0 (q). and

(d) ^{13}C-NMR for C16-NTA characterized as: (125 MHz, DMSO) δ 174.4 (s), 173.7 (s), 172.5 (s), 64.7 (d), 53.7 (t), 38.6 (t), 35.8 (t), 31.7 (t), 29.4 - 29.1 (t), 25.7 (t), 23.5 (t), 22.5 (t), 14.4 (q). The solvent peak (not integrated) appears at 2.5 ppm for the ^{1}H-NMR and at 39-40 ppm for the ^{13}C-NMR. Source: April Marple

Construct	DNA Sequence
WTBfr	NNNNNgNNcgtaNcaTTcCnCTCTaganaTAATTTTGTTtaACTTTAAGAAGG AGATATACATATGAAAGGTGATACTAAAGTTATAAATTATCTCAACAA ACTGTTGGGAAATGAGCTTGTCGCAATCAATCAGTACTTTCTCCATGC CCGAATGTTTAAAAACTGGGGTCTCAAACGTCTCAATGATGTGGAGTA TCATGAATCCATTGATGAGATGAAACACGCCGATCGTTATATTGAGCG CATTCTTTTTCTGGAAGGTCTTCCAAACTTACAGGACCTGGGCAAACT GAACATTGGTGAAGATGTTGAGGAAATGCTGCGTTCTGATCTGGCACT TGAGCTGGATGGCGCGAAGAATTTGCGTGAGGCAATTGGTTATGCCGA TAGCGTTCATGATTACGTCAGCCGCGATATGATGATAGAAATTTTGCG TGATGAAGAAGGCCATATCGACTGGCTGGAAACGGAACTTGATCTGAT TCAGAAGATGGGCCTGCAAAATTATCTGCAAGCACAGATCCGCGAAG AAGGTTGAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCA CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCC CGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCT
HisBfr	NNatgNGNcGtaNcNATTCCCCTCTAgAnaTAATTTTGTTtaACTTTAAGAA GGAGATATACATATGAAAGGTGATACTAAAGTTATAAATTATCTCAAC AAACTGTTGGGAAATGAGCTTGTCGCAATCAATCAGTACTTTCTCCAT GCCCGAATGTTTAAAAACTGGGGTCTCAAACGTCTCAATGATGTGGAG TATCATGAATCCATTGATGAGATGAAACACGCCGATCGTTATATTGAG CGCATTCTTTTTCTGGAAGGTCTTCCAAACTTACAGGACCTGGGCAAA CTGAACATTGGTGAAGATGTTGAGGAAATGCTGCGTTCTGATCTGGCA CTTGAGCTGGATGGCGCGAAGAATTTGCGTGAGGCAATTGGTTATGCC GATAGCGTTCATGATTACGTCAGCCGCGATATGATGATAGAAATTTTG CGTGATGAAGAAGGCCATATCGACTGGCTGGAAACGGAACTTGATCTG ATTCAGAAGATGGGCCTGCAAAATTATCTGCAAGCACAGATCCGCGAA GAAGGTCTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAAC AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATA ACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT GCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCC
Qtag1-Bfr	ATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAAAGAAACCG CTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCAGATCTACATA TGAAAGGTGATACTAAAGTTATAAATTATCTCAACAAACTGTTGGGAA ATGAGCTTGTCGCAATCAATCAGTACTTTCTCCATGCCCGAATGTTTAA AAACTGGGGTCTCAAACGTCTCAATGATGTGGAGTATCATGAATCCAT TGATGAGATGAAACACGCCGATCGTTATATTGAGCGCATTCTTTTTCT GGAAGGTCTTCCAAACTTACAGGACCTGGGCAAACTGAACATTGGTGA AGATGTTGAGGAAATGCTGCGTTCTGATCTGGCACTTGAGCTGGATGG CGCGAAGAATTTGCGTGAGGCAATTGGTTATGCCGATAGCGTTCATGA TTACGTCAGCCGCGATATGATGATAGAAATTTTGCGTGATGAAGAAGG CCATATCGACTGGCTGGAAACGGAACTTGATCTGATTCAGAAGATGGG CCTGCAAAATTATCTGCAAGCACAGATCCGCGAAGAAGGTCTCGAGCA CCACCACCACCACTGAGATCCGGCTG
Qtag2-Bfr	CATATGAAGGAGACCGCGGCGGCGAAATTCGAACGTCAGCACATGGA CAGCGGTGGCGGTGGCATGAAGGGTGATACCAAAGTGATCAACTACC TGAACAAACTGCTGGGCAACGAGCTGGTTGCGATTAACCAATATTTCC TGCACGCGCGTATGTTTAAGAACTGGGGTCTGAAACGTCTGAACGACG TGGAGTACCACGAAAGCATCGACGAGATGAAGCACGCGGATCGTTAT

