

Protein-based adsorbents for clearance of protein-bound uremic toxins

Dissertation

zur Erlangung der Würde des Doktors der Naturwissenschaften der
Fakultät für Mathematik, Informatik und Naturwissenschaften,
Fachbereich Chemie der Universität Hamburg

vorgelegt von

Hendrik Böhler

aus Gelsenkirchen

Hamburg 2022

Gutachter der Dissertation

Prof. Dr. Tobias Beck

Prof. Dr. Wolfgang Parak

Gutachter der Disputation

Prof. Dr. Tobias Beck

Prof. Dr. Andrew Torda

Dr. Maria Riedner

Tag der Disputation

11.11.2022

Druckfreigabe

14.11.2022

The work presented in this thesis was carried out at the Institute of Inorganic Chemistry of the RWTH Aachen University between January 2019 and December 2019 and at the Institute of Physical Chemistry of the University Hamburg between January 2020 and August 2022.

Publications:

Parts of this thesis are in the process of being published:

- I. Böhler, H.; Orth-Alampour, S.; Baaten, C.; Riedner, M; Jankowski, J.; Beck, T., Assembly of chemically modified protein nanocages into 3D materials for the adsorption of uremic toxins, *submitted*.
- II. Böhler, H.; Orth-Alampour, S.; Baaten, C.; Riedner, M; Jankowski, J.; Beck, T., Redesigned ferritin variants and their assembly into 3D materials for the adsorption of uremic toxins , *in preparation*.
- III. Patent filing: Neue Toxin-Bindemittel für Hämodialyse, *in preparation*

The author has also contributed to the following related publications:

- IV. Böhler, H.; Rütten, M.; Lang, L.; Beck, T., Crystalline biohybrid materials based on protein cages in *Methods in Molecular Biology*, Springer Nature, *in press*.

Abstract

Elevated concentrations of protein-bound uremic toxins (PBUT) are linked to high cardiovascular mortality in patients suffering from chronic kidney disease.^[1] Due to their partly hydrophobic character, PBUTs are bound to plasma proteins, which limits their removal in renal replacement therapy. Extracorporeal adsorbents are a promising addition to conventional treatment. However, the combination of high adsorption capacity and hemocompatibility is challenging. Here, an unprecedented approach for fabrication of a heterogeneous adsorbent based on the intrinsically biocompatible protein cage ferritin was investigated. The general feasibility of this strategy was demonstrated by creation of protein-based crystalline or non-crystalline materials, which were stabilized by chemical fixation to tolerate conditions in blood-like systems. First experiments in terms of biocompatibility were promising by verifying the absence of endotoxin contaminations and showing no activation of blood platelets. To enhance PBUT adsorption, strategies to modify the inner surface of the protein cage were developed. First, conjugation of hydrophobic molecules to genetically introduced cysteine amino acids was investigated. Protein variants modified with up to 96 phenyl or aliphatic guest molecules could be fabricated. The second strategy embodied extensive protein redesign guided by the Rosetta molecular modeling software suite. Ferritin variants with decreased negative charge on the inner surface, but with increased density of hydrophobic amino acids, expanded pores, and specific binding sites were designed *in silico*. Respective variants were expressed, purified, and characterized *in vitro*. The adsorption capacity of the developed protein-based adsorbents to selected PBUTs, namely indoxyl sulfate (IS), *p*-cresyl sulfate (*p*CS) and phenylacetic acid (PAA), was determined. Unmodified cages already showed adsorption capacity towards the PBUTs. Neither introduction of hydrophobic molecules nor hydrophobic amino acids revealed a significant increase in adsorption capacity. However, binding sites combining hydrophobic and polar positive amino acids increased the adsorption capacity significantly. Further, binding sites targeting unique feature of one toxin showed evidence for selectivity in their adsorption behavior. This finding indicates that the hydrophobic and polar negative charged groups of the PBUTs need to be stabilized to achieve effective adsorption. Mutations decreasing negative surface potential at the entry area or expanding the pores also increased the adsorption capacity indicating the importance of mass transport inside the cage. Adsorption capacities found for protein-based materials (IS: $985 \pm 11 \mu\text{g g}^{-1}$, *p*CS: $697 \pm 64 \mu\text{g g}^{-1}$, PAA: $13,440 \pm 840 \mu\text{g g}^{-1}$) were already in the same order of magnitude as some conventional adsorbents, while showing excellent biocompatibility. The conducted research identifies several possible routes for further optimization and highlight the great potential of protein-based adsorbents in renal replacement therapy.

Zusammenfassung

Erhöhte Konzentrationen proteingebundener urämischer Toxine (PBUT) stehen in Verbindung mit einer hohen kardiovaskulären Sterblichkeit bei Patienten mit chronischer Nierenerkrankung.^[1] Aufgrund ihres teilweise hydrophoben Charakters sind PBUTs an Plasmaproteine gebunden, was ihre Entfernung bei der Nierenersatztherapie einschränkt. Extrakorporale Adsorptionsmittel sind eine vielversprechende Ergänzung zur konventionellen Behandlung. Die Kombination aus hoher Adsorptionskapazität und Hämokompatibilität stellt jedoch eine Herausforderung dar. In diesem Projekt wurde ein noch nie dagewesener Ansatz zur Herstellung eines heterogenen Adsorptionsmittels auf der Grundlage des biokompatiblen Proteinkäfigs Ferritin untersucht. Die generelle Machbarkeit dieser Strategie wurde durch die Herstellung von kristallinen oder nicht-kristallinen Materialien auf Proteinbasis demonstriert, die durch chemische Fixierung stabilisiert wurden, um die Bedingungen in blutähnlichen Systemen zu tolerieren. Erste Experimente zur Biokompatibilität waren vielversprechend, da keine Endotoxin-Kontaminationen nachgewiesen werden konnten und keine Aktivierung von Blutplättchen auftrat. Um die PBUT-Adsorption zu verbessern, wurden Strategien zur Modifizierung der inneren Oberfläche des Proteinkäfigs entwickelt. Zunächst wurde die Konjugation hydrophober Moleküle an genetisch eingeführten Cystein-Aminosäuren untersucht. Es konnten Proteinvarianten hergestellt werden, die mit bis zu 96 phenylischen oder aliphatischen Gastmolekülen modifiziert waren. Die zweite Strategie bestand in einer umfassenden Umgestaltung der Proteine mit Hilfe der Rosetta-Software für die Molekularmodellierung. Ferritin-Varianten mit geringerer negativer Ladung auf der inneren Oberfläche, aber mit erhöhter Dichte an hydrophoben Aminosäuren, erweiterten Poren und spezifischen Bindungsstellen wurden *in silico* entworfen. Die entsprechenden Varianten wurden *in vitro* exprimiert, aufgereinigt und charakterisiert. Die Adsorptionskapazität der entwickelten proteinbasierten Adsorbentien für ausgewählte PBUTs, nämlich Indoxylsulfat (IS), p-Kresylsulfat (pCS) und Phenyllessigsäure (PAA), wurde bestimmt. Unveränderte Käfige zeigten bereits eine Adsorptionskapazität gegenüber den PBUTs. Weder die Einführung von hydrophoben Molekülen noch von hydrophoben Aminosäuren führte zu einer signifikanten Erhöhung der Adsorptionskapazität. Bindungsstellen, die hydrophobe und polar-positive Aminosäuren kombinieren, erhöhten jedoch die Adsorptionskapazität erheblich. Darüber hinaus zeigten Bindungsstellen, die ein einzigartiges Merkmal eines Toxins stabilisieren, Hinweise auf Selektivität in ihrem Adsorptionsverhalten. Dieses Ergebnis deutet darauf hin, dass die hydrophoben und polar negativ geladenen Gruppen der PBUTs stabilisiert werden müssen, um eine effektive Adsorption zu erreichen. Mutationen, die das negative Oberflächenpotenzial im Eingangsbereich verringerten oder die Poren vergrößerten, erhöhten ebenfalls die Adsorptionskapazität, was auf die Bedeutung des Massentransports ins Innere

des Käfigs hinweist. Die Adsorptionskapazitäten der proteinbasierten Materialien (IS: $985 \pm 11 \mu\text{g g}^{-1}$, pCS: $697 \pm 64 \mu\text{g g}^{-1}$, PPA: $13.440 \pm 840 \mu\text{g g}^{-1}$) lagen bereits in der selben Größenordnung wie die einiger herkömmlicher Adsorbentien und wiesen gleichzeitig eine ausgezeichnete Biokompatibilität auf. Die durchgeführten Untersuchungen zeigen mehrere mögliche Wege zur weiteren Optimierung auf und unterstreichen das große Potenzial von proteinbasierten Adsorbentien in der Nierenersatztherapie.

Table of contents

1 Introduction	1
2 Scientific background	2
2.1 Chronic kidney diseases	2
2.2 Protein-bound uremic toxins	3
2.3 Removal of protein-bound uremic toxins	4
2.4 Protein nano-compartments	7
2.4.1 The ferritin protein container	11
2.4.2 Bioconjugation	13
2.5 The Rosetta molecular modeling software suite	18
2.5.1 Basic concepts of Rosetta – the core library	18
2.5.2 Applications in this research	25
3 Basis of this work	27
4 Concepts and aim of this thesis	28
4.1 Covalent linkage of hydrophobic molecules to the inner surface	29
4.2 Redesign of the cavity by amino acid exchange	30
5 Results and discussion	31
5.1 Functionalization of the inner protein cavity	31
5.1.1 Rational design of Ftn ^(neg) -Cys variants	31
5.1.2 Mutagenesis, production and purification of ferritin cysteine variants	32
5.1.3 Functionalization with hydrophobic agents	34
5.1.4 Fabrication of heterogeneous adsorbent material	42
5.1.5 Biocompatibility	50
5.1.6 The adsorption capacity of functionalized Ftn ^(neg) variants	54
5.1.7 Encapsulation of fluorophore molecules for bioimaging and biosensing applications	62
5.2 Designing toxin binding Ftn ^(neg) variants	65
5.2.1 Computational design	65

5.2.2 Expression, purification and characterization of redesigned ferritin variants	80
5.2.3 Fabrication of heterogeneous adsorbent materials	90
5.2.4 Biocompatibility	92
5.2.5 Adsorption capacity of redesigned Ftn ^(neg) variants	94
5.3 Protein-Protein docking for interfaces in binary protein crystals	104
6 Summary and conclusion	109
7 Outlook.....	114
8 Experimental part	117
8.1 General.....	117
8.2 Chemicals	117
8.3 Analytic methods.....	117
8.3.1 TEM.....	117
8.3.2 ESI mass spectrometry.....	118
8.3.3 UV/Vis spectroscopy	118
8.3.4 Bradford assay	118
8.3.5 Toxin assay	119
8.4 Protein production and purification.....	121
8.4.1 QuikChange site-directed mutagenesis	121
8.4.2 Protein production and purification	122
8.4.3 Purification from exclusion bodies.	123
8.5 Functionalization of cysteines in the inner cavity.....	123
8.5.1 Functionalization with 2-iodo-N-phenylacetamide	123
8.5.2 Functionalization with 2-Bromo-N-decylacetamide	124
8.5.3 Functionalization with fluorophores.....	124
8.6 Crystallography	125
8.6.1 Hanging drop crystallization.....	125
8.6.2 Structure determination and refinement.....	125
8.6.3 Batch crystallization.....	126

8.6.4 Fabrication of non-crystalline adsorbent	126
8.6.5 Crystal fixation.....	126
8.7 Biocompatibility assays.....	127
8.7.1 Quantitative polymerase chain reaction (qPCR) analysis of mRNA expression in human aortic endothelial cells	127
8.7.2 Platelet activation assay	128
8.7.3 BSA adsorption assay	128
8.8 Computational methods	129
8.8.1 Preparation.....	129
8.8.2 Design with fixbb protocols	132
8.8.3 Design of binding sites	133
8.8.4 Protein-protein docking.....	137
9 Bibliography	140
10 Appendix	150
10.1 Amino acid sequences of proteins fabricated in the thesis	150
10.2 Additional information for chemical modification part.....	153
10.3 Additional information on protein redesign part	169
10.4 Results from protein-protein docking with binary protein container.....	189
10.5 Scripts and files used for computational protein design.....	197
10.5.1 Compilation of Rosetta	197
10.5.2 Generate symmetry definition file	197
10.5.3 Relax protocol	198
10.5.4 Fixbb protocol.....	199
10.5.5 Script to sort the Relax and Fixbb output	203
10.5.6 Protein-ligand docking	204
10.5.7 Protein-protein docking.....	215
10.5.8 Script for identifying the influence on single amino acids and the overall dG_separated	220

11 List of abbreviations.....	223
12 List of chemicals.....	225
13 List of figures.....	228
14 List of tables.....	232
Acknowledgment – Danksagung	233

1 Introduction

Approximately 9% of the world's population is suffering from chronic kidney disease (CKD).^[2] This condition arises from different disease pathways reducing the body's capability to clear the blood from harmful compounds. In end-stage CKD patients, the kidney is no longer capable to clear toxins properly and renal replacement therapies in form of dialysis or kidney replacement are necessary.^[3] The need of renal replacement therapies is expected to grow over the next decade, especially due to demographic and lifestyle shifts in combination with an increased demand in developing countries, for example, in East Asia.^[3-5]

People suffering from CKD have a significantly higher susceptibility to cardiovascular diseases.^[6,7] As shown in recent studies, these indications could be associated with the group of protein-bound uremic toxins (PBUT).^[1,8] Well-studied representatives of this group are indoxyl sulfate (IS) and *p*-cresyl sulfate (*p*CS). They can induce renal and cardiovascular toxicity,^[1] increase the risk for thrombotic events,^[9] inhibit endothelial proliferation and wound repair,^[10] cause endothelial dysfunction,^[11] kidney fibrosis^[12] and cardiovascular calcification.^[13] Due to their partially hydrophobic character, the PBUTs are bound to plasma proteins (e.g. serum albumin), which prevents efficient diffusion through the pores of dialysis membranes.^[14,15] As a result, even prolonged and more frequent therapies with common hemodialysis techniques can not sufficiently clear the PBUTs from the blood.^[16] Blood purification with adsorbents for clearance of PBUTs from the blood is a promising addition to conventional therapy. The adsorbent could be easily integrated in common dialysis techniques, thus reducing treatment costs and simultaneously allowing regular treatment of CKD patients.^[17] However, due to their low hemocompatibility, conventional adsorbents can induce blood coagulation, allergic reactions, and depletion of platelets and leukocytes.^[17]

These findings demonstrate that bio- and hemocompatible adsorbents possessing high affinity to PBUTs need to be developed. Current material research is mainly focused on increasing the hemocompatibility of conventional materials such as activated charcoal^[17,18], zeolites^[19], or metal-organic frameworks.^[20] In this work, an opposite strategy is demonstrated: The protein cage ferritin possess an intrinsic bio- and hemocompatibility due to its origin from the human organism. Therefore, ferritin was selected as a scaffold to investigate the fabrication of heterogeneous adsorbents based on biocompatible proteins. To increase the PBUT adsorption of the protein cage, strategies including chemical modification and genetic redesign were developed and investigated.

2 Scientific background

In this chapter, the theoretical background about the main topics of this work is described. The chapter is separated in three main sections including different subsections. Key points are summarized at the end of each main section. The first part consists of background information on chronic kidney disease and state of the art of PBUTs clearance. This is followed by section two with closer focus on protein cages and their advantages. In addition, strategies for modification of protein cages are presented. In section three, the Rosetta molecular modeling software suite, which was used in this work to design protein variants capable of PBUT adsorption, is described.

2.1 Chronic kidney diseases

The kidney plays an essential role in human physiology and is responsible for clearance of substances from blood. The kidney balances the electrolyte and water budget and is likewise further involved in human metabolism and hormone production.^[21]

Chronic kidney disease (CKD) is attributed to an irreversible loss in kidney function. It is defined as abnormalities with implications for health in kidney structure or function present for longer than 3 months.^[22] In high- to medium-income countries, diabetes and hypertension are the main causes leading to CKD.^[2] The human kidneys has metabolic, endocrine and excretory functions. Excretory function can be measured by the glomerular filtration rate (GFR), which is a common index for overall kidney function. GFR is commonly measured by the clearance of an exogenous marker after a single injection.^[3] However, markers for kidney damage are also used as measurement tool of CKD progression. These markers include i.e., increased concentrations of albuminuria, urinary sedimentations, renal tubular disorders or structural abnormalities. Based on these criteria, kidney disease is divided into 5 different stages.^[22] In early stages, patients are frequently asymptomatic or might have non-specific symptoms such as lethargy, itch or a loss of appetite. In later stages, more severe complications like anemia, bone disease and an increased risk for cardiovascular disease and cancer can emerge.^[3]

CKD patients have a high propensity to die from a cardiovascular event. In 2017 from 2.6 million deaths attributed to CKD, 1.4 million people died due to cardiovascular disease attributed to an impaired kidney function.^[2] Similar high mortalities can be also observed in patients on dialysis.^[7] In recent years, research demonstrated that this overall high risk for cardiovascular events could be referred to protein-bound uremic toxins.^[1,8]

2.2 Protein-bound uremic toxins

Uremic toxins or uremic retention solutes constitute the vast number of components, which are usually excreted by healthy kidneys. Their concentration is correlated to at least one symptom of the uremic syndrome.^[23] Uremic toxins are classified by the European Uremic Toxin Work Group into three categories.^[15,24]

1. Small water-soluble compounds with molecular weight up to 500 Da (e.g. urea, creatinine, uric acid);
2. Middle-sized molecules larger than 500 Da (e.g. β 2-microglobulin, leptin);
3. Protein-bound compounds with a molecular weight mostly below 500 Da (e.g. indoxyl sulfate, *p*-cresyl sulfate, phenylacetic acid)

The concentration of these solutes vary over a broad range from nanograms per liter to grams per liter.^[15] There are more than 100 known uremic toxins ^[25] from which 33 are protein-bound.^[26]

The further part of this section focuses on three PBUTs, which are used as model systems in this work; namely indoxyl sulfate (IS), *p*-cresyl sulfate (*p*CS) and phenylacetic acid (PAA). The chemical structures of these toxins are shown in Figure 2.1.

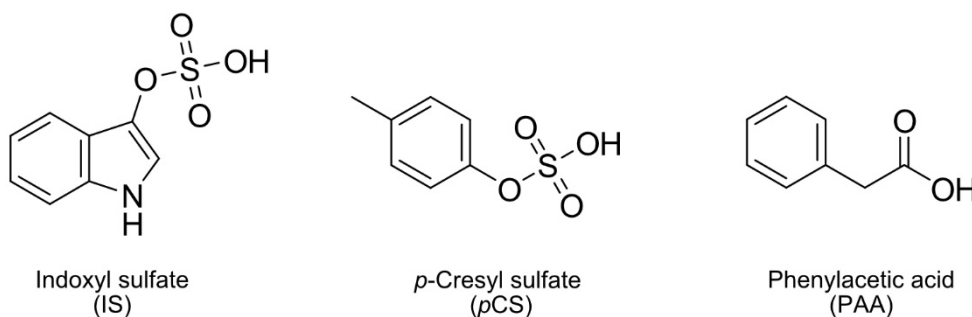


Figure 2.1: Structure of the protein-bound uremic toxins. The three shown PBUTs are used as model systems in this work. All of them have an acidic group, which is deprotonated and thus negatively charged at neutral pH values. Furthermore, each PBUT possess an aromatic ring system responsible for their partly hydrophobic character.

The main fraction of the investigated PBUTs originates from excess dietary proteins, specifically from excess aromatic amino acids tyrosine, phenylalanine, and tryptophan.^[27-29] The degradation of all three amino acids is started by gut bacteria in the colon. Phenylalanine is partly transformed to tyrosine or metabolized to phenylacetic acid.^[27,28] Gut bacteria transforms tyrosine over several intermediates to *p*-cresol, which is then metabolized in the liver to *p*-cresyl sulfate.^[28] Tryptophan is firstly converted to indole by gut bacteria and finally

metabolized to indoxyl sulfate in the liver.^[29] After entering the bloodstream, all three derivatives bind reversibly to plasma proteins, in particular human serum albumin (HSA). *p*CS and IS bind to drug binding sites in HSA, namely to the Sudlow site I and II.^[30] PAA also binds to HSA but only show weak affinity to the Sudlow sites, indicating the presence of more favorable binding site for PAA.^[31] The bound fraction depends on the derivative. Around 95% of *p*CS and IS^[32] are bound, while only 70% of the PAA are bound.^[33] The toxins are transported by serum proteins to the kidneys. Unbound fractions are directly secreted from the peritubular capillary into proximal tubule cells in healthy organisms. Bound fractions are uncoupled by organic anion transporters (OATs) at the basolateral membrane of the proximal tubule cell. Subsequently, toxins are transported by membrane-based proteins and are finally excreted in the urine.^[29]

Elevated PBUT concentrations are thought to have varieties of effects on the human organisms. IS can induce vascular inflammation,^[8] cardiovascular calcification,^[13] accelerate thrombotic responses after vascular injury^[9] and induces oxidative stress which can lead to endothelial dysfunction.^[11] The PBUT *p*CS could be linked to an increased risk of cardiovascular toxicity and mortality. IS and *p*CS inhibit endothelial proliferation and wound repair,^[12] additionally they induce renal and kidney fibrosis, causing a faster progression of CKD.^[1] PBUTs can also affect the immune defense like PAA by impairing leucocyte^[34] or macrophage^[35] function. Similar to the other PBUTs, PAA could also attributed to higher cardiovascular morbidity.^[33]

2.3 Removal of protein-bound uremic toxins

Renal replacement therapies (RRT) are frequently used as treatment for CKD patients. There is a number of different techniques in various variations. However, detailed descriptions of all methods cannot be included in the scope of this thesis. In-depth information can be found in reference 36 and 37. Most RRT are applied extracorporeal. Nonetheless, paracorporeal methods (i.e., peritoneal dialysis) exist. Commonly applied extracorporeal techniques are hemodialysis, hemofiltration and hemodiafiltration. In hemodialysis, the driving force for solute movement is based on a concentration gradient. This diffusive transport effectively removes the small-water soluble solutes from the blood. In hemofiltration, a convective clearance along a pressure gradient is embodied. In contrast to hemodialysis, hemofiltration has shown improved clearance for middle-sized water-soluble molecules.^[36] A combination of diffusive and convective clearance is commonly used in hemodiafiltration^[37,38] All dialysis methods are based on the movement of solutes through pores of a dialytic membrane. Since the main fraction of PBUTs are bound to plasma proteins, their passing through the membrane pores is

hindered.^[14,15] This results in inefficient removal of these toxins with common renal replacement therapies.^[16,39] An overview of these techniques is summarized in Table 2.1.

Table 2.1: Renal replacement therapies. Overview of conventional renal replacement therapies with their mode of mass transfer and the class of solutes, which can effectively be cleared from the blood.