234

	ATCGAGCGTATTCTGTTTCTGGAAGGCCTGCCGAACCTGCAGGACCTG GGTAAACTGAACATCGGCGAAGATGTTGAGGAAATGCTGCGTAGCGA CCTGGCGCTGGAGCTGGATGGTGCGAAGAACCTGCGTGAAGCGATTG GCTACGCGGACAGCGTGCACGATTATGTTAGCCGTGACATGATGATCG AAATTCTGCGTGACGAGGAAGGTCACATCGATTGGCTGGAGACCGAA CTGGATCTGATTCAGAAGATGGGTCTGCAAAACTACCTGCAGGCGCAA ATTCGTGAGGAAGGCCTCGAG
Qtag3-Bfr	CATATGGAGCGTCTGCAGCAACCGACCGGTGGCGGTATGAAGGGCGA CACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGT TGCGATTAACCAGTATTTCCTGCACGCGCGTATGTTTAAGAACTGGGG CCTGAAACGTCTGAACGATGTGGAGTACCACGAAAGCATCGACGAGA TGAAGCACGCGGATCGTTATATCGAGCGTATTCTGTTCCTGGAAGGTC TGCCGAACCTGCAAGACCTGGGCAAACTGAACATCGGTGAAGATGTTG AGGAAATGCTGCGTAGCGACCTGGCGCTGGAGCTGGATGGTGCGAAG AACCTGCGTGAAGCGATTGGTTACGCGGACAGCGTGCACGATTATGTT AGCCGTGACATGATGATCGAGATTCTGCGTGACGAGGAAGGCCACATC GATTGGCTGGAGACCGAACTGGATCTGATTCAGAAGATGGGCCTGCAA AACTACCTGCAGGCGCAAATTCGTGAGGAAGGTCTCGAG
Qtag4-Bfr	CATATGGAGCGTCTGCAACAACCGACCGGCGGCGGCGGCATGAAGGG CGATACCAAAGTGATTAACTACCTGAACAAACTGCTGGGCAATGAGCT GGTTGCGATTAACCAGTACTTCCTGCACGCGCGTATGTTTAAGAACTG GGGTCTGAAACGTCTGAACGACGTTGAGTACCACGAAAGCATCGACG AGATGAAGCACGCGGATCGTTATATCGAGCGTATTCTGTTCCTGGAAG GCCTGCCGAACCTGCAAGACCTGGGTAAACTGAACATCGGCGAAGAT GTGGAGGAAATGCTGCGTAGCGACCTGGCGCTGGAGCTGGATGGTGC GAAGAACCTGCGTGAAGCGATTGGCTACGCGGACAGCGTGCACGATT ATGTTAGCCGTGATATGATGATCGAAATTCTGCGTGACGAGGAAGGTC ACATCGATTGGCTGGAGACCGAACTGGACCTGATCCAAAAGATGGGC CTGCAAAACTACCTGCAAGCGCAAATCCGTGAAGAGGGCCTCGAG
SortTag1-Bfr	CATATGCTGGTTCCGCGTGGTGGCGGTGGCGGTATGAAGGGCGACACC AAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAGCTGGTTGCG ATTAACCAGTATTTCCTGCACGCGCGTATGTTTAAGAACTGGGGCCTG AAACGTCTGAACGATGTGGAGTACCACGAAAGCATCGACGAGATGAA GCACGCGGATCGTTATATCGAGCGTATTCTGTTCCTGGAAGGTCTGCC GAACCTGCAAGACCTGGGCAAACTGAACATCGGTGAAGATGTTGAGG AAATGCTGCGTAGCGACCTGGCGCTGGAGCTGGATGGTGCGAAGAAC CTGCGTGAAGCGATTGGTTACGCGGACAGCGTGCACGATTATGTTAGC CGTGACATGATGATCGAAATTCTGCGTGACGAGGAAGGCCACATCGAT TGGCTGGAGACCGAACTGGATCTGATTCAGAAGATGGGCCTGCAAAA CTACCTGCAGGCGCAAATTCGTGAGGAAGGTCTCGAG
WTAfFtn	CATATGGCGAGCATTAGCGAGAAAATGGTGGAGGCGCTGAATCGTCA AATCAACGCGGAAATCTATAGCGCGTATCTGTATCTGAGCATGGCGAG CTACTTCGACAGCATCGGTCTGAAGGGCTTTAGCAACTGGATGCGTGT GCAGTGGCAAGAGGAACTGATGCACGCGATGAAGATGTTCGACTTTGT TAGCGAACGTGGTGGCCGTGTGAAACTGTACGCGGTTGAGGAACCGCC GAGCGAATGGGATAGCCCGCTGGCGGCGTTTGAGCACGTGTATGAAC ACGAGGTGAACGTTACCAAGCGTATTCACGAACTGGTTGAGATGGCGA TGCAGGAAAAAGATTTCGCGACCTACAACTTTCTGCAGTGGTACGTTG CGGAGCAAGTTGA

235

|	
	GGAAGAGGCGAGCGCGCTGGACATCGTGGAAAAGCTGCGTCTGATTG GCGAGGATAAACGTGCGCTGCTGTTTCTGGACAAGGAACTGAGCCTGC GTCAATTTACCCCGCCGGCGGAAGAGGAAAAATAACTCGAG
HisAfFtn	CATATGGCGAGCATTAGCGAGAAGATGGTGGAGGCGCTGAACCGTCA GATCAACGCGGAGATTTATAGCGCGTACCTGTACCTGAGCATGGCGAG CTACTTCGACAGCATCGGTCTGAAGGGCTTTAGCAACTGGATGCGTGT GCAGTGGCAAGAGGAACTGATGCACGCGATGAAGATGTTCGACTTTGT TAGCGAACGTGGTGGCCGTGTGAAACTGTACGCGGTTGAGGAACCGCC GAGCGAATGGGATAGCCCGCTGGCGGCGTTTGAGCACGTGTATGAAC ACGAGGTGAACGTTACCAAGCGTATTCACGAACTGGTTGAGATGGCGA TGCAGGAAAAGATTTCGCGACCTACAACTTTCTGCAGTGGTACGTTG CGGAGCAAGTTGAGGAAGAGGCGAGCGCGCTGGACATCGTGGAAAAG CTGCGTCTGATTGGCGAGGATAAACGTGCGCTGCTGTTTCTGGACAAG GAACTGAGCCTGCGTCAATTTACCCCGCCGGCGGAAGAGGAGAAGCT CGAG
AfFtnC54	CATATGGCGAGCATTAGCGAGAAGATGGTGGAGGCGCTGAACCGTCA GATCAACGCGGAGATTTATAGCGCGTACCTGTACCTGAGCATGGCGAG CTACTTCGACAGCATCGGTCTGAAGGGCTTTAGCAACTGGATGCGTGT GCAGTGGCAAGAGGAACTGTGCCACGCGATGAAGATGTTCGACTTTGT TAGCGAACGTGGTGGCCGTGTGAAACTGTACGCGGTTGAGGAACCGCC GAGCGAATGGGATAGCCCGCTGGCGGCGTTTGAGCACGTGTATGAAC ACGAGGTGAACGTTACCAAGCGTATTCACGAACTGGTTGAGATGGCGA TGCAGGAAAAGATTTCGCGACCTACAACTTTCTGCAGTGGTACGTTG CGGAGCAAGTTGAGGAAGAGGCGAGCGCGCTGGACATCGTGGAAAAG CTGCGTCTGATTGGCGAGGATAAACGTGCGCTGCTGTTTCTGGACAAG GAACTGAGCCTGCGTCAATTTACCCCGCCGGCGGAAGAGGAGAAGCT CGAG
HisGFP	NNNNNNNNNNNNNNNNTTCCCTCTNNAATAATTTTGTTTNACTTTAA GAAGGAGNTATACATATGGCTAGCCACCATCACCATCACCATATGGTG AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG GCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCA CCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCA CCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGC ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA CCATCTTCTTCAAGGACGACGGCAACTACAAgACCCGCGCCGAGGTGA AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACG GCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGC GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC CCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTG AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT CGTGACCGCCGCCGGGATCACTCACGGCATGGACGAGCTGACAAGTA ATGCAATGATTCGTCAGACCCTCTGGTCGTGGCTGCCTCCATTATTGGT ATTCTGCACCTGATCCTGTGGAT TCTGGATCGTCTGTAACTCGAGCACCACCACCACCACTGANNTCC GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCNCCG CTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCNNGAG GGGTTTTTTGCTGAAGG