Renal replacement therapy	Mass transport	Effective clearance of		
		small-sized water-soluble solutes	middel-sized water-soluble solutes	PBUTs
Hemodialysis	Diffusive	✓	X	X
Hemofiltration	Convective	X	✓	X
Hemodiafiltration	Diffusive and convective	✓	✓	X

Separation of PBUTs can be improved through treatment with oral or intravenous supplements. Intravenous supplements are designed to displace toxins from the binding sites of plasma proteins, thus increasing the soluble fraction that can be separated by dialysis. This can be achieved, for example, by ibuprofen,^[40] mesna^[41] or acetylcysteine^[42]. One promising approach of increasing the unbound fraction of PBUTs without addition of further drugs could be treatment with high-frequency fields which is already investigated and patented^[26]. Oral supplements such as the carbon-based material AST-120 (Kremezin™) adsorb toxins produced by microbial processes in the gut before they enter the bloodstream^[43]. Moreover, therapeutics like synbiotic NATUREN G™ can inhibit PBUT production.^[44] In a systematic comparison of all studies published until January 2020 focused on the reduction of protein-bound uremic toxins^[26], the above mentioned methods reached reduction rates for the toxins IS and pCS of up to 44%.^[43] However, significantly higher reduction rates between 71% and 78% were achieved with a combination of fractional plasma separation, adsorption and dialysis (FPAD).^[26,45] This technique is based on separation of HSA and other proteins from blood serum by an initial membrane based separation step followed by contact with a hydrophobic adsorbent.^[46] This fractionated plasma separation is crucial to prevent blood coagulation and other side-effects, which result from contact of blood cells and platelets with the hydrophobic adsorbent. Systems based on the FPAD principle are already commercially available for treatment of liver failure (Prometheus®).^[47] However, demand of additional equipment and the attributed cost make the FPAD technique too expensive for routine use and broad-scale

treatment of dialysis patients.^[17] Nevertheless, this demonstrates the great potential of adsorption based techniques for removal of PBUTs.

Several common adsorbents were investigated for their ability to bind uremic toxins including zeolites,^[19,48] metal-organic frameworks^[20] or carbon-based adsorbents.^[50] Adsorption of urea, uric acid, creatine, *p*-cresol and indoxyl sulfate are demonstrated for different types of zeolites.^[19,48] Reported adsorption capacities for the zeolite type P87 is 1000 $\mu\text{g g}^{-1}$ for indoxyl sulfate.^[49] Carbon-based adsorbents with a dual porosity were also investigated for their potential in clearance of uremic toxins and cytokines. These materials possess adsorption capacities between 3200 and up to 3700 $\mu\text{g g}^{-1}$ as demonstrated in recent publications.^[50] For zirconium-based metal-organic frameworks (MOF), high adsorption capacities of 156 (IS) and 166 (*p*CS) mg g^{-1} were reported.^[20] Single-crystal X-ray diffraction (SCXRD) of such materials revealed that *p*CS prefers binding sites where its aromatic ring is bound by π - π interactions, whereas the negatively charged sulfate group forms hydrogen bonds with polar parts of the material.^[20] These findings indicate that for efficient adsorption of these toxins, a mixture of hydrophilic and hydrophobic interactions is required. Instead of using the adsorbent as a free standing material it is also possible to incorporate them into polymer-based hollow fiber membranes like demonstrated for carbon particles^[51,52] or zeolites.^[49] Extensive bio- and especially hemocompatibility studies for the presented materials are not available yet. Efforts to increase hemocompatibility of adsorbents were done for activated charcoal based materials either for adsorber particle with a polymer coating^[17] or monolithic adsorbents based on a lignin binder route.^[18] Both materials showed no activation of blood platelets and induction of blood coagulation. Nonetheless, further investigations for inflammation, allergic reactions or potential side effects with more sensitive CKD blood need to be performed for clinical translation.^[17,18]

All reported studies focus on improving the compatibility of conventional materials for applications in blood. In this thesis, the issue is approached from reverse direction by choosing the protein cage ferritin with its intrinsic biocompatibility as matrix material for a PBUT adsorbent. The structure of protein cages makes them exceptionally modular and allows tailoring their properties for specific applications. In the following section 2.4, properties and applications for protein cages in general, and ferritin in particular, are discussed.

Summary

- Elevated PBUT concentrations increase the risk for cardiovascular events and accelerate overall progression of CKD.
- Main fraction of PBUTs are bound to plasma proteins hindering their clearance by conventional renal replacement therapies.

- Application of adsorbents display a promising strategy for PBUT removal. Suitable materials with high biocompatibility and affinity for PBUTs are not developed yet

2.4 Protein nano-compartments

Protein cages can be found in almost every organism playing important roles in various biological processes, for instance, protein folding,^[53,54] proteolysis,^[55] catalysis,^[56] or storage.^[57] During these processes, they provide high local concentrations of reaction partners and sequester toxic or volatile intermediates.^[55]

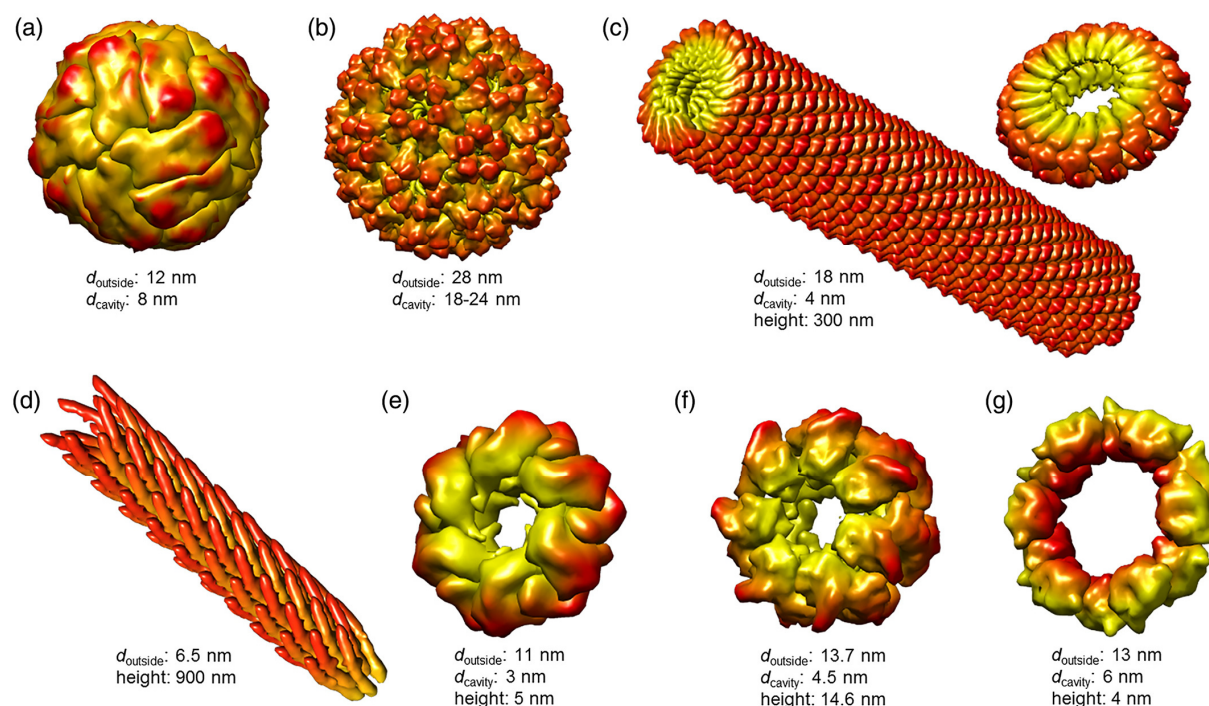


Figure 2.2 : Dimensions and shapes of different protein cages. The structural variety of protein cages is demonstrated by the following examples. a) horse spleen apoferritin (PDB ID: 1IER), b) cowpea chlorotic mottle virus (PDB ID: 1CWP), c) tobacco mosaic virus, rod (PDB-ID: 3J06) and its two-ring circular permutant (PDB ID: 3KML), d) M13 phage (PDB ID: 1IFI); (e) stable protein 1 (PDB ID: 1TR0); f) GroEL (PDB ID: 1SS8), and g) bovine mitochondrial peroxiredoxin (PDB ID: 1ZYE). (Adopted from reference 58 with permission from Wiley & Sons, copyright 2019)

Most protein containers adopt spherical morphologies, but rod-, ring- or tube-like structures are known as well.^[58,59] Despite of varieties in size and shape depicted in Figure 2.2, all protein structures share basic design principles and are assembled by a different number of small-scaled proteins (subunits). Their subunits self-assemble under formation of various intermediates to the full cage structure. Depending on container size, the total number of subunits can vary significantly. This concept allows minimization of genetic burden, while maximizing the space for cargo.^[57] Cage assembly is entropically favored by an increase in

solvent disorder through release of ordered water molecules, achieved by burying of hydrophobic surfaces upon subunit association.^[60] Various assembly mechanisms were observed or modeled:^[61-64] Reported examples are a nucleation and growth pathway or an unordered protein binding on the cargo followed by rearrangement into ordered cage structures.^[65] Due to their inherent design, protein cages possess three distinct interfaces: An outer surface, an inner surface and a surface at the intersection of subunits (inter-subunit surface), visualized in Figure 2.3. Different surfaces govern key properties of the cage independently from each other, which can be selectively altered by chemical or genetical modification.^[66] These features in combination with the atomic precision of protein-biosynthesis make protein compartments to a valuable tool in bio-nanotechnology.

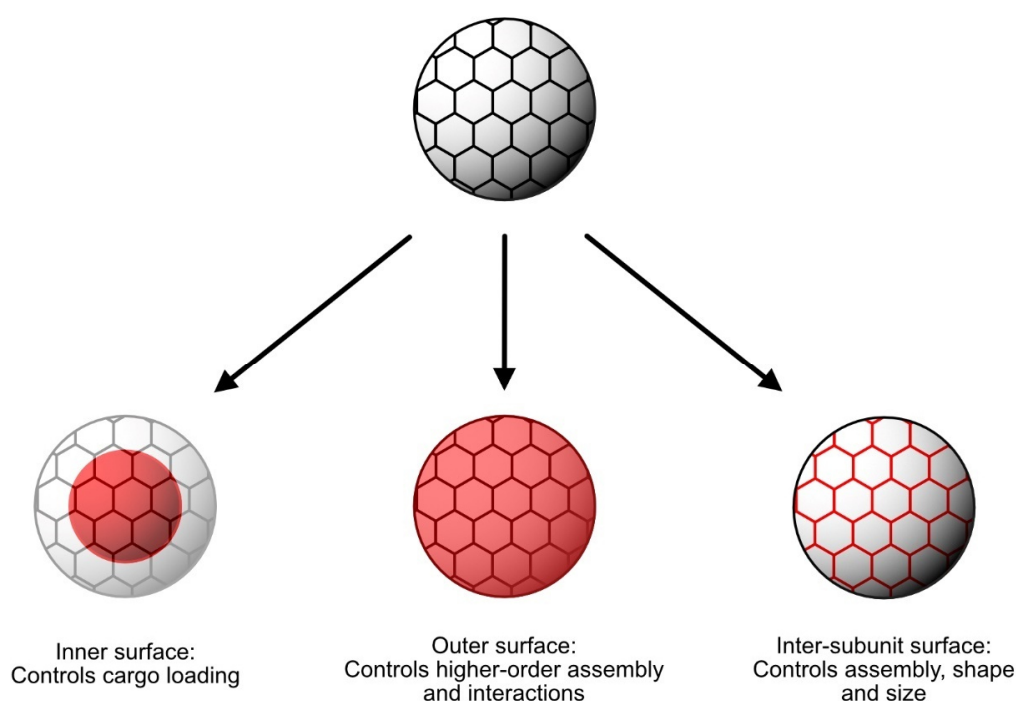


Figure 2.3: Schematic visualization of the three distinct surfaces of a protein cage. Each surface controls different properties of the cage. The inner surface controls cargo loading. The outer surface interacts with environment or higher-order assembly of the cages and inter-subunit surface controls assembly of subunits into cages affecting shape and size of the cage. (Adapted from reference 66 with permission from Wiley and Sons, copyright 2007)

The inter-subunit surface controls assembly of the container. Modifications at this site can result in tremendous changes in size,^[67] symmetry,^[68] or assembly and disassembly behavior. Introducing only a few mutations at the inter-subunit surface of *Aquifex aeolicus* lumazine synthase (AaLS), which is natively built from 60 subunits, leads to the construction of cages built up from 180 or 360 subunits.^[69,70] The inter-subunit interface is crucial for *de novo* construction of artificial cages which can be either achieved by fusion of proteins, yielding an

appropriate symmetry,^[53,71,72] or by virtual docking and redesign with computational macromolecular modeling programs.^[73-75] On the other hand, the outer surface is responsible for the interaction of the container with its environment and governs formation of higher-order structures. The uniform size, structure and physiochemical properties of the cages are highly beneficial for the formation of well-defined structures.^[32] Higher-order cage assemblies could first be observed in protein crystals built from rod-like viruses.^[71,76] Electrostatic interactions or metal coordination are main driving factors of protein assembly.^[77] Formation of higher ordered structures from intrinsic negatively charged cages with other positively charged materials could be observed for cowpea chlorotic mottle virus (CCMV) and tobacco mosaic virus (TMV) e.g. with positive charged gold nanoparticles.^[78,79] Similar results were achieved with cationic dendrimers,^[80-82] divalent metal ions^[83,84] or positively tagged monomeric proteins.^[85] The surface charge of proteins can be altered by the exchange of surface amino acids with charged amino acids.^[86-88] By use of this strategy, positive and negative variants of the human heavy chain ferritin could be fabricated, which are able to self-assembly into a three-dimensional binary lattice.^[89] The ability of some amino acids to form complexes with specific metal ions can be used to create unique architectures. This was demonstrated by layer-by-layer assembly of AaLs,^[90] the formation of one-dimensional nanotubes from chaperonin GroEL,^[91,92] or by the creation of a library from different ferritin lattices in presence of various ligands or metal ions.^[93]

Modification of protein cages with viral antigens on the outer surface allows to mimic viruses to serv as safe substitutes for whole viruses. This technique was recently applied as a potential vaccine candidate for SARS-Cov-2 virus by fusing the SARS-CoV-2-spike protein to the outer ferritin surface.^[94] In first trials, the obtained constructs were able to trigger a protective immune response.^[95,96] Another intriguing application involving expression of peptides on protein cages outer surface is phage display. This screening method displays a commonly used method for epitope mapping,^[97] analysis for protein-protein interactions,^[97] selection of receptor agonists and antagonists,^[98] or development of peptides sequences that can catalyze the mineralization of inorganic materials like zinc sulfide, selenide^[99] or silica.^[100] Mineralization peptides fused to the outer surface of a protein cage can also be used to build an inorganic matrix around a protein container.^[101] Changing the composition of amino acids on the outer surface can also be used to “camouflage” the protein cage and improve immune tolerance, which is important for several biomedical applications and was successfully demonstrated with adeno-associated virus, a common delivery vehicle for gene therapy.^[102] Outer-surface modifications can enable the cage with cell targeting properties useful for drug-delivery applications. Accordingly, receptor-mediated endocytosis could be shown for capsids modified with an Arg-Gly-Arg motif or other peptide based moieties^[103,104] and small molecules.^[105]

The inner surface controls the interaction between container and cargo. A variety of different types of native and non-native cargo can be encapsulated in protein compartments, for instance: nucleic acids,^[106] proteins and enzymes,^[107,108] small molecules,^[109] drugs,^[110] but also inorganic^[111] or polymeric materials.^[112] Consequently, multiple encapsulation strategies can be adopted. Cargo loading can be achieved *in vivo* by co-expression of the cage with a protein cargo.^[113-115] Loading can likewise be achieved *in vitro*, allowing a more precise control of reaction conditions.^[57] Smaller cargo, like small molecules, can be encapsulated by passive diffusion, if the container process a sufficient porosity for efficient loading.^[113] For larger cargo, such as nanoparticles or proteins, the cage can be disassembled by weakening interactions between subunits followed by a reassembly around the cargo.^[116,117]

High encapsulation efficiency of natural and non-natural encapsulation strategies can be achieved by facilitating some sort of interaction of the inner surface with the cargo. The simplest of these interactions is charge complementarity observed, for example, by encapsulation of negatively charged nucleic acids in the positively charged lumen of viral capsids.^[106] Metal nanoparticles with a ligand-mediated surface charge can also be encapsulated through electrostatic interactions, as recently demonstrated by the encapsulation of gold nanoparticles in ferritin.^[118] The efficiency of encapsulation can be increased by the presence of a binding pocket at the inner surface that binds to specific peptide sequences through hydrophobic or ionic interactions.^[119,120] Inspired by these systems, tagging strategies for *in vivo* and *in vitro* encapsulation based on natural^[108] or artificial^[107] peptide tags were developed for proteins^[107] or nanoparticles.^[89] Additionally, varieties of further encapsulation strategies are known based on nucleic acids,^[121] polymers,^[116,122] or even the SpyTag/SpyCatcher technology.^[123]

Encapsulation of small molecule cargo has therapeutic and diagnostic applications in medicine, but also experiences potential use in agriculture for the delivery of pesticides.^[109,124] Loading of small molecules is typically performed *in vitro* by passive diffusion through the pores and binding to the inner surface mediated by covalent or non-covalent interactions.^[117,125] For such loading strategies, a certain degree of cage porosity is required. Cage porosity can further be enhanced by changes in pH, chelating agents or metal ions, inducing a swelling or contraction of the cage.^[126] Charge complementarity allows encapsulation of cationic or anionic drug-like molecules through electrostatic interactions.^[127,128] For the encapsulation of more hydrophobic cargo, integration of a lipid-based nanoparticle in a protein cage was done to obtain a cage with a hydrophobic core. This can be achieved either by reassembly of a cage around a lipid nanoparticle guided by targeting-peptides^[129], by charge complementarity^[125,130] or by passive diffusion of the lipid surfactants. For passive diffusion, the protein cage serves as a template for the formation of a lipid-particle.^[131] For covalent linkage of small-molecules to the

inner surface of cages, genetic introduction of cysteines can provide a site-specific reactive anchor site. This is for example applied to encapsulate MRI contrast agents^[132], chemotherapeutics^[133,134] or other therapeutic active agents.^[135] Additional conjugations to the side chains of lysine^[136] and tyrosine^[137] residues are possible. A detailed in-depth description about the conjugation reactions used in the experimental part of this thesis can be found in chapter 2.4.2. The protein cage human heavy chain ferritin used as a scaffold in this work is further described in the following section.

2.4.1 The ferritin protein container

Ferritin belongs to the ferritin superfamily, which can be divided into three subfamilies: classical ferritins (Ftn), the bacterioferritins (Bfr) and the DNA-binding proteins from starved cells (Dps). All members of this family share the same characteristic four-helix bundle fold and self-assemble into spherical cage like structures.^[138-140] The classical ferritin containers are essential for iron mineralization and storage.^[141] Moreover, they can be found in all domains of life except yeast.^[139]

The ferritin subunit is composed of four parallel right-handed helices (A, B, C, D) and a shorter helix (E), which is oriented nearly perpendicular to the other helices^[141,142] as shown in Figure 2.4a. Two subunits assemble antiparallely to each other, forming a dimer. Further assembly involves hexameric and dodecameric intermediates from these dimeric building blocks and is followed by the association of two dodecamers to the complete cage shell.^[57,142-144] The final protein cage is composed of 24 subunits and possesses the geometry of a rhombic dodecahedron with an octahedral 432 point symmetry.^[145] It has an outer diameter (Figure 2.4b) from approximately 12-13 nm and the size of the inner cavity is 6-8 nm.^[146] Small pores with a size from 0.3 to 0.5 nm are located at the intersection of the subunits in front of the 3-fold and 4-fold symmetry axis,^[147,148] controlling the mass transport inside the cavity.^[149] In total, eight hydrophilic 3-fold channels and six hydrophobic 4-fold channels are present.^[150]

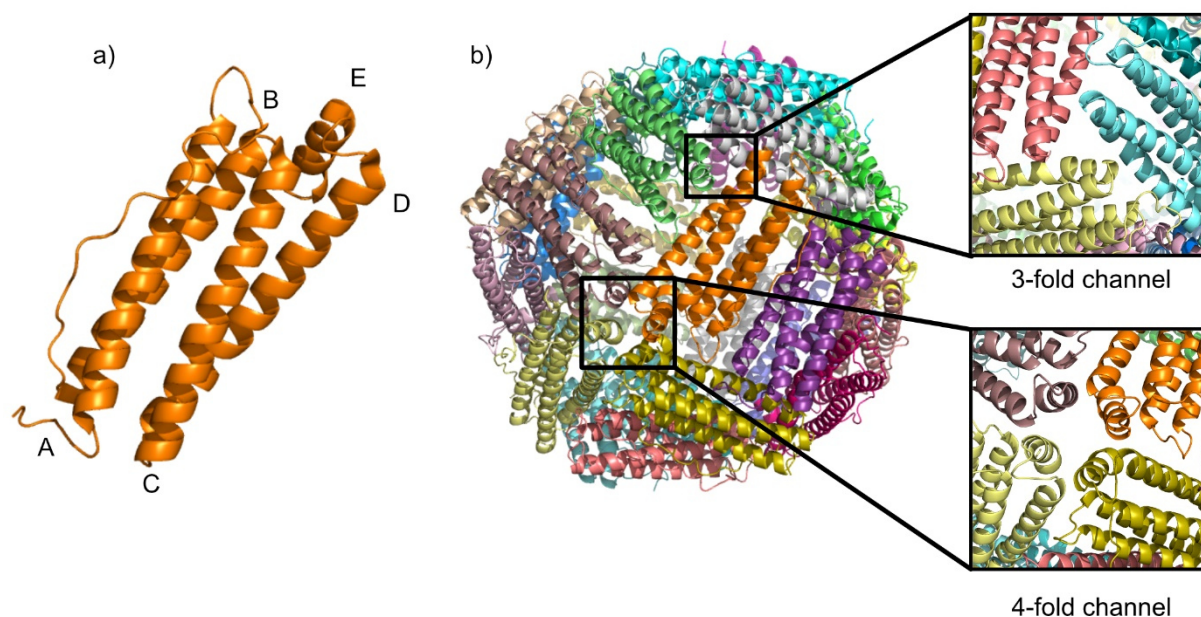
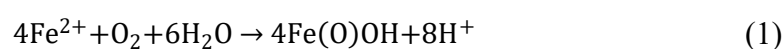


Figure 2.4: Structure of the ferritin cage. a) Subunit of the human heavy chain ferritin (PDB ID: 2CEI) with its 4 main helices (A-D) and the short E helix b) Complete container assembled by 24 subunits with its 3-fold and 4-fold channel located at the intersection of three and four subunits.

The key feature of ferritins lies in their ability to catalyze oxidation of soluble ferrous (Fe^{2+}) to ferric (Fe^{3+}) iron ions and its storage as ferric oxyhydroxide.^[151,152] Oxidation takes place at the oxidoreductase (ferroxidase) site located at the center of the four-helix bundle in the inner cavity. The active site consists of six highly conserved residues forming a carboxylate-bridged di-iron center.^[151,153] Iron ions enter the cavity via negatively charged hydrophilic 3-fold channels.^[154] Additionally, a field gradient around the pores guides metal ions towards the pore entry and the catalytic active ferroxidase site. Here, the ferrous iron is oxidized with molecular oxygen.^[155] The overall reaction can be described by equation (1).^[156]



Even though the actual biological process is much more complex and not fully understood,^[157] there are at least three reaction pathways depending on the rate of iron uptake with intermediate H_2O_2 formation.^[158] The resulting core consist of up to 4500 $\text{Fe}(\text{II})$ ions atoms.^[121]

Mammalian ferritins are heteropolymers consisting of two different subunits: the so-called Heavy chain (H-chain) and Light chain (L-chain) subunits. In contrast to its heavier counterpart, the L-chain lacks the catalytic active ferroxidase site but provides a favorable nucleation site on their inner surface.^[158,159] The ratio between L- and H-chain governs the function of the ferritin. The L-chain rich ferritin is suitable for long term iron storage and can be found mainly

in heart tissue. Thus, higher concentration of H-chain containing ferritin subunits are abundant in tissue where a fast iron metabolism is required, for instance in liver tissue.^[153]

Ferritin is an intriguing system for biomedical application and a multitude of applications in the field of drug delivery, cancer therapy, and bioimaging have been investigated. Advantages of ferritin-based formulations are for instance their safety profile defined *in vivo* pharmacokinetics and easy scale-up.^[160] The outer surface of ferritin can be functionalized with targeting ligands,^[161] but unfunctionalized ferritin is already capable of specific targeting of tumor cells enabling applications in tumor diagnostic and therapy.^[162] This selectivity is possible due to overexpression of the native binding receptor for human ferritin (transferrin receptor 1)^[163] in tumor cells. Ferritin can also be integrated in polyelectrolyte microgels to increase their bioavailability, thus protection from enzymatic degradation.^[164] The inner cavity of ferritin has been loaded with a variety of species depending on the desired application. Loading strategies can be divided into three groups: 1) passive diffusion through the pores, 2) disassembly and reassembly of the cage around the cargo, and 3) biomineralization of inorganic materials inside the cavity. Biomineralization can be used for imaging or diagnostic applications by the internalization of Gd_2O_3 ^[165], Co_3O_4 ^[111] or Fe_3O_4 ^[166]. Additionally, encapsulation of CuS for photothermal therapy^[167] or $LuPO_4$ for radioimmunotherapy^[168] is possible. The disassembly/reassembly strategy has been used for encapsulation of various drug like small molecules like doxorubicin,^[110] epirubicin,^[169] or curcumin.^[170] The encapsulation of siRNA for gene slicing in tumor cells was demonstrated.^[171] Moreover, encapsulation of larger cargo like proteins, for example, cytochrome c^[172] or green fluorescent protein (GFP) fused to different enzymes^[173], or dendrimers such as poly(amidoamine)^[174] have been published.

In this work, a negatively supercharged version of the human heavy chain ferritin^[89] is used as a building block for functional materials with applications in blood purification. The outer surface will be left unchanged allowing to rely on already established assembly methods, for instance crystallization. Functionality will be introduced to the inner cavity through functionalization of introduced anchor sites and protein redesign guided by computational methods. The theoretical basis for these approaches will be briefly described for the functionalization strategy and more in detail for the protein redesign.

2.4.2 Bioconjugation

Adding new functionalities to proteins through site and chemo selective modifications has become an indispensable tool for life science and therapeutic research.^[175] A great variety of reactive agents forming covalent bounds with functional groups present in proteins are known. Most of them focusing on the amino group of lysine or the thiol group of cysteine.^[176] However,

new reaction pathways targeting tyrosine are emerging.^[177] Introduction of non-canonical amino acids further expand the toolbox of possible chemical reactions for bioconjugation.^[178] Nonetheless, a complete overview of bioconjugation techniques would exceed the scope of this subsection. Therefore, the upcoming section focuses only on the reactions used in this thesis.

Within this work, bioconjugation is facilitated for modification of thiol groups to introduce hydrophobic molecules to the protein cavity and for cross-linking of amino groups to enhance the stability of the generated adsorbents. Hydrophobic molecules were conjugated to protein cysteine groups using an α -halocarbonyl (α -haloacetamide) linker. Those linkers possess a halogen atom in β -position relative to the amid nitrogen atom, which can be easily displaced by an attacking nucleophile.^[179] These reagents are widely used in mass spectrometry for alkylation of cysteines, which is useful in peptide mapping.^[180] A typical scheme for such reactions is shown in Figure 2.5. The relative reactivity with respect to the halogen atom increases with higher homologues (I > Br > Cl > F). The reaction can also proceed with amines, but the reactivity can be governed by the degree of ionization and thus via pH value. At neutral pH values, reaction with thiol groups are favored.^[181]

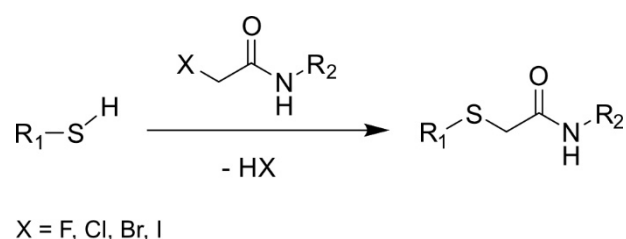


Figure 2.5: Bioconjugation with an α -halocarbonyl agent. The thiol group of a cysteine in the protein sidechain forms a covalent linkage with an α -halocarbonyl agent.

For conjugation of various dye molecules to the protein, a thiol selective maleimide linker was chosen. The thiol-maleimide click reaction follows the mechanism of a Michael-type addition reaction. During the initial step, a thiolate anion is generated by deprotonation of the thiol group caused by catalytic concentrations of a weak base or nucleophile. The strong nucleophilic thiolate anion attacks the double bond of the maleimide forming an enolate intermediate. The basic enolate ion deprotonates another thiol group forming the desired product and another reactive thiolate species.^[182] The mechanism is shown in Figure 2.6.

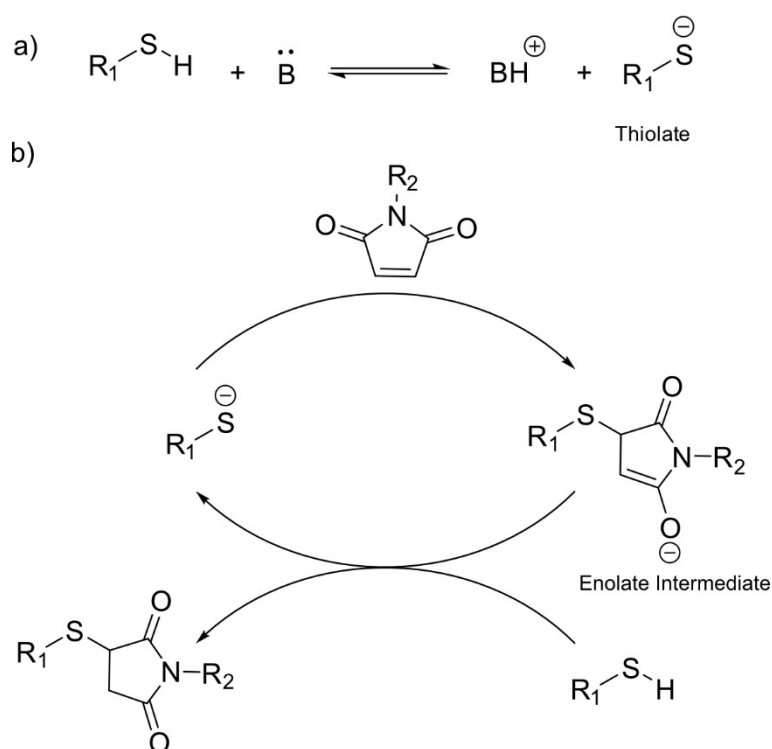


Figure 2.6: Mechanism of thiol-maleimide click reaction. a) Initial formation of a thiolate anion by catalytic amounts of base. b) Catalytic cycle for the addition of a thiolate to an N-substituted maleimide. (Adopted from reference 182 with permission of the Royal Society of Chemistry, copyright 2010)

The reaction is pH dependent and selective for thiols at pH values between pH 6.5 to 7.5. At more basic conditions above pH 8.5, reactions with primary amines are favored.^[183] This side reaction is used in this work for cross-linking of protein crystals with a bifunctional sulfosuccinimidyl-4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) crosslinker. Besides the maleimide group, the sulfo-SMCC linker carries an *N*-hydroxysuccinimide (NHS) ester group. This active ester group is susceptible towards nucleophilic attacks by primary amines present in the protein forming a zwitterionic intermediate followed by displacement of the NHS group and the formation of the acylated product. Respective reaction scheme is shown in Figure 2.7. Side reactions with thiol or hydroxyl groups present in the protein are possible but not thermodynamically stable, thus the reaction is selective for amino groups.^[181] Sulfo-SMCC was used for chemical fixation of protein crystals, since the more common glutaraldehyde yield crystals which possess no adsorption capacity.

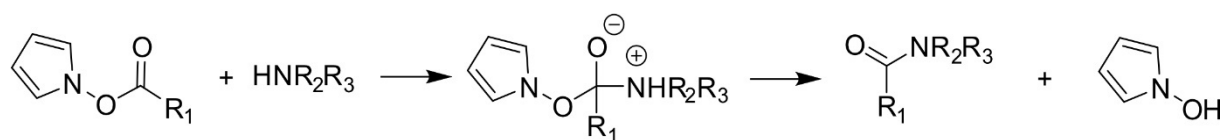


Figure 2.7: Reaction scheme for active ester with primary or secondary amines. First the amines perform a nucleophilic attack on the ester group leading to the formation of a zwitterionic intermediate. Upon displacement of the NHS group the acrylated product and *N*-Hydroxysuccinimide are formed

Glutaraldehyde is a bifunctional reagent with two terminal aldehyde-groups (Figure 2.8). In aqueous solution, multiple monomeric and polymeric reactive species of glutaraldehyde are in equilibrium.^[184] Each of them can react via a different mechanism with the amino group of lysine residues. In addition to amino groups, the reaction is also possible with thiol, phenol or imidazole moieties. Currently, it is not known which structure and mechanism is mainly responsible for cross-linking. As shown in Figure 2.8, under acidic and neutral conditions free glutaraldehyde (I) can react under formation of a Schiff base. Nevertheless, the resulting product is not stable. Taken together, a nucleophile substitution of hydroxyl-groups from monomeric (II) or polymeric cyclic hemiacetals (II) with the protein amino groups is more likely.^[184] Under basic conditions, internal aldehyde groups of polymeric α,β -unsaturated aldehydes react either under formation of a resonance stabilized Schiff bases^[185] or the C-C double bond undergoes Michael addition with the amine group.^[186] Glutaraldehyde is widely used for inter- and intramolecular cross-linking and is also applied in clinical applications.^[187,188] To enhance the stability of the protein-based materials used in this work, chemical cross-linking via glutaraldehyde is applied. Beside increasing the stability of protein crystals, glutaraldehyde is used in this thesis to produce a chemical aggregated solid protein material. This technique has also been used to produce solid biocatalyst from trypsin,^[189] papain,^[190] and alkine proteinase.^[191]

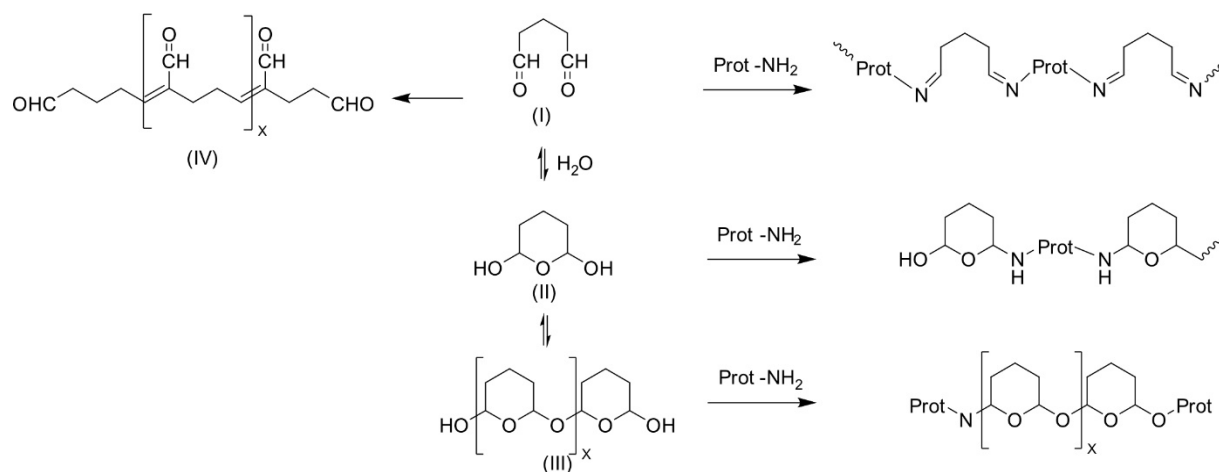


Figure 2.8: Glutaraldehyde structures and possible reaction pathways. Free glutaraldehyde (I) can react under acidic and neutral conditions with proteins under formation of Schiff bases but can also undergo intermolecular aldol condensations to α, β-unsaturated oligomeric aldehydes (IV) or form hemiacetals (II) or hemiacetal oligomers (III). Hemiacetal derivatives can cross-link proteins by nucleophilic substitution of their hydroxy groups with amino groups from the protein.^[184,192] (Adapted from reference 184 with permission of the Royal Society of Chemistry, copyright 2011)

Summary

- Key properties like cargo-loading or higher-order assembly of protein cages can be independently modified by altering of the distinct surfaces (outer, inner, or inter-subunit surface) of the cage.
- Ferritin has an outer diameter of 12 nm and a 7-8 nm large cavity. Up to 14 pores allow the diffusion of small molecules like the PBUTs into the inner cavity. It has a high biocompatibility and is already investigated for many biomedical applications.
- The stability of proteins and protein crystals can be increased by chemical cross-linking.
- The inner cavity can be modified to increase binding affinity towards the PBUTs, for instance by covalent linkage of molecules via bioconjugation techniques using the thiol-groups of cysteines as a reactive handle.

Another important technique for increasing the affinity to the PBUTs is the exchange of amino acids at the inner protein surface. Exchanging many amino acids while retaining the structure of the cage is a difficult challenge. For guidance of protein design decisions, the molecular modeling software Rosetta is used described in the following chapter.

2.5 The Rosetta molecular modeling software suite

Rosetta is a software suitable for modeling biological macromolecular structures such as proteins or nucleic acids. Rosetta is not a monolithic program, but a large collection of computer code from different authors. Originated as property of the University of Washington, Rosetta transformed to a collaborative project with developing laboratories at over 60 institutes worldwide (www.rosettacommons.org).^[193] Rosetta was originally developed for *de novo* protein structure prediction, but is constantly improved and extended. It can evaluate and rank the physical plausibility of biological macromolecular structures, and although alter these structures.^[194] From this basis, multiple extensions towards other applications have been developed, for instance: loop-modeling,^[195] protein design,^[196,197] interaction of peptides and nucleic acids,^[198,199] design^[200,201], and modeling^[202,203] of antibodies, and protein-protein docking^[204] or protein-small molecule docking.^[205,206]

The basic workflow of Rosetta is based on the assumption that naturally observed biological macromolecule conformations are always in the lowest free energy state.^[207,208] The folded states are located in minima on the energy landscape, possessing a net favorable change in Gibbs free energy relative to the unfolded state.^[209] As the direct consequence of this assumption, structure prediction is the problem of finding the conformation with the lowest-energy for the given amino acid sequence, whereas designing is the problem of finding an amino acids sequence with the lowest energy level for a given structure. The key to solve both of these problems lies in an accurate free-energy function and a sampling method that is capable to locate minima in the energy function.^[210]

2.5.1 Basic concepts of Rosetta – the core library

Rosetta is residue-centric, meaning that all atoms in the macromolecule are aligned to a given residue while the residues display the smallest unit of scoring. Information about the macromolecule is consequently stored as a set of residues in the so-called pose class.^[211] Rosetta uses internal coordinates to describe atom positions in the given pose instead of Cartesian coordinates. In internal coordinates, each atom position can be described by three parameters (bond length, bond angle and torsion angle) in dependency of a set of reference atoms (typically bounded neighbor atoms). Since the bond angle and length can be approximated as fixed, the use of internal coordinates allow a drastic reduction in degrees of freedom (DOF), from three per atom (x,y, and z position) to one (only the dihedral angles about rotatable bonds). This procedure reduces the search space and increases computational performance.^[212] Since changes in one atom's internal coordinates also influences neighboring atoms, propagation of changes through the atomic system must be explicitly defined. This

kinematic connectivity can be defined by a tree (a connected acyclic graph). The nodes of this tree represent atoms and the edges display kinematic connections. The internal coordinates in this so-called kinematic tree are mapped onto the individual atoms relative to a reference frame, which is defined by the three upstream atoms. In Rosetta, this tree is called AtomTree. By starting at a defined root node and moving downstream, the AtomTree allows conversion of the internal coordinates to Cartesian coordinates. To enhance computational speed, the AtomTree also tracks which DOF have changed since last scoring and if positions of some atoms with respect to each other have not changed, thus their already calculated energy can be reused.^[211]

2.5.1.1 Rosetta protocols, mover and sampling

For modeling biological macromolecules, Rosetta searches the energetic landscape, simulated by the energy functions, for global minima. The sampling strategy performed by Rosetta depends on the specific scientific objective. The majority of the sampling algorithms are based on the Metropolis Monte-Carlo algorithm, followed by gradient-based minimization for the last refinement.^[213] Subsequently, Rosetta randomly samples the search space, making it necessary to generate a large number of independent trajectories, followed by careful evaluation of the results.^[194]

A typical sampling protocol is outlined in Figure 2.9. It starts with a structure from a biological macromolecule (the pose). Dependent on the desired application, the pose can be altered by movers. Movers can alter the whole pose or only one residue. Most movers exclusively change 3D conformation. However, EnzymaticMover, a tool to exchange the amino acid residues necessary for design applications, exists. Analysis Movers that do not change the pose but report about geometric energies are another type of movers. So-called ContainerMovers can also be used to arrange other movers for more complex applications. Some movers also contain whole protocols¹, which is useful for some applications such as protein docking.^[214]

¹ A list of all available mover can be found under https://new.rosettacommons.org/docs/latest/scripting_documentation/rosettascripts/movers/movers-rosettascripts

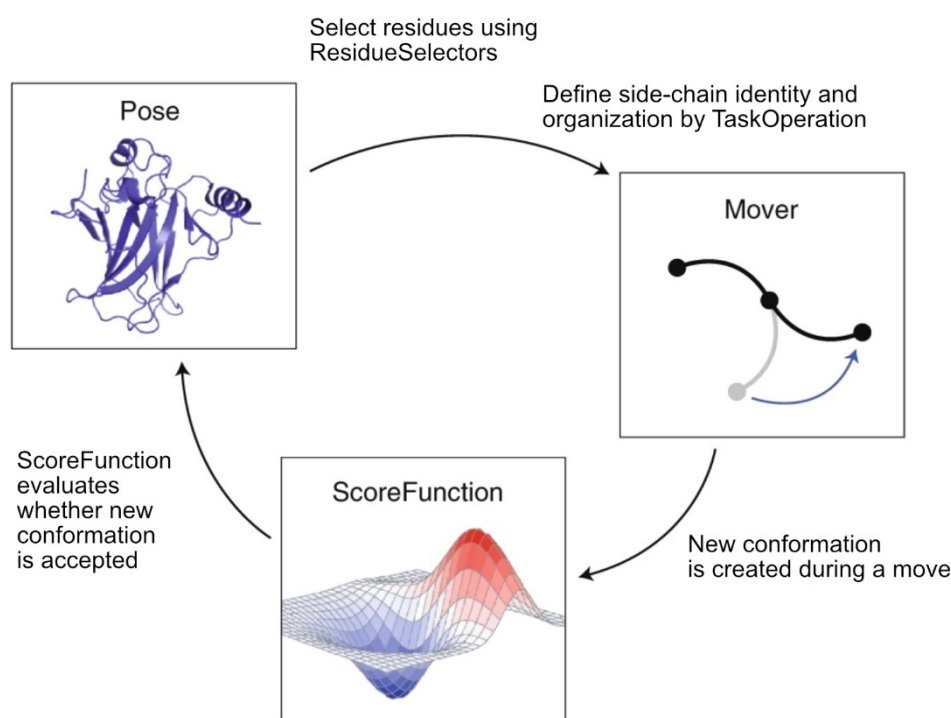


Figure 2.9: Main elements in a Rosetta protocol. The structure of a biomacromolecule (pose) is altered by a defined mover and subsequently scored by the Rosetta score function. The scores of the new and old structures are compared and the move is either accepted or rejected. This process is repeated a several times until no further score improvement can be found. (Adopted from reference 194 with permission of Springer Nature, copyright 2020)

After altered by the Mover, the pose is evaluated by the Energy or Score function. Depending on the energy score, the altered pose is either accepted or rejected based on the Metropolis criterion shown in equation (2). If the energy of the new pose is lower than the original, the pose is directly accepted. If the energy of the original pose is higher, the move can still be accepted, but only with a given probability P dependent on the difference between the original energy score E_{orig} and the new one E_{new} along with the temperature T .^[194]

$$P = \exp\left(-\frac{E_{new} - E_{orig}}{T}\right) \quad (2)$$

Rosetta contains predefined protocols for various applications. Examples used in this work are Fixed backbone (Fixbb) design for protein redesign or the Relax protocol for energetic optimization of the structure. Additionally, it provides script interfaces to write own xml-based protocols. RosettaScripts allows users to tailor Rosetta for a specific problem without

² Since this is only applied when $E_{new} < E_{orig}$ the term in the exponential function will always be negative. As a result the probability to accept the move declines rapidly with increasing difference in energy.

requirement of extensive C++ knowledge to change the source code of Rosetta.^[215] These xml-scripts are used in this work to model protein-ligand and protein-protein docking.

2.5.1.2 The energy functions

Rosetta's energy function, which combines knowledge- and physics-based terms, is constantly improved aiming for enhancement of computational speed, extensibility and accuracy.^[194,216]

The first version developed 1997 by Simon *et al.* used statistic potentials describing individual residue environments and frequent residue-pair interactions derived from the Protein Data Bank.^[217] The energy function was later on expanded by terms for van der Waals forces and hydrogen bonding.^[218] Nonetheless, these terms are only suitable for low-resolution modeling, meaning that only coordinates of the protein backbone atoms are included in the calculated energy. Within this equation, side chains and their interactions were treated only implicitly. In this low-resolution energy function, amino acid side chains are treated as centroids.^[217] This makes the energy landscape rather smooth at the expense of its accuracy.^[213]

Higher-resolution modeling, required for design and docking applications, was enabled by Kuhlman and Baker with implementation of an all-atom energy function. This function is capable of modeling atomic packing, hydrogen bonding, solvation and protein torsion angles commonly found in folded proteins. Combination of traditional molecular mechanics energies with statistical derived potentials energies (see Table 2.2) allowed Rosetta to reach several milestones in structure prediction and design.^[209]

The most recent version of Rosetta's all atom energy functions is called Rosetta Energy Function 2015 (REF15). It calculates the energy ΔE_{total} of a biological macromolecule by linear combination of energy terms E_i , which are determined as function of geometric degrees of freedom (Θ) and chemical identities (aa). In addition, each individual energy term has an attributed weight w_i to enable fine tuning of the different terms. Equation (3) describes the relationship between the terms.^[194]

$$\Delta E_{total} = \sum_i w_i E_i(\Theta_i, aa_i) \quad (3)$$

All energy terms necessary for protein modeling in REF15 are summarized in Table 2.2 at the end of this section and are briefly described in the following.