HTEKBfr	ATGAAGGGCGACACCAAGGTTATTAACTACCTGAACAAACTGCTGGGC AACGAACTGGTTGCGATCAATCAATACTTTCTGCACGCGCGTATGTTC AAGAACTGGGGTCTGAAACGTCTGAACGACGTGGAGTACCACGAAAG CATTGACGAGATGAAGCACGCGGATCGTTATATCGAGCGTATTCTGTT TCTGGAAGGCCTGCCGAACCTGCAGGACCTGGGTAAACTGAACATCGG CGAAGATGTTGAGGAAATGCTGCGTAGCGACCTGGCGCTGGAGCTGG ATGGTGCGAAGAACCTGCGTGAAGCGATTGGCTACGCGGACAGCGTG CACGATTATGTTAGCCGTGACATGATGATCGAGATTCTGCGTGACGAG GAAGGTCACATCGATTGGCTGGAAACCGAACTGGATCTGATTCAAAAG ATGGGTCTGCAAAATTACCTGCAAGCGCAAATCCGTGAGGAAGGCTA A
HisPfFtn	ATGCTGAGCGAACGTATGCTGAAGGCGCTGAACGACCAACTGAACCG TGAACTGTATAGCGCGTACCTGTACTTTGCGATGGCGGCGTATTTCGA AGACCTGGGTCTGGAGGGCTTTGCGAACTGGATGAAGGCGCAGGCGG AGGAAGAGATCGGTCACGCGCTGCGTTTCTACAACTATATTTACGACC GTAACGGCCGTGTGGAACTGGATGAGATTCCGAAGCCGCCGAAAGAA TGGGAGAGCCCGCTGAAGGCGTTCGAAGCGGCGTATGAACACGAGAA GTTTATCAGCAAAAGCATTTACGAGCTGGCGGCGCTGGCGGAAGAGG AAAAAGATTACAGCACCCGTGCGTTCCTGGAATGGTTTATCAACGAGC AGGTGGAGGAAGAGGC GAGCGTTAAGAAAATTCTGGACAAGCTGAAATTCGCGAAAGATAGCC CGCAAATCCTGTTTATGCTGGATAAGGAACTGAGCGCGCGTGCGCCGA AACTGCCGGGTCTGCTGATGCAAGGCGGCGAA