The terms `fa_atr` and `fa_rep` describe van der Waals interactions, which depend on the distance between the observed atom pair. At larger distances, attractive forces, arising from

the cross-correlated motions of electrons in neighboring nonbonded atoms, are dominant. At very short distances, repulsive forces become more prevalent, since electrons cannot occupy the same orbitals because of Pauli's exclusion principle. Rosetta uses a Lennard-Jones 6-9 potential to calculate the interaction energy of atoms. Though, potential is split at its minima yielding one term for attractive (fa_atr) and one term for repulsive (fa_rep) interactions, enabling separate weighting of both interactions. An additional term fa_intra_rep similar to fa_rep , but for repulsive van der Waals forces between atoms of the same residue, is present.^[209]

Electrostatic interactions between fully or partially charged residues are calculated by Coulomb's law with partial charges and a few modifications. One modification represents the difference in dielectric constant between the protein core and the solvent-exposed surface. Substitution of the dielectric constant by a sigmoidal function, increasing from the core to the solvent, allows simulation of this circumstance.^[209,219] The electrostatic energy is represented by the fa_elec term in the energy function.

Burial of hydrophobic groups away from water, to a close, nearly void-free packing is a main driving force in protein folding.^[210,220] Nevertheless, modeling interaction of all solvent molecules with the protein is computationally too expensive. That is why Rosetta displays the solvent as bulk water based on the Lazardis-Karplus implicit Gaussian exclusion model.^[221] To account for distinct chemical environment for the solvent atoms, the solvation energy has different components. An isotropic solvation energy, called fa_sol , deduces uniformly distributed water around all atoms with an additional intraresidue version fa_intra_sol . For modeling solvation shells near polar atoms, an anisotropic solvation energy lk_ball_wtd is used.^[209]

Hydrogen bonds are crucial for protein folding, function and interaction. However, due to their hybrid covalent-electrostatic character, modeling is difficult. The electrostatic proportions of the hydrogen bonds are calculated by using the already described fa_elec term.^[209] The covalent character depends on geometric concerns to maximize orbital overlapping.^[222] Rosetta uses an empirically model derived from polar contacts in approximately 8000 high-resolution crystal structures^[223,224] (Top8000 data set) to consider these geometries.^[225] The hydrogen bonding energy is subdivided in four different terms: long-range ($hbond_lr_bb$) and short-range ($hbond_sr_bb$) backbone hydrogen bonds, hydrogen bonds between backbone and side-chain atoms ($hbond_bb_sc$) and hydrogen bonds between side-chain atoms ($hbond_sc$).^[209]

Disulfide bonding is crucial for certain protein folds. The `dslf_fa13` term is modeling these interactions for pairs of cysteines and computes their energy. The model itself is again knowledge-based and empirical derived from the Top8000 data set similar to hydrogen bonds.^[209]

To model backbone and side-chain torsions, Rosetta uses several knowledge-based terms. The term `rama_prepro` evaluates the backbone dihedral angles ϕ , ψ . It is based on Ramachandran maps for each amino acid making use of torsions from 3985 protein chains.^[209,226] The angle probabilities are converted to energies via the inverted Boltzmann relation.^[227] Depending on the existing ϕ, ψ backbone conformation, the propensity to observe an amino acid relative to the other 19 canonical amino acids is represented by the `p_aa_pp` term.^[209,228] The probability to find a specific side-chain conformation for given backbone ϕ, ψ values is described by the term `fa_dun`.^[209] The probabilities are derived from the 2010 backbone-dependent rotamer library, which contains the frequencies, means and standard deviations of individual χ angles.^[229] The dihedral angle ω of peptide bonds has a partial double-bond character and is fixed in cis or trans conformation. This rotational energy barrier is modeled by the term `omega` with a backbone-dependent harmonic penalty centered near 0° for cis and 180° for trans.^[209]

Proline rings require careful attention, since they cannot be represented by a kinematic tree used in Rosetta. Therefore, the term `pro_close`, ensuring proline closure, is present. The hydroxyl hydrogen atom of tyrosine amino acids prefers to stay in plane of the aromatic ring. Considering this preference, the `yhh_planarity` terms implements a sinusoidal penalty to avoid deviation of the X_3 angle.^[209]

For protein design applications, to compare sequences that differ in their amino acid composition for the same target sites, the energy gap between the folded and unfolded state is required. The unfolded state reference energy is implemented in the term `ref`, calculated as a sum of individual unfolded-state reference energies ΔG_{ref} for each amino acid.^[209] These values are derived from empirical optimization, maximizing native sequence recovery during design simulations on high-resolution crystal structures.^[216,230] The `ref` energy term contains weights for each amino acid. By changing these weights accordingly, designs favoring a specific amino acids, as done for supercharging of ferritin variants, are possible.^[88,231]

Table 2.2: Terms in REF15 energy function (Adopted from reference 210 [209] with permission from the American Chemical Society, copyright 2017)

Term	Description
Physic-based terms	
fa_atr	Attractive energy between two atoms on different residues
fa_rep	Repulsive energy between two atoms on different residues
fa_intra_rep	Repulsive energy between two atoms in different residues
fa_sol	Solvation energy between protein atoms in different residues
lk_ball_wtd	orientation-dependent solvation of polar atoms assuming
fa_intra_sol	solvation energy between protein atoms in the same residue
fa_elec	energy of interaction between two nonbonded charged atoms
hbond_lr_bb	energy of short-range hydrogen bonds
hbond_sr_bb	energy of long-range hydrogen bonds
hbond_bb_sc	energy of backbone–side-chain hydrogen bonds
hbond_sc	energy of side-chain–side-chain hydrogen bonds
dslf_fa13	energy of disulfide bridges
Statistic-based terms	
rama_prepro	probability of backbone ϕ , ψ angles given the amino acid type
p_aa_pp	probability of amino acid identity given backbone ϕ , ψ angles
fa_dun	probability that a chosen rotamer is native-like given backbone ϕ , ψ angles
omega	backbone-dependent penalty for dihedrals that deviate from cis or trans isomerism
pro_close	penalty for an open proline ring and proline ω bonding energy
yhh_planarity	sinusoidal penalty for nonplanar tyrosine χ_3 dihedral angle
ref	reference energies for amino acid types

2.5.2 Applications in this research

2.5.2.1 Protein relaxation

In this work, high-resolution crystal structures were used as input poses for further modeling. To ensure reliable calculations, conformation of these structures first need to be optimized to reach a global energy minimum. This was achieved with the Rosetta Relax protocol. During this iterative relaxation protocol, the ϕ and ψ backbone angles are slightly perturbed but the overall global conformation of the protein is maintained. Additionally, the side chain rotamer angles are adjusted by a simulated annealing Metropolis Monte-Carlo search of the rotamer space and a gradient minimization to all torsion degrees of freedom. Thus, the structure is incrementally moved to its nearest energy minimum by increasing the repulsive part of the van der Waals potential.^[213]

2.5.2.2 Protein redesign

For changing given amino acids of the pose, fixed backbone sequence optimizations were performed. Here, a Monte-Carlo simulated annealing search was performed over the whole amino acid search space. Every amino acid is considered at each position in the sequence and rotamers are constrained to the Dunbrack Library.^[213] This protocol can be performed to enhance thermal stability of a protein.^[232] Additionally, by specifying the amino acids and positions for redesign, it can be used to alter other properties of the protein like implementing surface charges^[88,231] or to increase the density of hydrophobic side chains as carried out in section 5.2.1.1.

2.5.2.3 Small molecule docking

Many ligand-docking applications, developed to seek predictions of interactions between a protein and a small molecule, struggle to simulate the flexibility of proteins. Since Rosetta is designed for *de novo* structure prediction, it can treat protein flexibility with high accuracy.^[213,232] Additionally, small-ligand flexibility can be introduced to Rosetta.^[233] The Rosetta Ligands algorithm starts with a random ligand conformer from an ensemble of provided ligand conformations, which is moved to a user-defined putative binding site. First, a low-resolution shape complementary search is performed, where the ligand is rotated and translated to fit the protein binding site, searching to optimize attractive and repulsive score terms. Next, the ligand pose is perturbed and side chain rotamers are optimized through Monte-Carlo minimization cycles by a high-resolution docking algorithm. In the last step, a gradient minimization of all torsion degrees of freedom is carried out.^[234] In this thesis, an adopted version of the script in form of an RosettaScript interface is used.^[235] This allows

additional sampling over different amino acids to improve the binding site, scriptable change of the starting position, and the change of the energy function from REF2015 to Talaris14, which performed better at small molecule docking.^[236] This technique was used for the fabrication of PBUTs binding sites in the ferritin cavity in section 5.2.1.2.

2.5.2.4 Protein-protein docking

A variety of biological processes in living cells are determined by interactions of different proteins. Generating the 3-D structure of attributed protein-protein complexes and modeling their interactions can generate invaluable insights into underlying mechanism, functionality, and organization, and guide development of diagnostic and therapeutic targets. Due to the great research interest, varieties of tools for prediction of protein-protein interaction have been developed, one of them is integrated in the Rosetta software suite.^[237,238]

In this work, a slightly adapted version of the Monte Carlo based multi-scale docking algorithm RosettaDock in form of a flexible Rosetta script is used.^[239] This script allows easy changes to the algorithm and further opens possibility to expand the protocol for later design applications. The algorithm starts with low-resolution docking step, in which protein side chains are abstracted by a single unified pseudo-atom (centroid). Moreover, one protein is translated and rotated relative to the other.^[239] Resulting poses are scored by residue pair interaction and environment statistics as well as repulsive and attractive van der Waals forces by Rosetta's low-resolution scoring function. This aims to find a threshold acceptance for prior performed rigid-body moves.^[215] The lowest-energy structure is forwarded to a high-resolution docking step with complete modeled protein side-chains. It is performed by allowing random rigid-body translation, Monte Carlo based side chain rotamer sampling and gradient-based rigid-body minimization.^[213,239] The protocol based on this algorithm is used to determine the interactions at the interfaces between differently charged ferritin cages in a protein crystal as described in section 5.3.

Summary:

- Rosetta mainly uses a Metropolis Monte-Carlo algorithm as sampling method for changing biomacromolecules structure and evaluates the output with a score function to search for a global energy minimum. Therefore, many independent trajectories need to be calculated and the output is analyzed based on the scientific objective.
- Rosetta was originally designed for protein structure prediction, but by adopting different sampling-strategies, functionality was expanded to many different applications such as protein design, protein-protein docking and protein ligand docking.

3 Basis of this work

The basis for this work was established in the Beck group in previous studies and is shortly summarized here. As a scaffold for production of a protein-based adsorbent, the negative supercharged variant of the human heavy chain ferritin, Ftn^(neg), was used.^[89] This variant was designed by extending a method first described by Miklos *et al.*^[231] using the Rosetta software suite to the ferritin nanocage.^[89] For the design, a fixed backbone protocol was used and the weights for the reference energies given in the energy function were gradually changed to favor negative charged amino acids.^[89] Additionally, in previous work optimal protein production parameters and a tag-free purification protocol was established to produce the protein variant in large quantities and high purity. Furthermore, conditions for hanging-drop crystallization yielding protein crystals suitable for structure prediction were established and published in literature.^[89] Following these results, in unpublished works by *Marcel Lach* these conditions were adapted to a batch crystallization process based on a protocol from *Rayment*.^[240] This procedure allowed the scale-up of the protein crystal production.

These previous results constitute the basis for the present project: The assembly strategy for the production of a solid protein-based material is further extended and the main focus can be applied to enhance the PBUTs binding affinity to the inner surface. The strategies investigated to reach this goal are introduced in the following chapter.

4 Concepts and aim of this thesis

Research conducted in this thesis focused on creating a protein-based adsorbent for the clearance of PBUTs from the blood of CKD patients. The three main requirements for the adsorbent are a high biocompatibility, a high binding affinity and an easy and cost-effective method to separate the adsorbent from the blood. As a scaffold, the negatively supercharged variant of the human heavy chain ferritin, Ftn^(neg), is used. Due to the small-sized pores of the ferritin, only small molecules such as the PBUTs can enter the cavity. Due to this size exclusion effect, the inner surface can be modified to bind the PBUTs, while plasma proteins and blood cells will only interact with the outer surface. This design principle ensures a high biocompatibility independently from the modifications of the inner surface. Since most of the serum protein are negatively charged, the decision was made to use the negatively supercharged variant to minimize unspecific adsorption on the adsorbents. No further changes should be made to the outer surface, since the human heavy chain ferritin is native to the blood serum and therefore possess an intrinsic high biocompatibility.

Through the outer surface, the assembly of the protein cages to free-standing 3D materials can be controlled, and the resulting solid material can be applied as a heterogeneous adsorbent ensuring easy separation. Due to its bio- and hemocompatibility, the material can potentially be immobilized in a cartridge and integrated in conventional dialysis machines, keeping additional cost low. Fabrication of this macroscopic material is achieved by relaying on already established batch crystallization routes, resulting in crystalline protein material with defined pores and channel through the material. The resulting highly ordered and uniform material is well-suited to compare changes in PBUT adsorption caused by different changes to the inner surface. The impact of the higher order assembly, especially in terms of diffusion through the material will be investigated by comparing this crystalline material with non-crystalline protein assemblies. The latter are synthesized on unordered precipitating the proteins through the use of glutaraldehyde. For clinical application, the habitus of the assembled material probably needs more adaptation, but for the early research stages the aforementioned material morphology will be suitable.

Modifications of the protein cage to enhance the adsorption of the PBUTs are limited to the inner surface or cavity of the protein cage. In this way, it is possible to independently investigate methods to increase the affinity towards the PBUTs without interfering with assembly for macroscopic materials or diminishing the biocompatibility. The capability of clearing the PBUTs will be investigated with three selected toxins namely indoxyl sulfate (IS), *p*-cresyl sulfate (*p*CS) and phenylacetic acid (PAA). IS and *p*CS are present in end-stage CKD patients in comparable concentrations between 41 - 44 mg L⁻¹ allowing conclusion in terms of possible selectivity

arising from the different aromatic ring structures. PAA is present in 10-fold higher concentration and comparison to the other PBUTs can allow indirect conclusions about PBUT diffusion through the material. To increase the PBUT adsorption, two different strategies are investigated, namely coupling of hydrophobic molecules to the inner surface, and genetic redesign guided by the molecular modeling software suite Rosetta, which will be described in detail in the following subsections.

4.1 Covalent linkage of hydrophobic molecules to the inner surface

Increasing the hydrophobicity of the inner surface is a promising strategy to increase the affinity towards the PBUTs, which can be achieved by linking hydrophobic molecules to the inner protein shell. The strategy is outlined in Figure 4.1. The PBUTs have hydrophobic and polar characteristics. Therefore, it is not directly obvious, which kind of ligand is ideally suited for toxin adsorption. To allow adaptability, the system is designed to be as variable as possible. The thiol group of the amino acid cysteine can be selectively addressed by various chemicals and is used as an anchor site for the chemical modification. Multiple anchor sites per ferritin subunit can be introduced at various solvent-accessible sites in the cavity, ensuring a high density of linked molecules at the inner surface.

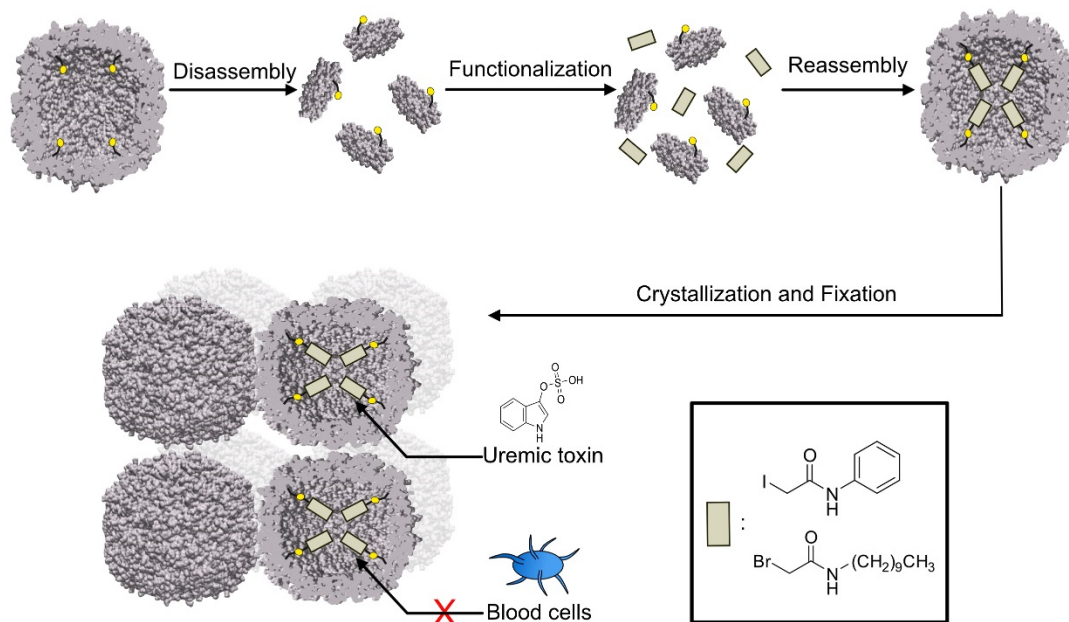


Figure 4.1: Schematic overview of the strategy for protein functionalization. The protein cage is disassembled into subunits and incubated with respective ligand. After reassembly, cages self-assemble into a 3D-material and are chemically fixed.

As reactive agents, α -halocarbonyl derivatives commonly used for the alkylation of cysteine residues in mass spectrometry peptide mapping are viable candidates. For effective coupling,

the protein cages will be disassembled into their subunits, to increase the accessibility of the reactive thiol groups. Derivatives investigated in this work bear an aliphatic chain of ten carbon atoms or a phenyl group and therefore marking an extreme case in hydrophobicity.

4.2 Redesign of the cavity by amino acid exchange

The aforementioned strategy depends on post-processing of the protein cage to allow adsorption of the PBUTs. This is always linked to higher production cost and possible unwanted side-effects from not fully removed byproducts. Redesigning the inner surface of ferritin through exchange of amino acids represents a more elegant strategy to enable the protein cage to adsorb PBUTs without further chemical modification. Research conducted towards this topic is aimed to evaluate the general feasibility of this route and develop strategies for its implementation. Since exchange of many amino acids could lead to improperly folded proteins, computational methods based on the Rosetta software suite are applied to guide the redesign. Two general approaches were investigated. The first one includes selecting a range of possible mutation sites and possible amino acids, and determines the stability of the resulting protein with the Rosetta software. This approach was done to create a ferritin cage with decreased negative surface potential, increased density of hydrophobic amino acids and expanded pores. The other concept embodied consecutive use of ligand-docking protocols to design binding sites combined of positive charged and hydrophobic amino acids capable of stabilizing the toxins at the inner surface. Selected designs will be tested in *in vitro* experiments, evaluating the stability of the designed protein cages and their PBUT adsorption capacities.

5 Results and discussion

This section can be broadly divided into two main parts focusing on the two strategies mentioned in the previous section. First, chemical functionalization of ferritin cages is investigated (section 5.1). Ferritin subunits containing reactive anchor sites are designed and protocols for coupling of hydrophobic molecules to these sites are established. The assembly of the cages to a heterogeneous adsorbent and investigation of its biocompatibility are described in detail. To determine the adsorption capacity of protein-based adsorbents, an analytic method is established and applied to the chemical modified materials. The first part ends by demonstrating the variability of the new protein variant through encapsulation of various dye molecules for future application in cell imaging. The second part is focused on redesigning the protein cage (section 5.2). First, different design ideas are developed, and respective models are generated by computational methods. Next, selected variants were expressed and characterized. Higher-order assembly and biocompatibility is described in short due to its similarity to previous results. Finally, adsorption capacity of adsorbents based on the new variants are determined. In a short final part, a sub project is presented (section 5.3). Here protein-protein docking protocols are established. These are applied to help understand the interactions between protein cages in binary protein crystals.

5.1 Functionalization of the inner protein cavity

In this section, chemical modifications of the inner ferritin surface are investigated. First, general criteria and design ideas are summarized. Next the production and characterization of a Ftn^(neg) variant with reactive thiol groups exclusively at the inner surface is described. Protocols for the functionalization with hydrophobic molecules and characterization methods for respective products are established. Protein-based adsorbents are fabricated from the functionalized Ftn^(neg) and their ability to adsorb PBUTs are quantified. Finally, the variability of the Ftn^(neg)-Cys variants are demonstrated by encapsulation of various fluorophores and possible applications in cell imaging are presented

5.1.1 Rational design of Ftn^(neg)-Cys variants

In order to conjugate hydrophobic agents in high density to the inner surface, a reactive anchor site is required. Multiple functional groups present in the protein can be targeted by reactive agents as summarized in section 2.4.2. In this thesis, thiol groups will be used as anchor sites, which can be targeted by α -halocarbonyl or maleimide derivatives. Native cysteines were removed to prevent functionalization at undesired positions. In Ftn^(neg) only one cysteine is present near the 3-fold channel at position 130. It was replaced with alanine in all Ftn^(neg)-Cys variants.

It is hypothesized that the reactivity of the thiol group increases with increasing accessibility. To discriminate buried from accessible amino acid positions, the surface accessible surface area (SASA) for each amino acid at the inner surface was determined with the *GetArea* webtool.^[241] This web tool compares the accessibility of an amino acid in the protein structure of Ftn^(neg) (PBD ID 5JJK) with the accessibility of the same amino acids in a theoretical tripeptide flanked by two glycine amino acids. The results for each position are given as a percentage of this ideally accessible extreme case and are listed in the appendix in Table 10.1. Up to four positions per subunit are selected adding up to a maximum of 96 reactive sites per cavity as shown in Figure 5.1.

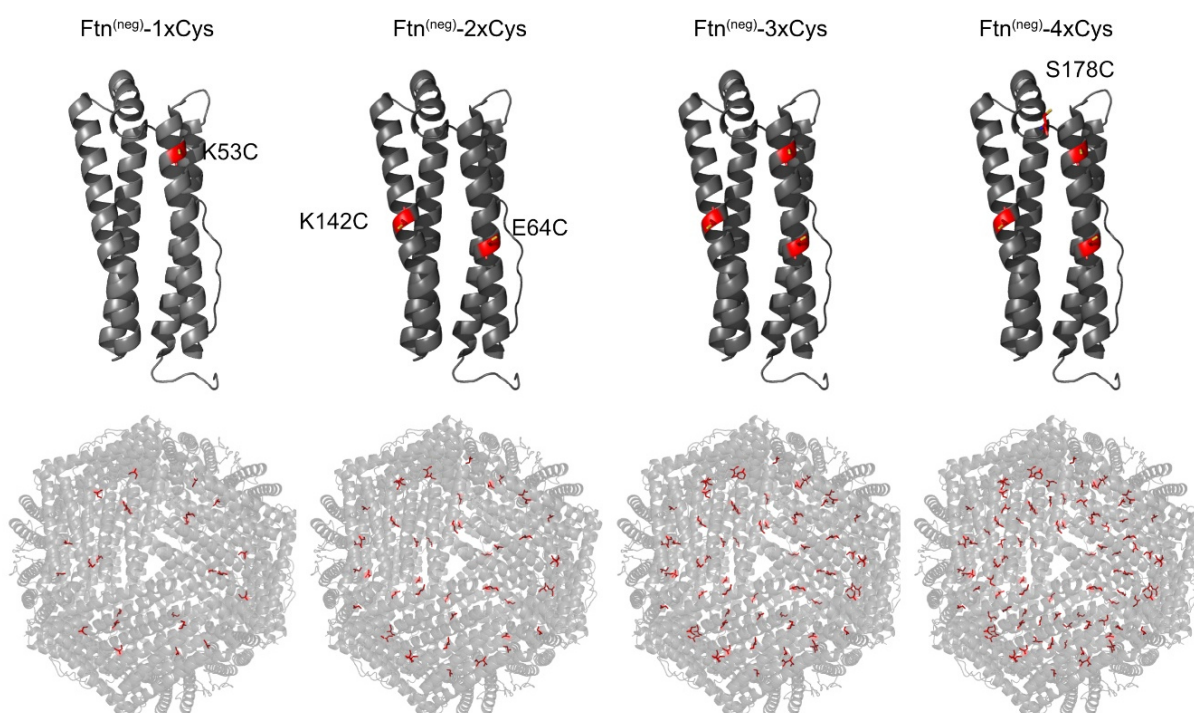


Figure 5.1: Schematic representation of the cysteine anchor sites. Ferritin subunit of the respective Ftn^(neg) cysteine variant with the mutated position highlighted in red. The location of the anchor sites in the fully assembled protein cage is shown in the lower panel.

The SASA of the finally selected positions is listed in Table 10.2. The distances between the possible anchor sites are an additional design criterion since already occupied sites could hinder the reaction at a neighboring site. The distance between the selected anchor sites is at least 8.8 Å as shown in Figure 10.1. The introduction of the cysteine amino acids in the genome of Ftn^(neg) as well as expression and purification of the respective ferritin variants are described in the following chapter.

5.1.2 Mutagenesis, production and purification of ferritin cysteine variants

QuikChange PCR methods were used for the introduction of desired mutations. The complete protocol can be found in section 8.4.1 and the sequences of the used primers is summarized

in Table 8.3. Most part of the mutagenesis was done by *Made Budiarta*. The resulting gene constructs were transformed in *E. coli* production strains using heat-shock techniques. The proteins were produced and purified using an already established protocol for Ftn^(neg) summarized in section 8.4.2.

Chromatography methods for protein purification allow first characterizations of the protein variants. Ion-exchange chromatography (IEC) is based on the surface charge of the protein, which has been bound to the oppositely charged column material and was eluted by an increasing salt gradient. Since all introduced mutations were located at the inner surface, no change in the elution behavior of the mutated variants was expected. Indeed, no significant changes to the conductivity during protein elution were detected as shown in Table 10.3. The full chromatograms can be found in the appendix in Figure 10.2. The elution volume during SEC chromatography was identical for cysteine-containing variants and the unmodified variants, proving a complete assembly of the container as seen in Figure 5.2.

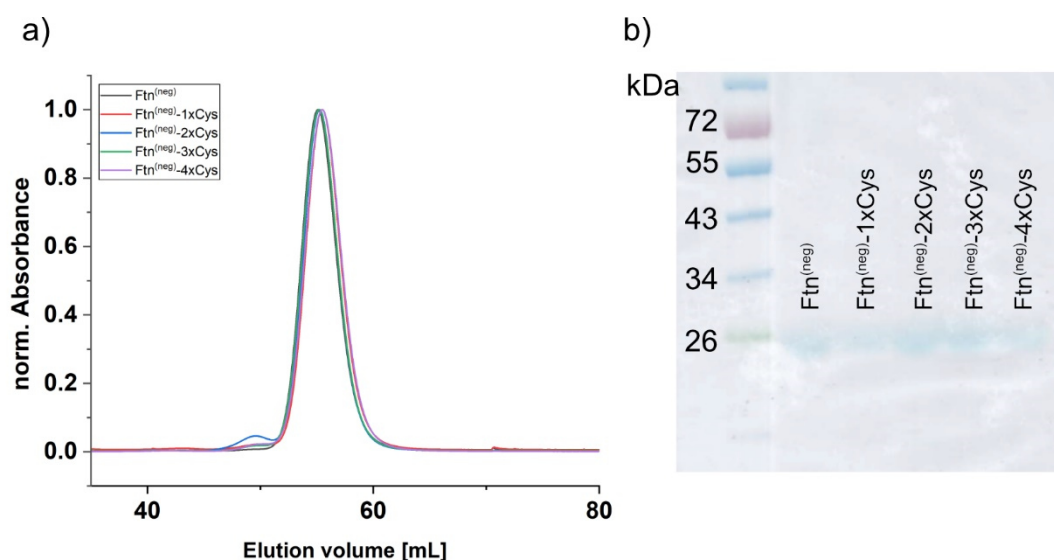


Figure 5.2: SEC chromatogram and SDS-PAGE for Ftn^(neg)-Cys variants. a) SEC chromatogram for ferritin variants with increasing number of introduced cysteine amino acids. b) SDS-PAGE of eluted SEC fractions proved the presence of ferritin.

The presence of the mutations was verified by comparing the molecular mass measured by ESI-MS to theoretical values as summarized in Table 5.1. The results validate the presence of the desired mutations. Corresponding ESI-MS spectra can be found in the appendix in Figure 10.3.

Table 5.1: Theoretical and measured molecular mass of Ftn^(neg)-Cys variants. The molecular mass of the variants as determined by ESI-MS fits well with theoretical values for the given amino acid sequence proving the presence of the desired mutations.

Protein	MW _{Theo.} ^[a]	MW _{Meas.}
	[kDa]	[kDa]
Ftn ^(neg)	21.196	21.196
Ftn ^(neg) -1xCys	21.138	21.139
Ftn ^(neg) -2xCys	21.114	21.113
Ftn ^(neg) -3xCys	21.088	21.088
Ftn ^(neg) -4xCys	21.102	21.102

[a] Determined with ProtParam tool^[242]

Ferritin variants with up to 4 cysteines per subunit exclusively located at the inner surface could be fabricated. Elution during SEC indicated the formation of intact protein cages. The presence of the mutations was proven by ESI-MS. In the following section, reaction conditions to functionalize these reactive sites with hydrophobic agents are investigated

5.1.3 Functionalization with hydrophobic agents

The investigated PBUTs possess a combination of hydrophobic and polar (negatively charged acid moiety) properties. Increasing the hydrophobicity of the inner cavity by conjugation of hydrophobic molecules could be a viable strategy to increase PBUT adsorption. Solubility is a limiting factor for efficient functionalization since both reaction partners need to be present in the same phase, which becomes challenging with increasing hydrophobicity of the ligand. To demonstrate the encapsulation of mainly hydrophobic species, α -halocarbonyl derivatives bearing a long aliphatic chain (2-bromo-N-decylacetamide) or a phenyl ring (2-iodo-N-phenylacetamide) were used. Chemical structures of these agents are shown in Figure 5.3. The functionalization agents were insoluble in water but could be dissolved in ethanol. The solution was stable for at least 24 h in mixtures with up to 50% ethanol. Complementary experiments with the protein cages proved its stability for at least 24 h in up to 80% ethanol. These results indicate that both reaction partners can be dissolved together in a mixture of ethanol and water. For initial coupling experiments, the Ftn^(neg)-1xCys variant was selected. 2-Iodo-N-phenylacetamid (Phe) was selected as a ligand, because of its higher reactivity due to the more favorable leaving group (see section 2.4.2). To increase the accessibility of the reactive thiol groups, the cage was disassembled into its subunit under acidic conditions.



Figure 5.3: Chemical structure of hydrophobic ligands. The chemical structure of hydrophobic ligands, which were coupled to cysteine residues to increase the hydrophobicity of the ferritin cage.

Since the coupling reaction only proceeds at a pH value between 7 and 8 as mentioned in chapter 2.4.2, the pH was increased by dilution with an appropriate buffer. The solution was concentrated achieving a higher reactant concentration for efficient conjugation. Next, the aqueous solution was mixed with equal amounts of ethanol. 20 equivalents Phe ligand with respect to the cysteine groups were added. The reaction mixture was stirred in the dark for 1 h, followed by dilution with an aqueous buffer to allow reassembly of the protein cage. The solution was kept at rest overnight, followed by purification through SEC. Elution volume in the respective chromatogram shown in Figure 5.4a indicates fully assembled protein cages. In pure Ftn^(neg), the ratio of protein UV absorption at 260 nm to absorption at 280 nm is expected to be around 0.5. After conjugation it was increased to 0.68. This shift could be caused by the additional UV absorption from the coupled Phenyl group. To determine the number of functionalized thiol groups, a second coupling step with a dye molecule linked to a maleimide group was performed. The ferritin sample was disassembled again and reacted with the dye Alexa Fluor 488 (AF488) using an protocol established by *Made Budiarta*.^[164] The full protocol as well as the calculation made to determine the amount of dye molecules per cage is described in section 8.5.3. The UV/Vis spectrum of a variant already functionalized with Phe and an unmodified Ftn^(neg)-1xCys after treatment with the dye are shown in Figure 5.4b. The UV/Vis spectrum of the unmodified cage after dye conjugation showed a strong absorption signal at 500 nm. The ratio of dye molecule per cage could be determined to be nearly 24 to one indicating conjugation of all reactive groups to the dye. For the Ftn^(neg)-1xPhe variant the number of dye molecules per cage was nearly zero, proving that all cysteine groups were occupied by the hydrophobic molecules. The downside of this method is the high cost of AF488 together with the time, work and sample demand. Additionally, experiments for variants with more than one cysteine revealed that not all positions could be functionalized with the dye. A more suitable characterization method is needed for further experiments. However, these first results showed that the general outline of the protocol using ethanol-water mixtures as a reaction medium is a promising starting point.

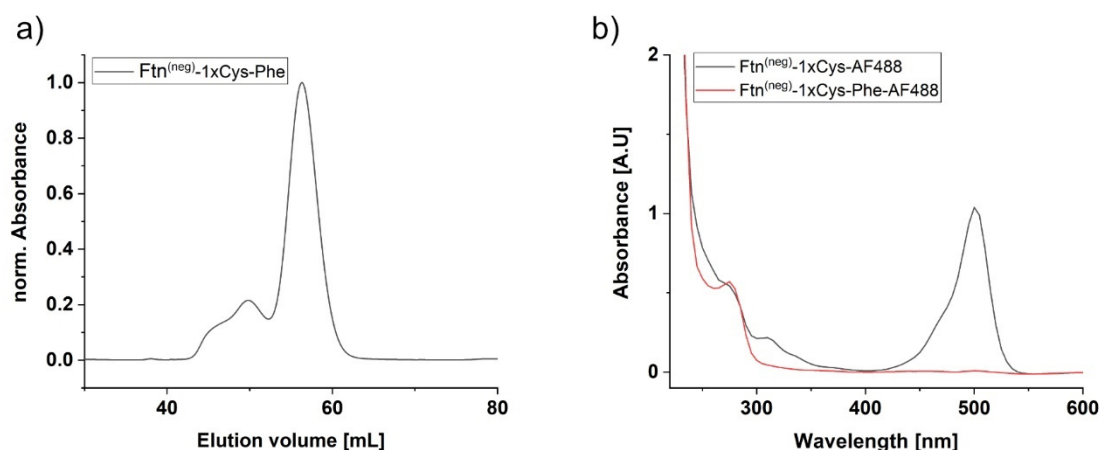


Figure 5.4: Functionalization of Ftn^(neg)-1xCys with Phe-ligand. a) SEC results for Ftn^(neg)-1xCys after coupling reaction with the Phe molecule. b) Comparison of UV/Vis spectra of Ftn^(neg)-1xCys conjugated to AF488 dye for untreated variant and variant already functionalized with Phe.

For functionalization of Ftn^(neg)-3xCys and 4xCys, the protocol was expanded by the use of a reducing agent (Tris(2-carboxyethyl)phosphine, TCEP), to keep the thiol groups reduced and reactive. Tests with 10 and 20 equivalents of TCEP with respect to the thiol groups were performed for both variants. In both cases, the SEC showed elution at the expected volume indicating complete reassembly, as shown in Figure 10.4 in the appendix. Again, the ratio of absorbance at 260nm/280nm of the resulting proteins was increased indicating at least partial conjugation, as shown in Table 10.4. Higher values for the 4xCys variants indicated that more Phe groups were present. Treatment with 10 eq. TCEP led to higher values than treatment with 20 eq., suggesting a higher degree of functionalization. ESI mass spectrometry was performed to investigate the functionalization degree. In Figure 5.5a, the spectrum for the untreated Ftn^(neg)-3xCys is shown. Signals represent differently charged ions of the subunit. The mass of the subunit was calculated from its m/z value as described in section 8.3.2 and showed the expected mass of 21.088 kDa. Similarly, the mass of the subunit in Figure 5.5b could be calculated to be 21.487 kDa, which is exactly the mass of the subunit (21.088 kDa) plus three times the mass of the functionalization agent minus three times the iodide leaving group (134.160 Da) minus the mass of three leaving hydrogen atoms (1.007 Da). Since no other signals were present in this spectrum, it is evident that under these conditions all thiol groups are functionalized with the respective ligand. On the other hand, the spectrum of the reaction performed in the presence of 20 eq. TCEP (Figure 5.5c) showed a strong signal with an m/z value corresponding to a mass of 21.355 kDa fitting to a subunit bearing only two coupled Phe molecules. Additionally, three less intense signals near the main signal could be observed, which can all be attributed to either the non-functionalized subunit or subunits with one or three coupled groups. The same pattern could be observed for the Ftn^(neg)-4xCys variant

shown in Figure 10.5 in the appendix. The ESI-MS results fit well to the trend observed in the absorbance ratios shown in Table 10.4. Apparently, the higher concentration of TCEP in the reaction mixture was not beneficial for the reaction. A possible explanation could be that the high amount of TCEP displaces or blocks the functionalization agent.

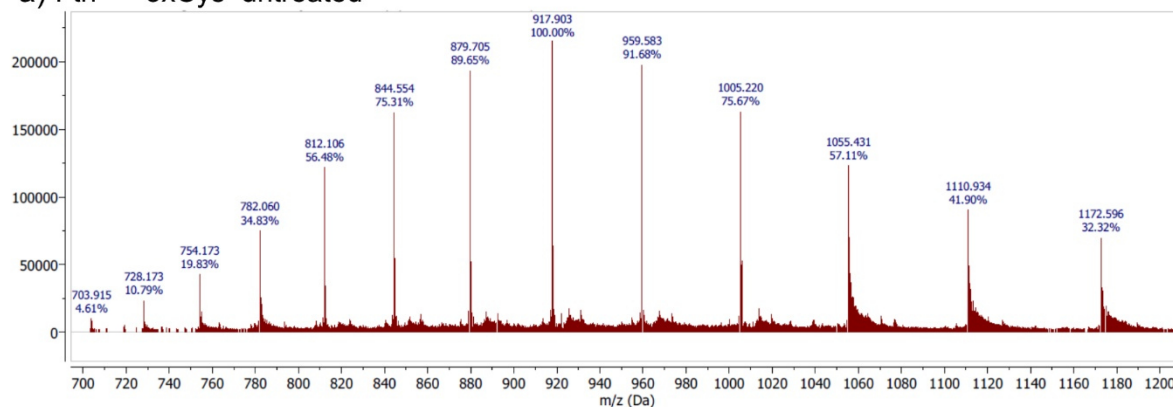
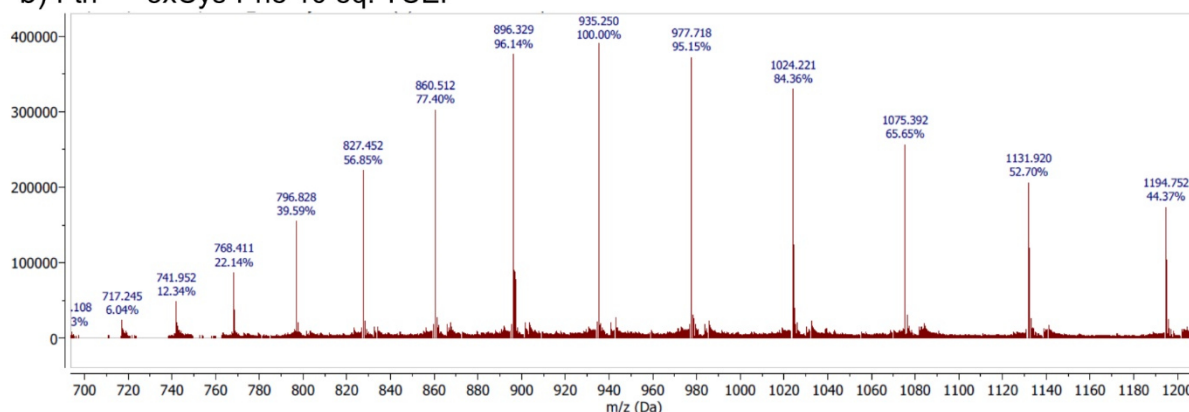
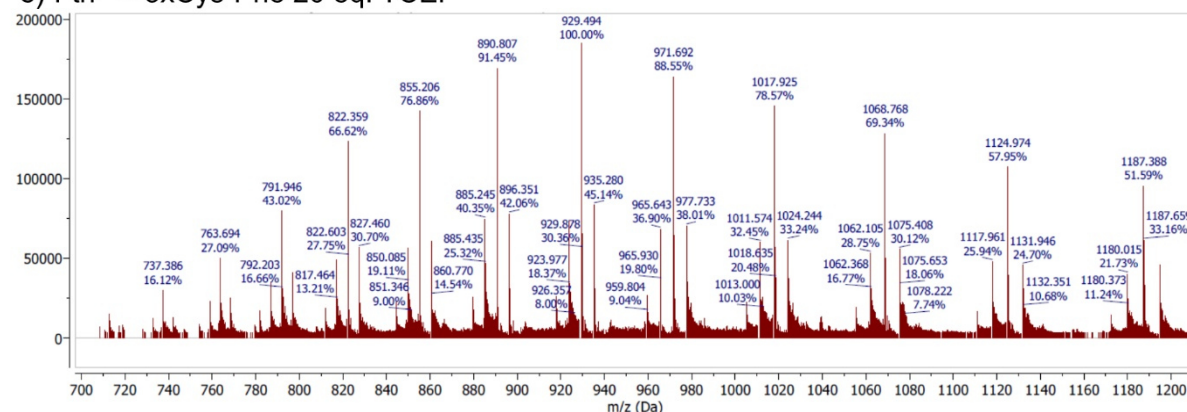
a) Ftn^(neg)-3xCys untreatedb) Ftn^(neg)-3xCys-Phe 10 eq. TCEPc) Ftn^(neg)-3xCys-Phe 20 eq. TCEP

Figure 5.5: ESI-MS spectra of functionalized Ftn^(neg)-3xCys variants. ESI-MS spectra for a) untreated Ftn^(neg)-3xCys, b) Ftn^(neg)-3xCys coupled to Phe at the presence of 10 eq. and c) Ftn^(neg)-3xCys coupled to Phe at the presence of 20 eq.

Applying the established protocol to the 2-bromo-N-decylacetamide (C10) yielded only a mixture of incomplete functionalized subunits, as observed in the respective ESI-MS spectrum in the appendix in Figure 10.6b. However, the respective SEC chromatogram in Figure 10.6a showed elution at the expected volume without larger aggregate peaks, indicating that even the sterically more demanding ligand did not interfere with cage assembly. The incomplete functionalization was most likely due to the lower reactivity of the bromo-acetamide derivate or its lower solubility. To increase its solubility, the ethanol content in the reaction mixture was increased to 80%. Again, the resulting SEC chromatogram revealed complete assembled container, suggesting that the protein cage tolerates this condition. However, incomplete functionalization could be observed in ESI-MS spectrum, shown in Figure 10.7 in the appendix. To account for the lower reactivity of the derivate, its concentration in the reaction mixture was doubled. The adapted reaction yielded a fully assembled container with all cysteine residues functionalized with the C10 ligand as observed in the respective chromatogram and ESI-MS spectrum in Figure 10.8. The developed protocol could also be applied to the Ftn^(neg)-4xCys yielding completely functionalized and intact protein cages. In conclusion protocols could be developed capable of functionalizing all thiol groups present in the Ftn^(neg)-Cys variant with either phenylic or aliphatic hydrophobic ligands.

Characterization of the functionalized derivatives of Ftn^(neg)-3xCys and 4xCys were performed. Respective variants that have proven to be fully functionalized are in the following termed Ftn^(neg)-3xPhe or 3xC10, or Ftn^(neg)-4xPhe or 4xC10. A comparison of the elution volume in SEC, shown in Figure 5.6, revealed no deviation between functionalized and non-functionalized variants. Therefore, the cage architecture appears to be unchanged by the functionalization of the inner cavity. This is further verified by TEM-images in Figure 5.7 and DLS measurements in Table 5.3.

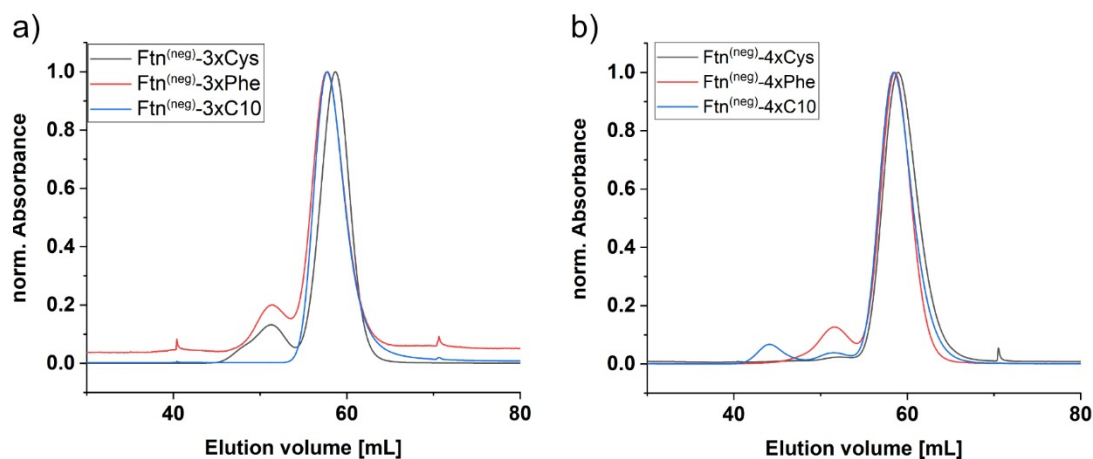


Figure 5.6: SEC chromatogram comparing the elution volume of functionalized and non-functionalized Ftn^(neg) cysteine variants. a) Comparisons for derivatives of Ftn^(neg)-3xCys. Neither functionalization with the phenylic nor the aliphatic ligand leads to a significant change in elution volume. b) Similar results are observed for derivatives of Ftn^(neg)-4xCys.

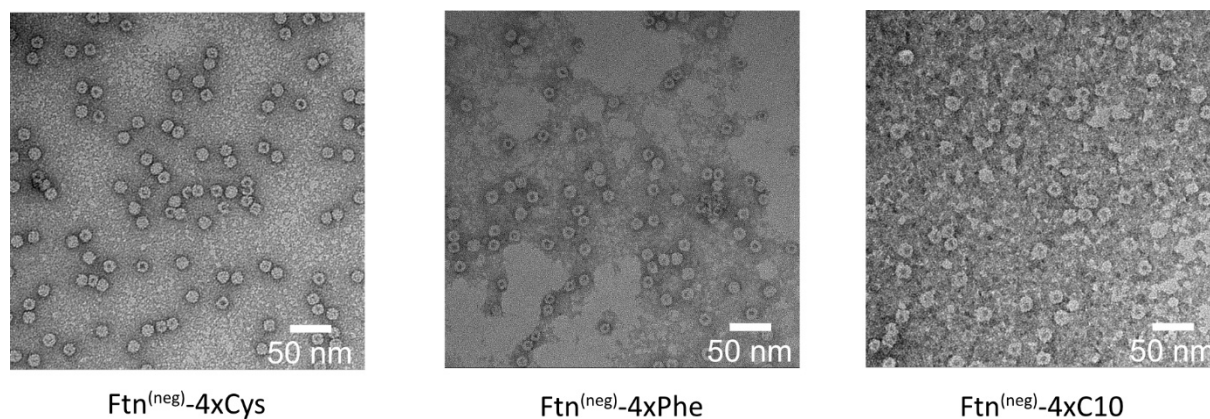


Figure 5.7: Negatively stained TEM images of functionalized and non-functionalized Ftn^(neg) cysteine variants. In all images, the characteristic cage like architecture of the ferritin is visible, which proves the intact cage geometry after coupling of the ligand molecules.