**Appendix 8. The DNA sequence for all the constructs used in this work.**

Sample 1	Monomer MW (Da)	Expected 24-mer MW (Da)	Vbar (mL/g)	ρ (g/mL)	η (cP)
100% HisBfr + heme	19,560	469,440	0.735	1.0083	1.0474
Sample 2					
50% HisBfr, 50% wt Bfr + heme	19,560 (His) 18,495 (wt)	443,880-469,440	0.736	1.0083	1.0474
Sample 3					
100% HisBfr + heme + AuNP + Ni2+	19,560	1,257,440	0.735	1.0083	1.0474
Sample 4					
100% His Bfr + heme + C16-NTA	19,560 ~40,000 (C16)	509,440	0.735	1.0083	1.0474
Sample 5					
50% HisBfr, 50% wt Bfr + C19-NTA	19,560 (His) 18,495 (wt) ~43,360 (C19)	487,240-512,800	0.736	1.0083	1.0474
Sample 6					
100% AfFtnC-NTA	20,288	486,912	0.735	1.0019	1.0300

Appendix 9. Initial parameters for AUC. The estimated with molecular weights calculated from amino acid sequences Partial specific volumes (**vbar**) were also calculated from their amino acid compositions, using the program SEDNTERP. SEDNTERP was also used to calculate buffer densities(ρ), and viscosities(η).