The complete loading of all variants containing 72 or 96 molecules of the respective ligand was proven by the presence of a single functionalized subunit in the ESI-MS spectrum with a mass perfectly fitting to the calculated mass of a subunit plus the mass of the respective amount of coupled molecules, shown in Table 5.2 and Table 5.3. The corresponding ESI-MS spectra can be found in the appendix in Figure 10.9 and Figure 10.10.

Table 5.2: Theoretical and measured molecular mass for Ftn^(neg)-3xCys derivatives. The theoretical and expected mass for a subunit linked to three times the respective hydrophobic molecule fit very well the masses determined by ESI-MS measurements, indicating a complete functionalization of the protein cage.

Protein	MW_{Theo.}^[a]	MW_{Meas.}
	[kDa]	[kDa]
Ftn ^(neg)	21.196	21.196
Ftn ^(neg) -3xCys	21.088	21.088
Ftn ^(neg) -3xPhe	21.487	21.487
Ftn ^(neg) -3xC10	21.680	21.680

[a] Determined with ProtParam tool^[242]

Table 5.3: Theoretical and measured molecular mass for Ftn^(neg)-4xCys derivatives and their diameter according to DLS measurements. The mass determined by ESI-MS measurements fits with the calculated masses for coupling of four times the respective molecule to a subunit, indicating complete functionalization of the inner protein cavity. The DLS measurements reveal almost identical diameters for the variants with a small increase for the loaded variants.

Protein	MW_{Theo.}^[a]	MW_{Meas.}	Diameter	PDI
	[kDa]	[kDa]	[nm]	
Ftn ^(neg)	21.196	21.196	13.75	0.010
Ftn ^(neg) -4xCys	21.102	21.102	13.80	0.130
Ftn ^(neg) -4xPhe	21.636	21.637	14.11	0.005
Ftn ^(neg) -4xC10	21.893	21.892	14.18	0.380

[a] Determined with ProtParam tool^[242]

Because native cysteine residues on the outer surface were removed, all modifications are limited to the inner cavity. It was assumed that changes to the inner cavity have no impact on the outer surface and the higher-order assembly of the protein cages. Therefore, hanging-drop crystallization under identical conditions used for unmodified Ftn^(neg) was performed (experimental details in section 8.6.1). Crystals with similar morphology were obtained as shown in Figure 5.8.

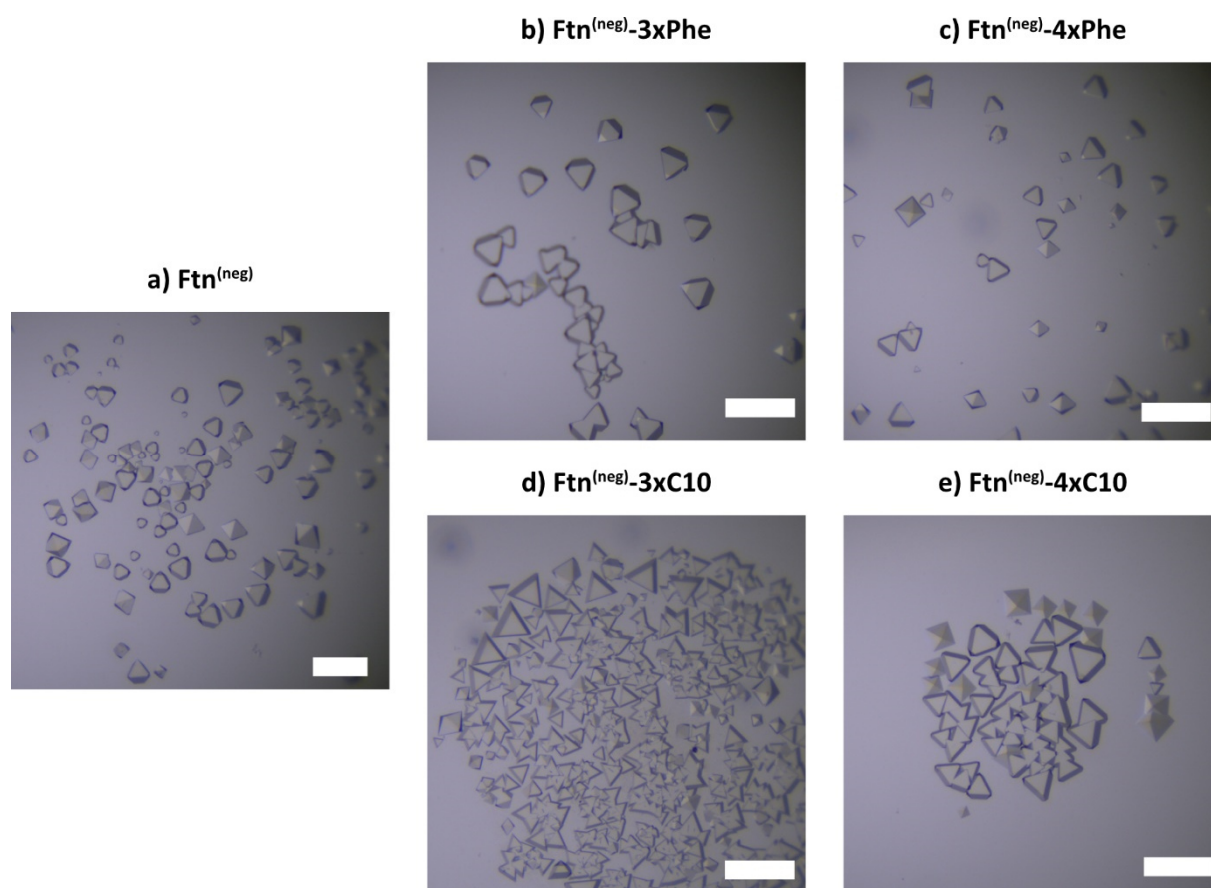


Figure 5.8: Crystals of functionalized Ftn^(neg)-Cys variants. Crystallization under conditions optimized for Ftn^(neg) yield crystals for (b-e) all Ftn^(neg)-Cys variants functionalized with Phe or C10 ligands with similar size and morphology as a) unmodified Ftn^(neg) crystals. Scalebar 400 μm .

Diffraction data for the respective crystals were collected at the *Deutsche Elektronen-Synchrotron DESY* and the crystal structures were determined. Omit maps were calculated to prove the presence of the ligand. The resulting images are shown in Figure 5.9. Compared to the unmodified variant, additional electron density was found at the positions (C53 and C64) of a Ftn^(neg)-3xC10 crystal, further indicating the presence of the ligand. The additional density is long enough for an aliphatic chain of 3 instead of 10 carbon atoms. The reason for that is most likely the high flexibility, allowing the ligand to adopt various configurations in the crystal, leading to a poorly defined electron density. Nevertheless, the crystal structure was extended by a shortened C10 ligand with only three carbon atoms. Crystallographic details on the data set are summarized in the appendix in Table 10.5. The $2F_o - F_c$ map (in blue) fits very well with the geometry of the ligand. The $F_o - F_c$ difference map (in green) generated for a model where the extension is missing showed unmodeled density and therefore proves the existence of a chemical entity at this position.

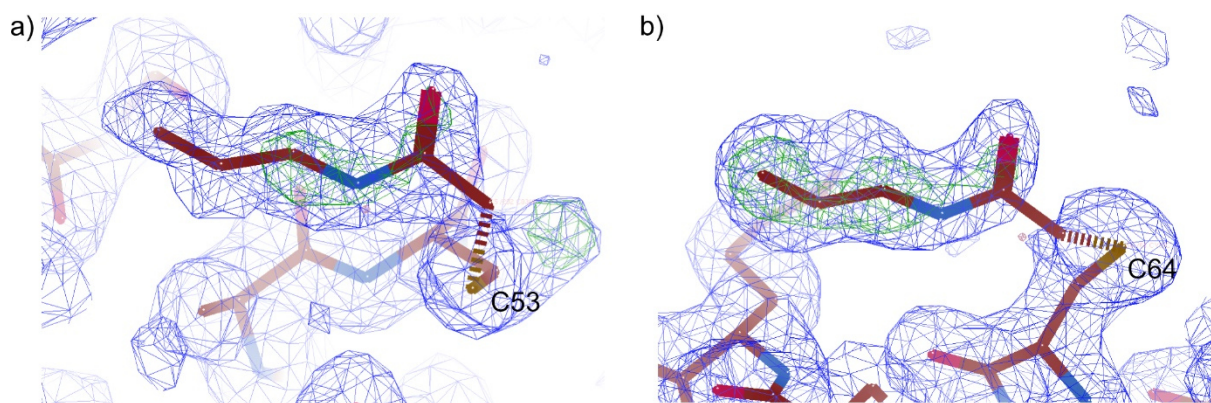


Figure 5.9: Electron density omit maps for Ftn^(neg)-3xC10**.** Electron density ($2F_o-F_c$ omit map, blue) and difference electron density (F_o-F_c , green) map for two functionalized cysteine residues with the aliphatic molecule. The electron density was only clearly visible at positions 53 and 64. $2F_o-F_c$ map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated without atoms of the coupled agents.

In summary, two building blocks for a protein-based hydrophobic adsorbent were fabricated. They were based on modified ferritin variants with up to 96 reactive thiol groups in the inner cavity allowing the modification of the inner surface with a high density of foreign molecules. Protocols for complete functionalization of the variants with hydrophobic agents were established and the loading was proven with ESI-MS techniques. The container architecture remains unchanged after functionalization as shown by size-exclusion chromatography, DLS measurements and negative stain TEM images. Since modifications are limited to the inner surface, established hanging-drop crystallization conditions yield similar results for the modified variants. In the next section, fabrication of a 3D material from these building blocks with the aim of creating a heterogeneous adsorbent are investigated.

5.1.4 Fabrication of heterogeneous adsorbent material

Hanging-drop crystallization yields low amount of crystals and is not suitable to produce the amounts of sample needed for the desired experiments. Therefore, a batch crystallization method suitable for production of several milligram protein crystals was applied. The difference between conventional hanging drop and batch crystallization can be explained using the phase diagram in Figure 5.10. The diagram is sharply divided by the solubility curve in a region of under- and supersaturation. The supersaturated region is further divided into the metastable, labile and precipitation regions. In hanging-drop crystallization, the protein is mixed with the precipitant solution on a cover slide and placed above a reservoir of the precipitant solution. The respective starting point in the phase diagram is located in the undersaturated phase. The concentration of protein and precipitant increases in the drop due to the diffusion of water. When the concentrations reach the labile region, formation of crystal nuclei occurs. Nuclei formation lowers the protein concentration shifting the system to the metastable region, where

new nuclei can no longer form, but existing nuclei can grow to protein crystals. In batch crystallization, the starting point needs to be located directly in the labile nucleation zone. Again, the decrease in protein concentration shifts the system into the metastable zone, where the crystals can grow till the undersaturated region is reached.

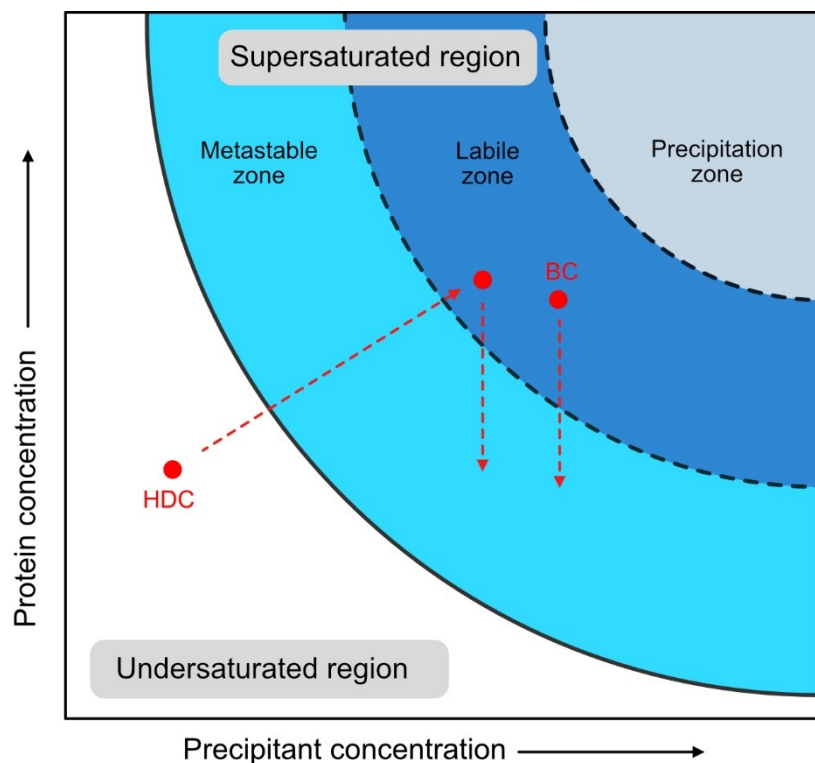


Figure 5.10: Phase diagram for crystallization of macromolecules. The diagram is sharply separated into the super- and undersaturated regions. Inside the supersaturated region a metastable, labile and precipitation zone exists. In the latter only precipitation occurs. In the labile zone, the formation of crystal nuclei can occur and in the metastable zone, nuclei can grow into crystals. The path for hanging drop (HDC) and batch (BC) crystallization techniques through the diagram is indicated in red. Adapted from reference 244 with permission of IUCr Journals, copyright 2014).

Initial conditions and a protocol based on already published methods for batch crystallization of $Ftn^{(neg)}$ were established in unpublished work of *Marcel Lach* and were used in this thesis. In short, the precipitant solution was added drop-wise to the protein solution under constant careful shaking. The mixture was kept at rest under ambient conditions. Crystal growth was stopped after 7 days by chemical fixation of the crystals and removing them from the protein-precipitant mixture (experimental details can be found in section 8.6.3). The size of the crystals could be tuned by changing the precipitant and protein concentration. This was done in initial experiments in small volumes of 20 μ L and the results are shown in Figure 5.11. From the experiments, it becomes evident that crystal size increased with an increasing precipitant concentration, in this case magnesium acetate (MgOAc). Furthermore, crystal size decreased

with increasing protein concentration, most likely due to the formation of more nuclei at the beginning. For later applications, a variability in crystal sizes can be advantageous because it allows further adaptation of the material. In general, a small crystal size should be beneficial for the overall adsorption capacity of the material since protein cages located in the center of the crystals can be easier reached by the toxins, which need to diffuse through the microporous crystal. On the other hand, for later application, the crystals need to be significantly larger than blood cells and other components of the blood to allow easy separation of the adsorbents.

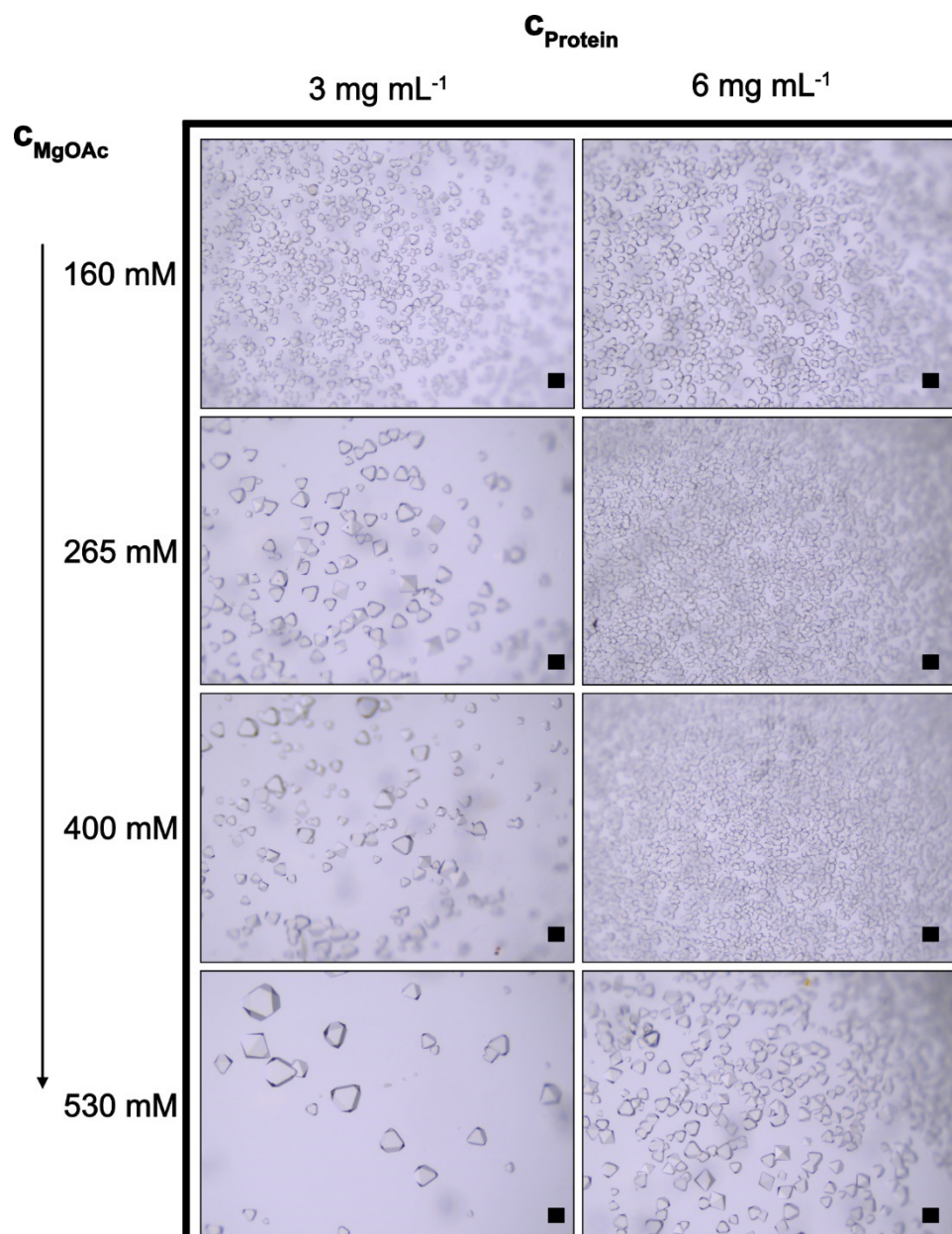


Figure 5.11: Crystal size at different precipitant and protein concentrations. Crystal size for $\text{Ftn}^{(\text{neg})}\text{-4xPhe}$ dependent on precipitate and protein concentration. The size of the crystals can be increased by increasing the concentration of the precipitant MgOAc. The crystal size can be decreased by increasing the protein concentration. (Scale bar 200 μm).

For the following experiments, the assembly condition with protein concentration of 3 mg mL^{-1} and 160 mM MgOAc yielding crystal sizes around $100 \text{ }\mu\text{m}$ was selected. To yield enough material for subsequent characterization methods, the batch crystallization procedure was upscaled, by increasing the volume of the reactants and keeping the concentrations constant. The volume of the total mixture was increased to 1 mL with a total protein mass of 3 mg . The resulting amount of crystals is shown in Figure 5.12.

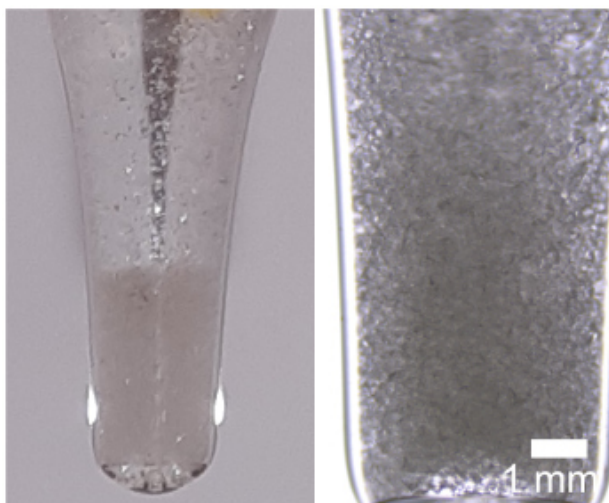


Figure 5.12: Protein crystals of Ftn^(neg)-4xPhe. Protein crystals were fabricated by upscaled batch crystallization technique with a total mass of 3 mg protein. Protein crystals were transferred into a glass vial and allowed to sediment.

Protein crystals are not stable when removed from the crystallization solution. To increase their stability, chemical cross-linking techniques were used. Crosslinking was performed by adding the cross-linker directly into the crystallization solution and incubating the crystals for a given amount of time (full experimental details can be found in section 8.6.5.2). The cross-linking was optimized for stability in a 60 mg mL^{-1} bovine serum albumin (BSA) solution with a pH of 7.4 and 0.13 M NaCl to mimic conditions in human blood. BSA was chosen since it possesses sufficient similarity to human serum albumin but is more affordable. In this thesis, two different crosslinkers were used, their chemical structure is illustrated in Figure 5.13.

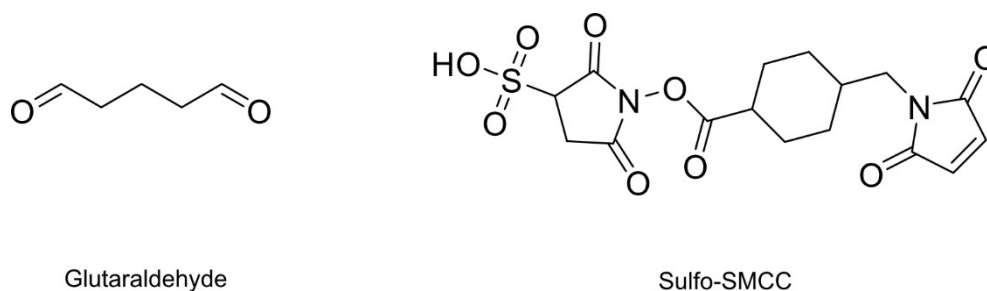


Figure 5.13: Chemical structure of applied crosslinker. Glutaraldehyde and Sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (Sulfo- SMCC) were used for chemical fixation of the protein crystals.

Cross-linked crystals were incubated in 60 mg mL^{-1} BSA solution and observed under a microscope. The results are shown in Figure 5.14. Without crosslinking, the crystals dissolved immediately after contact with the BSA solution as observed in Figure 5.14a. After incubation with glutaraldehyde for 4 h, the stability of the crystals significantly improves, but as observed in Figure 5.14b the crystals still dissolved after approximately one hour. The stability could be increased to the point until the crystals remain unharmed by the BSA solution by an additional crosslinking step with glutaraldehyde as seen in Figure 5.14c. The second incubation step is only performed for up to 10 minutes, because after longer incubation crystals showed yellow color, which is most likely due to the formation of polymeric glutaraldehyde species. In later experiments, it was observed that protein crystals after a second glutaraldehyde crosslinking step possess greatly reduced adsorption capacity in comparison to crystals just crosslinked once (Figure 10.13). This could be possibly explained by the formation of polymeric glutaraldehyde, blocking the entry of the crystal pores. Efforts to decrease glutaraldehyde concentration and incubation time during the second crosslinking step in order to find an optimal condition with high stability and adsorption capacity failed. Therefore, stabilization with another crosslinking agent, namely the sulfo-SMCC crosslinker, was established. This agent cannot react with itself, mitigating the risk of blocking the pores by forming polymeric species. The crosslinker has an NHS group capable of reacting with the amino groups of the protein. The other functional group is a maleimide group, which is usually used for the coupling of thiol groups. This reaction is only selective for thiols at pH values between 7 and 8. At a pH value of 8.5 present in the crystallization solution, reactions with amino groups are also possible. Crystals incubated once with sulfo-SMCC could withstand the BSA concentration for many hours as observed in Figure 5.14d and still showed measurable adsorption of PBUTs (Figure 10.13).