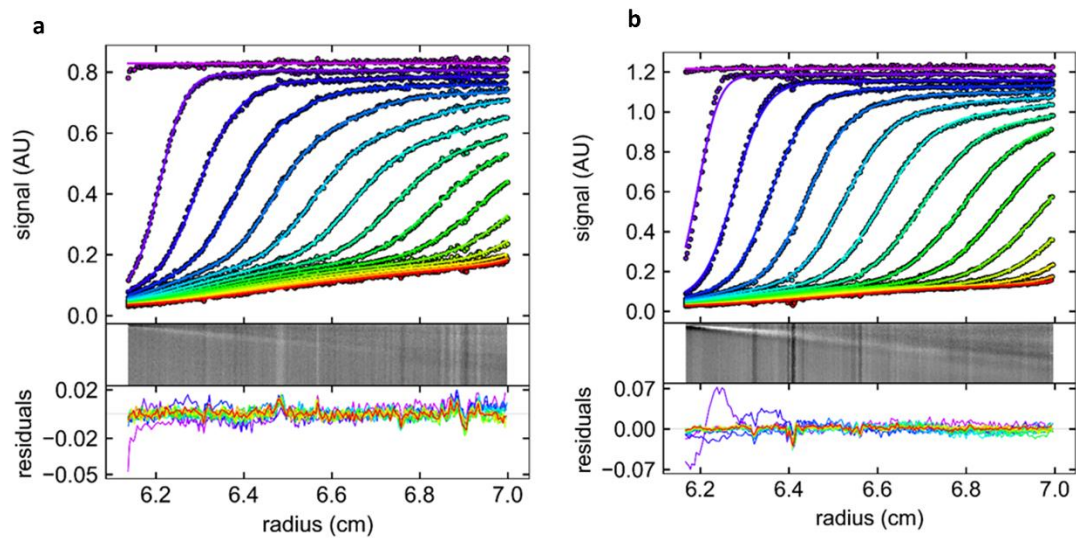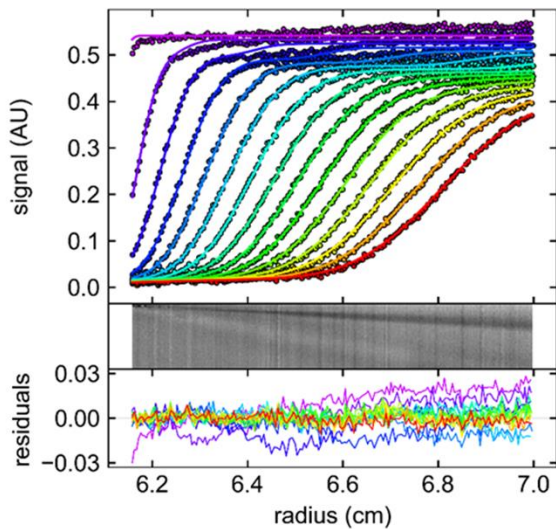
Appendix 10. Sedimentation velocity(top) and residuals curve (bottom). (a) 100 % HisBfr + heme (b) 50 % HisBfr + 50 % WTBfr + heme. All measurement were recorded at 20 000 rpm in a 1.2cm pathlength centerpiece cuvette at 280 nm wavelength.The sedimentation curves were deconvoluted to produce the size distribution plots.
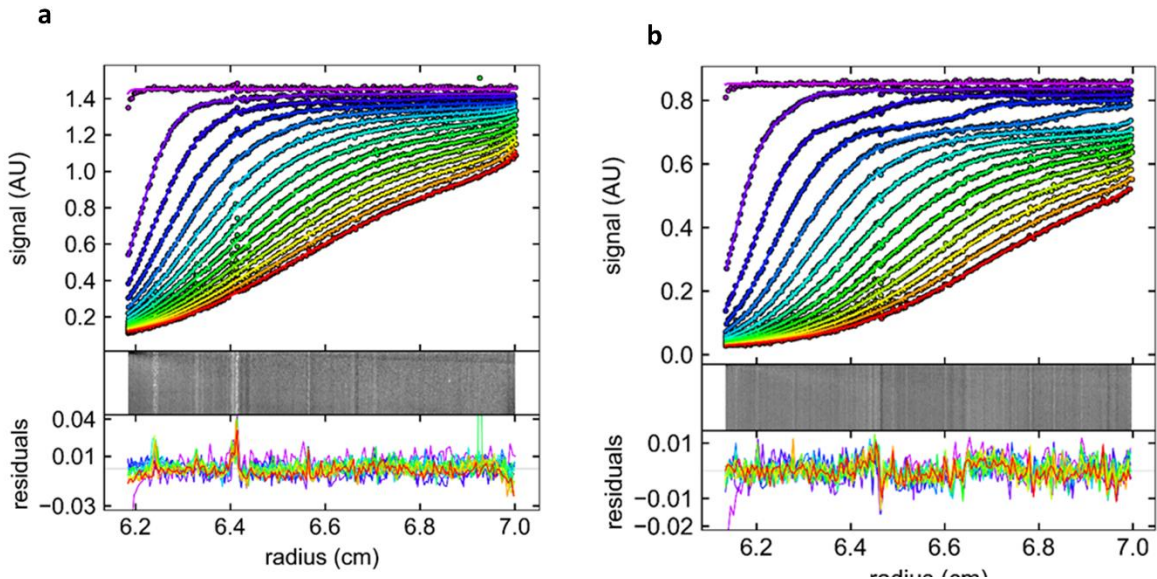
Appendix 11. **Sedimentation velocity(top) and residuals curve (bottom).** (a) 100 % HisBfr + heme (b) 50 % HisBfr + 50 % WTBfr + heme. Sample a and b measurement were recorded at 20 000 rpm and 30 000 rpm, respectively in a 1.2cm pathlength centerpiece cuvette at 280 nm wavelength.
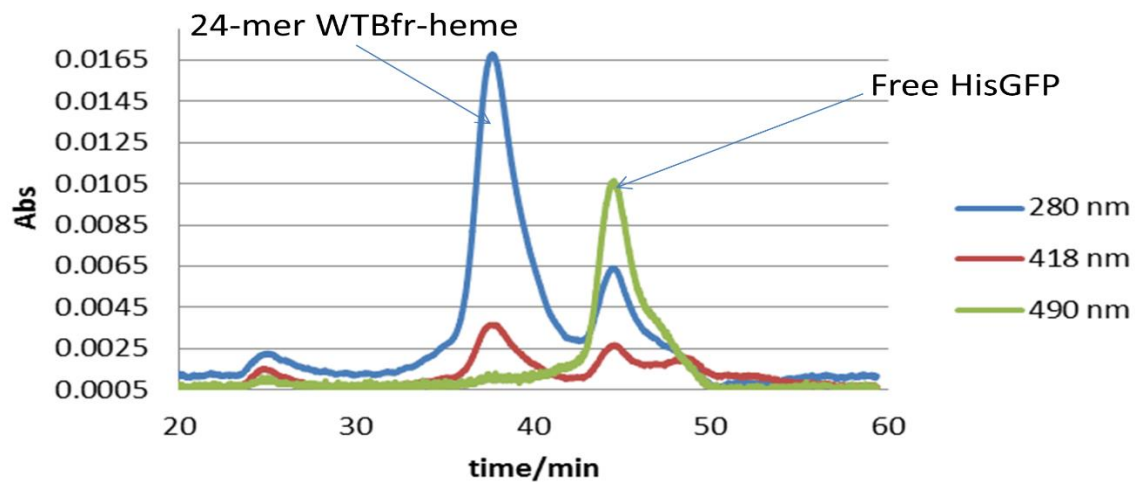

Appendix 12. Sedimentation velocity(top) and residuals curve (bottom) for AfFtnC54-NTA. AfFtnC54-NTA was dialyzed in 20 mM MgCl2, 25 mM HEPES-pH 7.5. The sample was recorded at 20 000 rpm in a 1.2cm pathlength centerpiece cuvette at 280 nm wavelength.

240

Sample 1 100% HisBfr + heme	Peak 1	Peak 2	Peak 3	Peak 4
MW (kDa)	68.3	208.1	498.8	1025.9
$s_{20,w}$ (S)	5.0	10.6	18.9	30.6
f/f$_o$	1.17			
RMSD	0.0048			

Sample 2 50% HisBfr, 50% wt Bfr + heme	Peak 1	Peak 2	Peak 3	
MW (kDa)	102.6	472.3	914.5	
$s_{20,w}$ (S)	6.4	17.7	27.6	
f/f$_o$	1.20			
RMSD	0.0078			

Sample 3 100% HisBfr + heme + AuNP + Ni2+	Peak 1	Peak 2	Peak 3	Peak 4
MW (kDa)	138.4	546.4	1686.1*	3906.8**
$s_{20,w}$ (S)	7.6	18.9	40.1*	70.3**
f/f$_o$	1.25			
RMSD	0.0086			
	*this broad 'peak' encompasses values of s = 25-50S, so the MW is an average of the population **this broad peak spanned from s = 55-80S, so the MW is an average			

Sample 4 100% His Bfr + heme + C16-NTA	Peak 1	Peak 2	Peak 3	Peak 4
MW (kDa)	95.0	253.0	495.0	994.1
$s_{20,w}$ (S)	4.4	8.5	13.3	21.1
f/f_o	1.67			
RMSD	0.0070			
Sample 5 50% HisBfr, 50% wt Bfr + C19-NTA	Peak 1	Peak 2		
MW (kDa)	83.9*	589.5		
$s_{20,w}$ (S)	4.7*	17.1		
f/f_o	1.44			
RMSD	0.0037			
	*I am unable to resolve this peak into multiple species, despite its broad, assymetric nature			
Sample 6 100% AfFtnC-NTA	Peak 1	Peak 2		
MW (kDa)	471.1	932.9		
$s_{20,w}$ (S)	17.7	27.9		
f/f_o	1.21			
RMSD	0.0065			

Appendix 13. Summary of results from all the AUC samples.

Appendix 14. Encapsulation of HisGFP into WTBfr using unmodified heme cofactor. The SEC column (Superose 6 ™ 10/300 GL) run shows empty WTBfr-heme capsules at ~ 38 min and free HisGFP at ~ 45 min