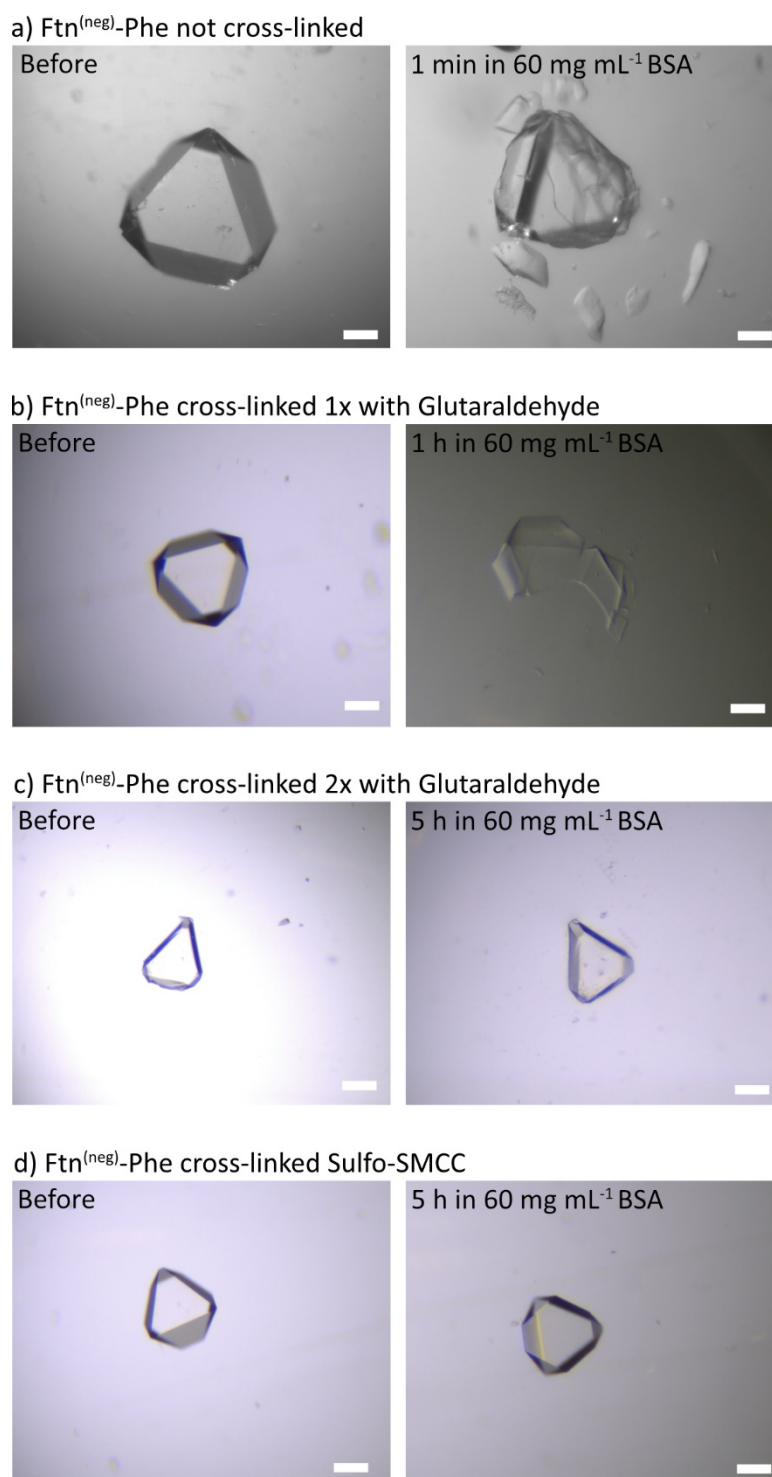


Figure 5.14: Stability of Ftn^(neg) crystals after different cross-linking procedures. a) Crystals without chemical fixation break and dissolve in a matter of minutes. b) Crosslinking with glutaraldehyde increased the stability, but after one hour the crystal is already heavily damaged. c) A second glutaraldehyde cross-linking yields crystals that can withstand the protein solution for multiple hours. d) Crosslinking with a Sulfo-SMCC cross-linker yields a crystal stable in the protein solution. Even after several days, the crystal shows no damage. Scale bar 200 μm .

In the crystalline material, solvent channels are present between the protein cages, allowing the PBUTs to diffuse in the material and reach cages located inside. Because of the highly

ordered character, these channels are evenly distributed. Their diameter is expected to be around 31 Å (Figure 10.11). It is possible that adsorption becomes limited by the diffusion through the material. To compare the crystalline adsorbent with a more open structure, a second method for the fabrication of heterogeneous material was explored. In chapter 2.4.2, it was shortly mentioned that glutaraldehyde was used to crosslink biocatalysts yielding a solid material. This technique was applied by incubating a protein cage solution with glutaraldehyde (experimental details can be found in section 8.6.4). After incubation for approximately 14 h, a white-yellowish protein precipitate formed, as depicted in Figure 5.15. If the incubation proceeds for too long, the color of the material changed more and more to yellow indicating the proceeding polymerization of the glutaraldehyde. The stability of the material was tested similar to the crystals in the BSA solution, showing no sizable mass loss after multiple hours in the BSA solution.

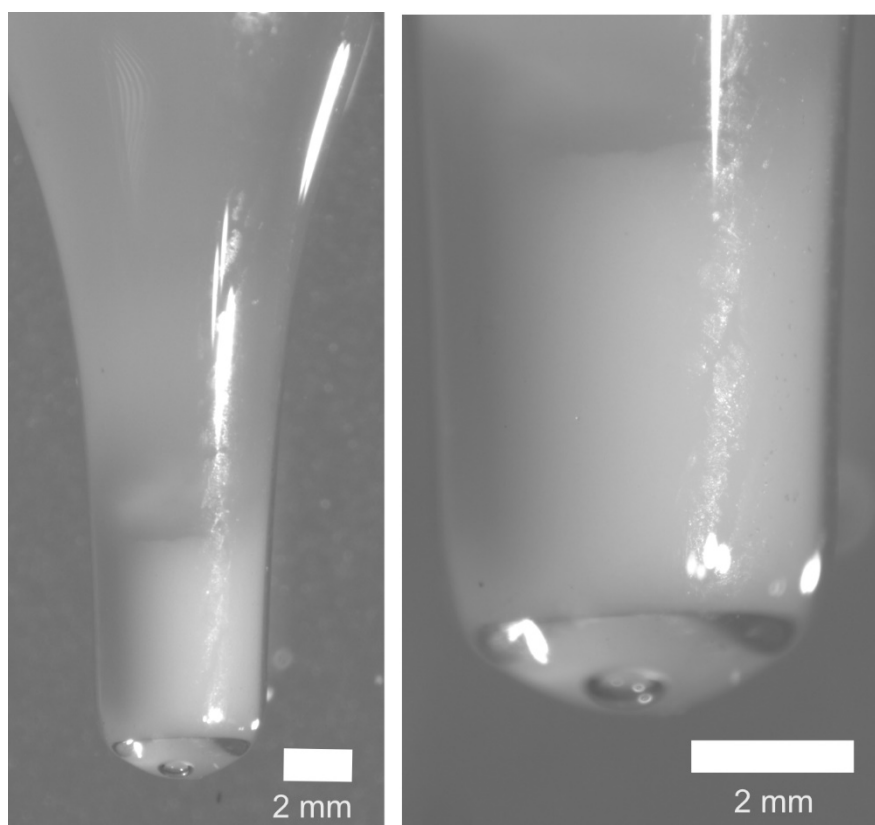


Figure 5.15: Non-crystalline protein material. Material is fabricated by mixing $\text{Ftn}^{(\text{neg})}$ -4xPhe with glutaraldehyde solution and incubation overnight. A white yellowish material is formed.

Precipitation was also tried with the sulfo-SMCC method, which also yields a material with a similar morphology. However, the material dissolved quickly in BSA and even in PBUT solution. The cross-linking was performed in an aqueous solution buffered with Tris(hydroxymethyl)aminomethane (Tris). The primary amino group of the buffer can also react

with the functional groups of the sulfo-SMCC. This unwanted reaction decreased the effective concentration of the cross-linking agent significantly. For the crystalline material the excess of cross-linking agent seems to be sufficient to still stabilize the material. However, for the non-crystalline material this is obviously not the case. The excess of crosslinker to cage was 120 to one for the sulfo-SMCC system and 1000 to one for the glutaraldehyde system. This difference probably explains why the precipitation with the glutaraldehyde in the Tris buffer still yield a stable material. For future experiments the cross-linking with sulfo-SMCC, but probably also with glutaraldehyde, should be conducted in amine free environments for example in phosphate buffer. However, the use of different buffer systems was not investigated in this work since the glutaraldehyde system yielded the desired material and did not show the same problems like diminished adsorption capacity or insufficient stability as observed for the crystalline material.

In summary, two routes to create a crystalline and non-crystalline insoluble material from the protein cages were established. One is based on batch crystallization followed by chemical fixation to increase the stability of the crystals. The other is based on an unordered crosslinking of the cages yielding a white precipitate-like material. It has to be mentioned that the characterization of these materials could not be completed in the scope of this work. Two crucial parameters for comparing the adsorbents with one another, namely the complete surface area and the pore size distribution, are needed. Efforts were made to determine these parameters for the crystalline material using gas adsorption techniques. 10 mg of protein crystals were fabricated using batch crystallization technique and dried under vacuum, but no surface area could be detected. It is possible that the sample amount was not enough. In the end stage of the vacuum drying, the surface of the liquid lies directly in the pores and channels of the container. Therefore, it is possible that due to the resulting capillary forces the pores collapse. This could be overcome by using more gentle drying methods like lyophilization. Due to the high sample amount needed for the analysis and the focus of this work on the inner cavity design, this experiment was not conducted yet. For future development of the material, these crucial parameters should be determined. Nevertheless, due to modular character of this system, techniques for efficient higher-order assemblies can be developed independently of the inner surface and can be combined.

To ensure that the material will induce no harmful side effects during its future application in human blood, biocompatibility assays are performed in the following section.

5.1.5 Biocompatibility

It was hypothesized that since the human heavy chain ferritin used as a scaffold is native to the human body, it possesses high biocompatibility. This is investigated in biocompatibility assays performed by *Setareh Orth-Alampour* from the *Jankowski group* at the *University Hospital Aachen*.

The protein cages were produced in *E. coli* bacteria. The outer cell membrane of these gram-negative bacteria contains lipopolysaccharides called endotoxins, which can cause a series of physiological reactions in the human body.^[243] During the decomposition of bacterial cells, these endotoxins are set free and need to be separated from the protein cages to prevent harmful side effects. To test if the protein-based sample contained any residual endotoxins, human aortic endothelial cells were incubated for 6 h with the respective crystals. The expression of human tumor necrosis factor-alpha (TNF- α) mRNA, which should be elevated in the presence of endotoxins,^[243] was monitored via quantitative polymerase chain reaction and compared to a positive control of 100 ng mL⁻¹ lipopolysaccharides (LPS). Experimental details can be found in chapter 8.7.1. The results for crystals crosslinked with glutaraldehyde and sulfo-SMCC are illustrated in Figure 5.16. The positive control (LPS) showed a significant increase in comparison to the negative control. On the other hand, all samples incubated with crystals showed no significant increase in TNF- α expression compared to the negative control. The results proved the absence of endotoxins contaminations in the material. The applied purification strategy was sufficient to remove potential harmful bacterial components.

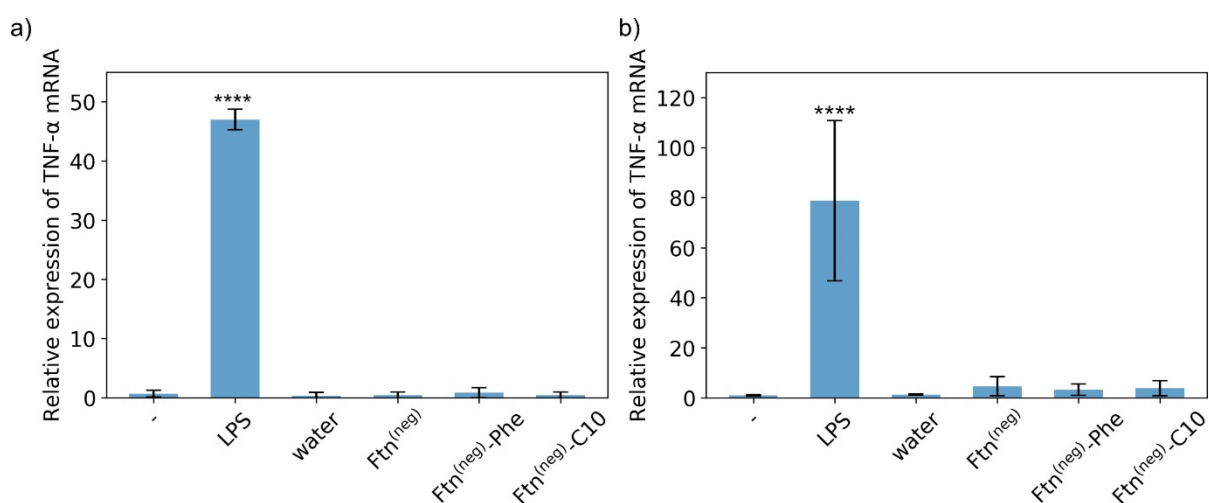


Figure 5.16: Endotoxin assay for the crystalline adsorbent. Relative expression of TNF- α mRNA for crystalline materials. The crystalline adsorbent is either crosslinked with glutaraldehyde a) or with sulfo-SMCC b). In both

cases, no significant expression of TNF- α is observed indicating no endotoxin contamination of the material. **** $P \leq 0.0001$ compared to negative control.

In later applications, the adsorbent will come into contact with human blood. Blood contact with materials unknown to the body can lead to blood coagulation, imposing dangerous side effects for the patient. Due to its human origin, it is likely that the protein cages surface does not induce blood coagulation. Importantly, the hydrophobic molecules are hidden inside the cavity. To verify that the protein-based adsorbent did not induce coagulation, the activation of platelets or thrombocytes upon incubation with the material was investigated. The platelet activation is regulated by the Ser-Thr kinase Akt.^[244] Experimental details can be found in section 8.7.2. In short, isolated platelets were incubated with a crystalline adsorbent or thrombin as a positive control. The platelet activation was measured by determining the ratio of phosphorylated to non-phosphorylated Akt¹ using Western blot techniques. The results for crystalline adsorbent crosslinked with either glutaraldehyde or sulfo-SMCC are illustrated in Figure 5.17. For both crosslinkers, no significant increase of the pAKT to AKT ratio was observed indicating that the material did not activate the blood platelets. Results for experiments conducted with pure ligands dissolved in ethanol are also depicted in Figure 5.17b. In the case of the C10 ligand, a slightly increased pAKT/AKT ratio was observed, but the finding is not very significant. Nevertheless, comparing it with the encapsulated C10 ligand, the mean value is nearly four times higher giving at least a small hint that encapsulation of the molecules in the cage improves their biocompatibility.

¹ The name AKT originated from the AKR mouse strain from which a transforming retrovirus encoding for an oncogene linked to “thymoma” (cancer originating from epithelial cells of the thymus) was isolated. Respective gene was called AKT and its gene product could be interpreted as a serine/threonine protein kinase. Later identified human a serine/threonine protein kinase were named accordingly.^[245]

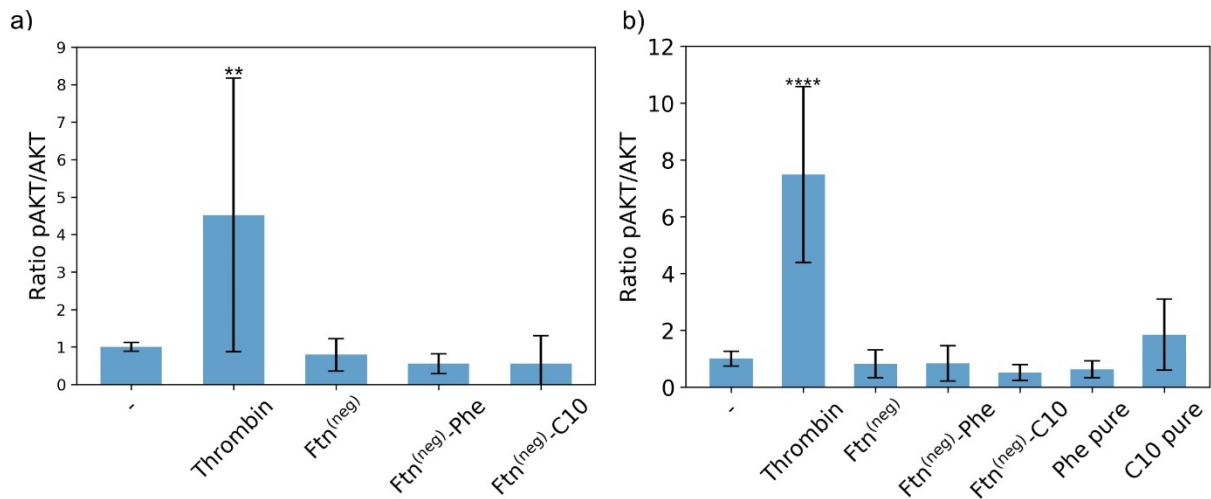


Figure 5.17: Platelet activation assay for crystalline protein-based adsorbent. Ratio of phosphorylated to unphosphorylated AKT as a measure for platelet activation. Platelets are incubated with crystals, which are a) crosslinked with glutaraldehyde or b) Sulfo-SMCC. ** $P \leq 0.01$, **** $P \leq 0.0001$ compared to negative control.

Similar tests for the presence of endotoxins or the induction of platelets were conducted for the non-crystalline material shown in Figure 5.18. Similar results were observed indicating that the higher-order assembly has no effect on the intrinsic biocompatibility of the protein cages.

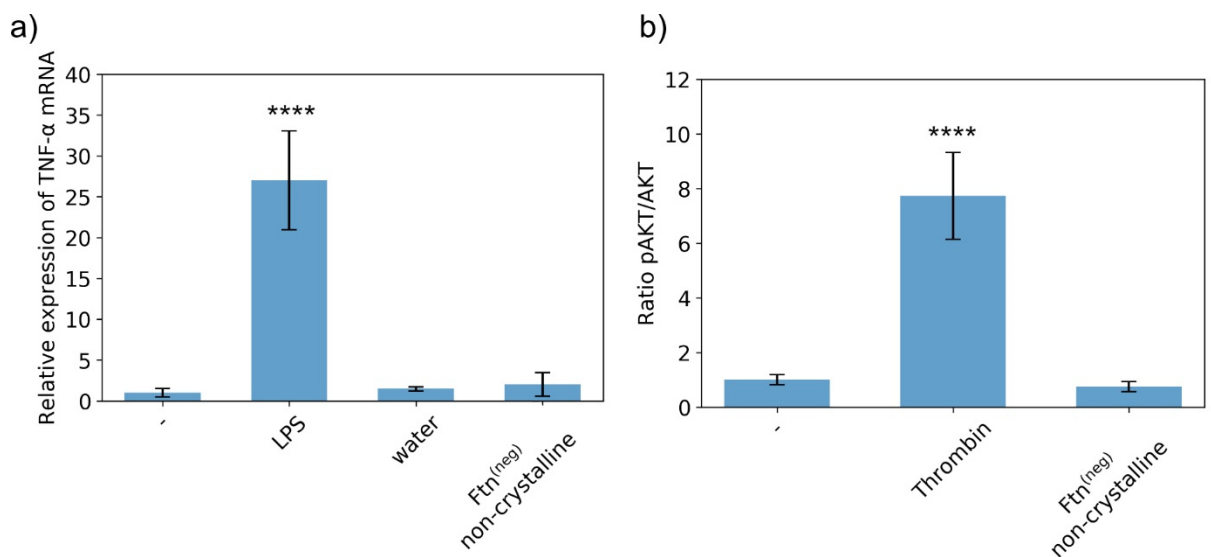


Figure 5.18: Biocompatibility of non-crystalline ferritin material. Endotoxin a) and Platelet activation b) assay for non-crystalline Ftn^(neg) material. No significant increase in TNF-α mRNA expression could be detected indicating no endotoxin contamination. No increase in pAKT/AKT ratio was observed indicating no activation of platelets. **** $P \leq 0.0001$ compared to negative control

Since it is aimed to contact the adsorbent with the whole blood, the adsorption of plasma proteins to the crystalline material was investigated. As a replacement for human serum albumin again BSA is used. Protein crystals chemically fixated with glutaraldehyde and sulfo-SMCC were incubated in a 60 mg mL^{-1} BSA solution. After 3 h, the BSA concentration was determined using a Bradford assay and compared to a control sample. Additionally, the concentration of a protein sample without BSA was determined to ensure that the Bradford assay is not falsified by dissolving protein-based material. More experimental details can be found in section 8.7.3. The results shown in Figure 5.19 revealed that the BSA concentration dropped by around 10%. Calculated on the protein mass around 300 to 600 mg BSA were adsorbed to each gram of adsorbent. Reduction of serum albumin levels of around 10% are also reported for other adsorbent materials and no harmful effects from BSA adsorption are reported, but BSA adsorption could block the pores of the adsorbent diminishing its capacity.^[17,50,245]

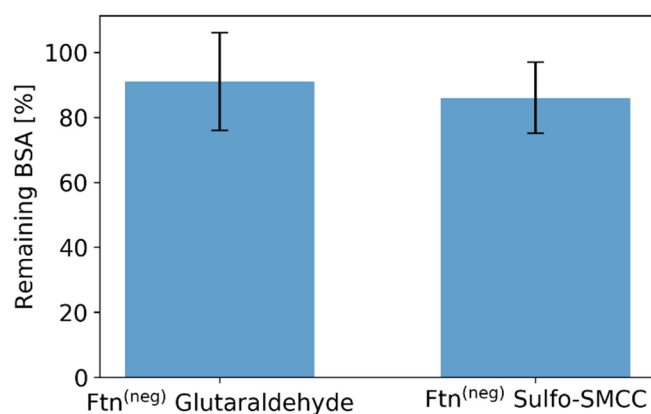


Figure 5.19: BSA adsorption assay. Percentage of remaining BSA in a 60 mg mL^{-1} solution after incubation with the crystalline protein-based material for 3 h. A decrease of around 10% BSA was found.

Initially it was expected that the negatively charged protein cage would show little interaction with the negatively charged BSA. However, the surface of proteins is heterogeneous and even in presence of an overall negative net charge, positively charged patches and groups are present. The binding of positively charged albumin patches to a negative charged patch of the Ftn^(neg) could explain the results. Additionally, binding events on the molecular level could be proceeding between functional groups of both proteins. Similar behavior was observed in molecular dynamics studies for the adsorption of BSA to negatively charged silica surfaces, where BSA formed a stable complex with the surface by binding via specific lysine residues.^[246] Other simulations towards BSA adsorption to surfaces also suggest that binding to carboxyl groups can lead to strong binding of BSA. This was compared to hydrophobic groups showing the weaker affinity and different geometries.^[247] Protein redesign could be a suitable tool to

decrease BSA adsorption, but understanding interactions between BSA and the ferritin cage is key for effective modification. *In silico* protein-protein docking could be a viable starting point for the investigation. Respective protocols executable in the Rosetta molecular modeling suite were established for docking of differently supercharged ferritin variants as a side project of this thesis (section 5.3). Here, these protocols could be applied to identify interacting side chains in both proteins. Respective amino acids at the cage could be either directly exchanged or buried by the introduction of amino acids with sterically demanding side chains in close vicinity. Here, the highly ordered crystalline material has the advantage that the orientation of the cages is always the same. This would allow designs where specific regions are pointed outwards, optimized to have a low affinity towards the BSA.

In summary, it could be proven that the protein-based material is biocompatible and possibly even hemocompatible. No residual amounts of endotoxins could be found, indicating that the applied purification methods are suitable for removing possible contaminations. Crystalline and non-crystalline materials showed no significant activation of blood platelets, indicating that the material is most likely not inducing blood coagulation. For a complete assessment of hemocompatibility for clinical applications, more specific experiments recommended by the ISO 10993-4 norm^[248] are needed. These embody monitoring complement factor 5a production, thrombin-antithrombin III complex level, thrombocyte concentration and possible other hematology parameters. Adsorption of BSA on the protein material remains an issue especially since it could block the entrance of the PBUTs to the protein cage. Further investigations as described above need to be conducted in order to address this problem. As a final characterization step the adsorption capacity of the protein-based material will be discussed in the next chapter.

5.1.6 The adsorption capacity of functionalized Ftn^(neg) variants

Characterization of the adsorption capacity is a critical part of this work and is needed to evaluate the impact of the modification done in the inner cavity. The adsorption of PBUTs to various modified variants will be compared by their respective adsorption capacity, which is defined as the quotient of mass of the adsorbed toxin and the overall mass of the protein material, as shown in equation (4).

$$\text{Adsorption capacity } [\mu\text{g g}^{-1}] = \frac{\text{Mass of adsorbed toxin } [\mu\text{g}]}{\text{Mass of adsorbent}[\text{g}]} \quad (4)$$

5.1.6.1 Establishing a protocol for determination of the PBUT adsorption capacity in protein-based materials

In general, the protein-based material was incubated in a PBUT solution with a pH value of 7.4 and 0.13 M NaCl. Afterwards the PBUT concentrations were determined and compared to a suitable control sample. As control sample, toxin solution was incubated without the addition of the protein adsorbent. Each sample was determined as a triplicate and the respective mean value was used for further calculation.

Early method development was focused on indoxyl sulfate. Toxin concentration was determined by an ultrahigh-performance liquid chromatography system (UHPLC) with a reversed-phase C18 column coupled to an electron-spray-ionization quadrupole-linear ion trap-mass spectrometer. The concentration of unknown samples was determined by comparing respective integrals with a calibration row. It was found that the integral of the peak derived from the HPLC-MS system was proportional to the PBUT concentration in a concentration regime up to $1 \mu\text{g mL}^{-1}$. A respective graph can be found in Figure 5.21a. For initial experiments, an IS concentration of $50 \mu\text{g mL}^{-1}$ was chosen. To determine the respective concentration, the solution needed to be diluted 100 times to reach a concentration located in the linear regime of the calibration row. Initial results revealed a high deviation between the three replicates of the same sample as visible in Figure 10.12 in the appendix. The high deviation could be tracked down to pipetting errors, caused by pipetting volumes of $1 \mu\text{L}$ of the sample. Increasing the minimal pipetted volume to $10 \mu\text{L}$ drastically reduces the standard deviation.

During the initial method validation, most control samples showed values significantly lower than the expected $50 \mu\text{g mL}^{-1}$, as seen in Figure 5.20a. A possible explanation could lie in insufficient sample stability under the conditions in the autosampler of the chromatography system. This was tested by monitoring the concentration of a freshly prepared stock solution over several hours. The results plotted in black in Figure 5.20b revealed a steady decrease in IS concentration over the course of 12 h.

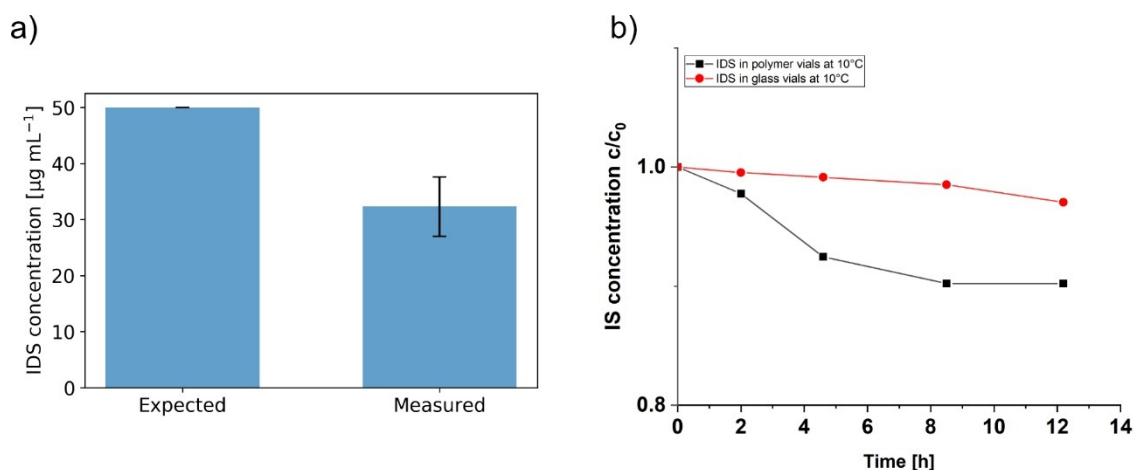


Figure 5.20: Decreasing IS control concentrations. Many initial experiments yield results with significantly lower control concentration as expected. a) comparison between expected concentration and measured concentration. b) Fraction of IS concentration in comparison to initial concentrations against time. The plot reveals a decrease of IS concentration when stored in polypropylene vials.

Based on this result, it was hypothesized that sample stability was the source of the problem. For further experiments, the samples were kept frozen and each sample was only added right before the measurement. However, control samples treated accordingly still showed significantly decreased concentrations, similar to Figure 5.20a. Another possible explanation for this observation was unspecific adsorption of the toxin to the walls of the storage vials. Up to this point, sample incubation as well as storage in the autosampler was done in polypropylene polymer vials. Freshly prepared IS stock solution were tested for its stability, but this time the samples were stored in a glass vial. Respective results, plotted in red in Figure 5.20b, show almost identical concentrations over the course of the whole experiment. The decrease in sample concentration observed in the earlier experiment (black line Figure 5.20b) was most likely not due to a decomposition of the toxin, but due to the adsorption of the IS to the walls of the polymeric vial. For further experiments, all samples were stored, incubated and measured in glass vials. Due to these adaptations, stable and reasonable control concentrations could be measured in all further experiments.

The intensity in mass spectrometry and therefore the respective integral can change gradually over time. To account for this during longer experiments, the calibration row was remeasured every 7 h. In this range, no significant changes between the calibration standards can be detected as observed in Figure 5.21a.

The IS concentration in later experiments was set to a level of 44 mg L^{-1} , which is a level expectable in end stage CKD patients.^[249] With the adapted protocol, this concentration was found for the used stock solution with only a minor deviation between three replicates (control

in Figure 5.21b). The same solution after 3 h incubation without protein adsorbent was remeasured as a negative control (negative control 3 h in Figure 5.21b). The concentration has fallen slightly but was still near the expected value of 44 mg L^{-1} . From the difference to concentration found in toxins samples treated with a protein-based adsorbent, the adsorbed toxin mass can be calculated. To ensure accurate results, the control sample was remeasured again before each sample.

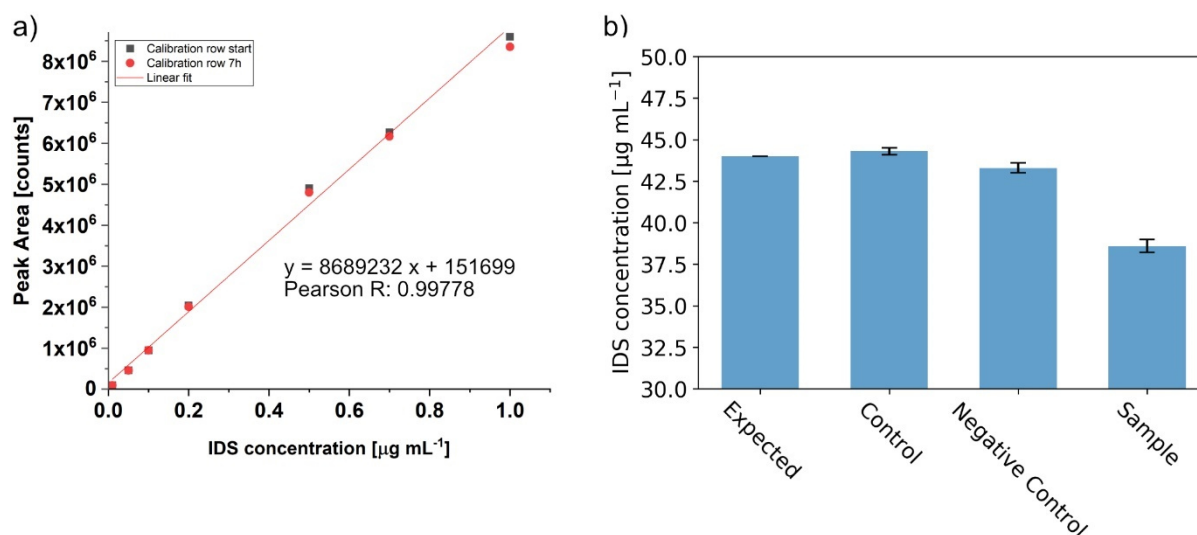


Figure 5.21: Results from toxin assays. a) Calibration row at the beginning of an experiment and after 7h. Values at the different times show no significant decrease. Linear regression was performed to calculate toxin concentration of unknown samples. b) Determined concentrations from toxin assays. The control showed the expected concentration with minimal deviation. The negative control showed nearly the same concentration with a minor decrease. Sample showed decreased concentration as a result of adsorption to protein material.

To determine the exact mass of the protein-based material, it was dried under vacuum until a constant mass was reached. The adsorption capacity was finally calculated by equation (4) from the mass of adsorbed toxin and the material mass. The method was originally developed for the determination of IS, but it was expanded to measurements of PAA and *pCS* without further obstacles. The adsorption of all toxins was investigated in concentrations expected in end-stage CKD patients and was set to 44 mg L^{-1} for IS,^[249] 41 mg L^{-1} for *pCS*,^[250] and 474 mg L^{-1} for PAA.^[33] Also, the determination of all three analytes in the same sample was possible since elution volumes sufficiently differ from one another.

In summary, an analytical method to determine the adsorption capacity of protein-based adsorbents was established. Complete experimental details are given in section 8.3.5. The protocol generated reliable values, but further improvements are advisable. The experiments were performed with a total adsorbent mass between 2 and 3 mg. Small changes to the mass

already have a great influence on the adsorption capacity and the used balance has an error of ± 0.1 mg. For future experiments, sample mass should be increased. Another drawback is the use of a standard calibration row and the repeated measurement of the control samples. Measuring time at the used devices is limited and this protocol does not use it in an efficient way. The application of internal standards like isotopically labeled toxins could help to tackle this problem. The adsorption capacity for the chemical-modified protein cages derived from this method is discussed in the next section.

5.1.6.2 Adsorption capacity of chemically modified protein cages

The impact of the different ligands on the overall adsorption capacity was determined. The crystalline adsorbent has a well-ordered structure and the resulting uniformity is beneficial for the comparability of the results. Crystalline material fabricated from Ftn^(neg) functionalized with the Phe and C10 ligand was investigated for their adsorption capacity towards the three PBUTs IS, *p*CS und PAA. Additionally, untreated Ferritin was used for comparison. Respective results are shown in Figure 5.22. Surprisingly, the unmodified protein container already showed adsorption capacity towards all investigated toxins. Respective adsorption capacities are ranging between 247, 283 and 2710 $\mu\text{g g}^{-1}$ for *p*CS, IS and PAA. The high capacity for PAA is most likely explained by the higher concentration of the toxin solution. In case of the *p*CS and IS, an increased adsorption capacity could be measured when the inner surface was modified with the Phe ligand. Respective adsorption capacities were 372 $\mu\text{g g}^{-1}$ for *p*CS and 458 $\mu\text{g g}^{-1}$ for IS. In contrast to that, the C10 ligand only showed a slight increase in the IS adsorption and no increase in *p*CS adsorption. Adsorption capacities for PAA were all in the same regime and it appears no real effect of the ligands could be observed. Despite a small trend towards higher adsorption capacity, the introduced hydrophobic ligands lead to no significant increase in adsorption capacity.

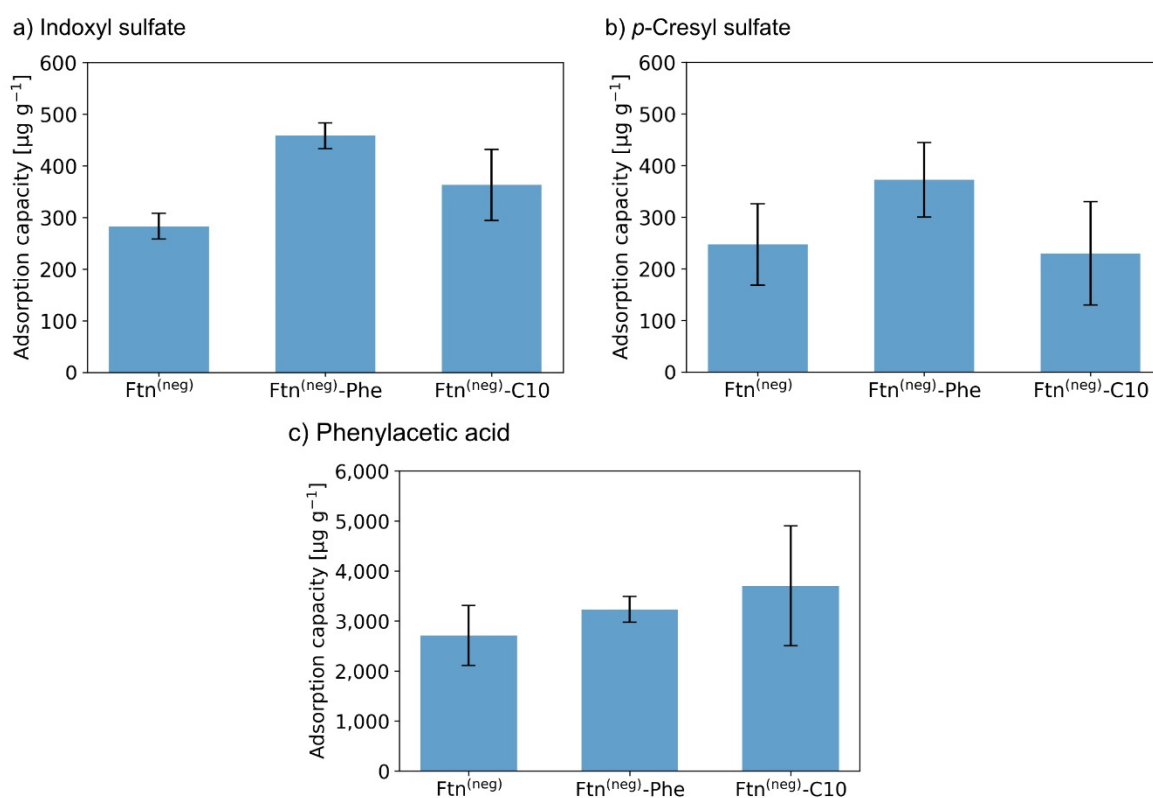


Figure 5.22: Adsorption capacity of crystalline material from chemical functionalized ferritin variants. The adsorption capacity of adsorbent derived from modified and unmodified ferritin variants towards a) IS, b) *p*CS and c) PAA. Respective chemical concentrations were 44 mg mL^{-1} for IS, 41 mg mL^{-1} for *p*CS and 474 mg mL^{-1} for PAA.

The results could be explained by the hydrophobic and hydrophilic properties of the PBUT. The used ligands only increase the hydrophobicity of the inner surface and the inner cavity is still negatively charged. So, no possibility for stabilization of the charged toxin groups is present. In future experiments, the adsorption could be improved by the use of ligands with a combination of positively charged and hydrophobic groups. The $\text{Ftn}^{(\text{neg})}\text{-Cys}$ system is designed for the variable exchange of the functionalization agent. The range of possible loaded molecules is mainly limited by the size of the agent and its solubility. Since already very hydrophobic molecules could be loaded, integration of similarly sized optimized ligands should be possible. In the following a few considerations for more efficient ligand systems are summarized. For ligands with an aromatic systems modulation of the π -systems electron density could be applied to increase the strength of the π - π interactions. The phenyl ring in *p*CS and PAA is electron rich. The electron density in the idol ring of the IS is even higher than in the phenylic ring due to the additional electrons from the nitrogen atom, which are integrated in the delocalized electron system. Additionally, the aromatic systems are linked to a sulfate or carboxylic acid group, which are deprotonated at physiological pH values. The resulting

+I effect is increasing the electron density even more. For effective π - π interactions the aromatic system of the ligand should be electron poor, which can be achieved by addition of electron withdrawing substituents for example trifluoromethyl or nitro groups. Respective aromatic systems could be combined with one or more primary to quarterly amino groups bearing a positive charge under physiological conditions to account for the amphiphilic character of the toxins. Depending on the geometry of the toxin the distance between the positive charged functional group and the aromatic system can be tuned by spacers consisting of one or more carbon atoms. Possible ligands are shown in Figure 10.14 in the appendix. In terms of aliphatic ligands addition of amino groups in the chain could led to an alteration of positively charged and hydrophobic motifs capable of binding the toxins. It should be also possible to increase the chain length. Assuming a diameter of the inner cavity of 6 nm^[146] an aliphatic chain could have a maximal length of 3 nm before contacting a similar ligand bound to the opposite surface of the protein cage. Chain length under ideal conditions were measured using the software *Avogardo* resulting in a total amount of atoms per chain of 22. However, due the curvaton of the cage the available volume per ligand decease closer to the center of the protein cage. As a result, ligands located at adjacent sites will sterically hinder each other if their size exceeds a certain threshold. Consequently, the theoretical maximal chain length of 22 atoms is most likely not practicable. Examples for possible ligands are shown in Figure 10.14 in the appendix. Additionally, the introduction of positively charged amino acids around the anchor sites could also help improve the performance of the ligand systems.

In further experiments, effects of the higher-order assembly of the cages were evaluated. For that the material obtained from the batch crystallization technique was compared to the material derived from the crosslinker precipitation method. The materials will be termed crystalline and non-crystalline in the following sections. The experiments were performed for unmodified and functionalized variants and compared in Figure 5.23 to the capacity determined in the prior experiment with the crystalline material. Again, only for the IS a significant improvement of the adsorption capacity could be found upon addition of the ligand.

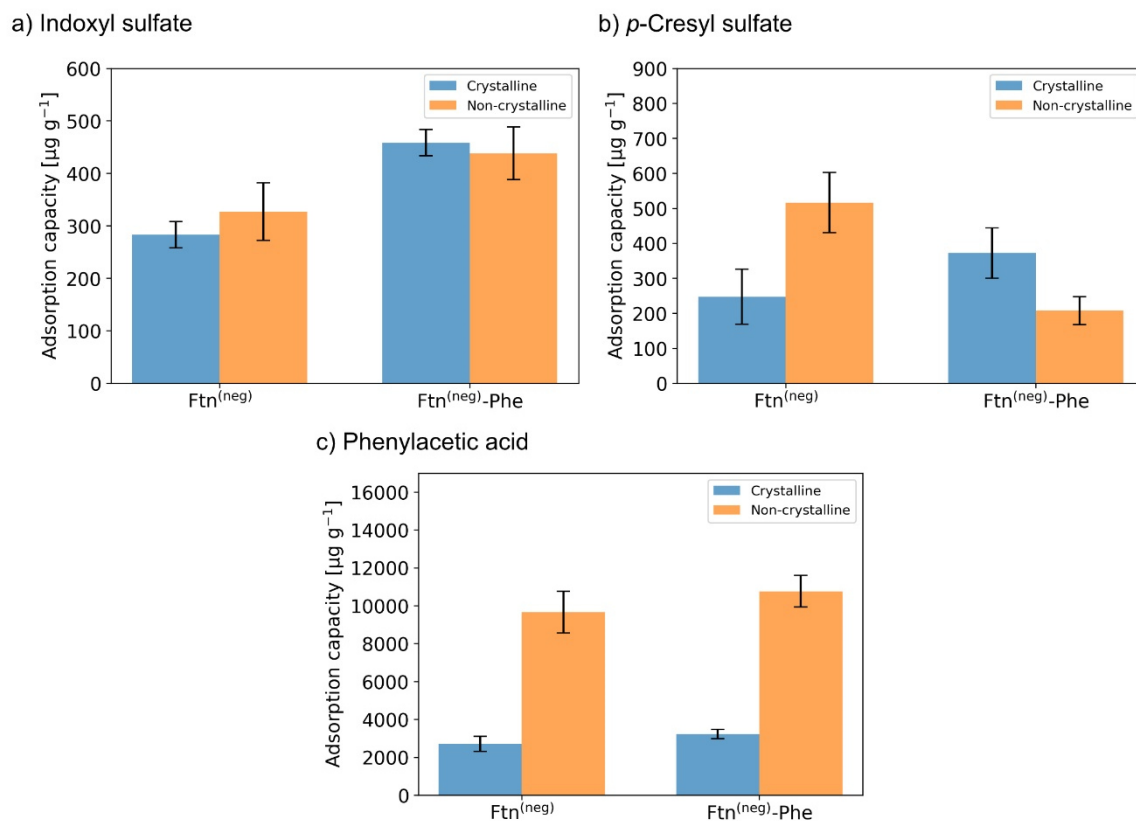


Figure 5.23: Comparison of adsorption capacity for the crystalline and non-crystalline adsorbents. The adsorption capacity for crystalline and non-crystalline adsorbent materials derived from non-functionalized and with Phe ligand functionalized ferritin variants. Values are determined for a) IS, b) *p*CS and c) PAA. Respective chemical concentrations were 44 mg mL^{-1} for IS, 41 mg mL^{-1} for *p*CS and 474 mg mL^{-1} for PAA.

In the case of the IS in Figure 5.23a, no significant difference between crystalline and non-crystalline material could be determined. For *p*CS adsorption (Figure 5.23b), non-crystalline materials derived from unmodified ferritin showed improved adsorption capacity. On the other hand, for the functionalized variant the opposite effect was observed. The adsorption of PAA appears to greatly benefit from the change in material morphology. Adsorption capacities for both variants showed a three to fourfold increase reaching values of nearly $10,000 \mu\text{g g}^{-1}$. These findings indicate the impact of the higher-order assembly on the overall adsorption capacity. But while experiments for PAA showed this quite clearly, the results derived from adsorption of IS suggest no difference between different material assemblies. Also, *p*CS adsorption results showed both trends. More data on this topic is available for redesigned protein cages (section 5.2.5) and the discussion about the possible trends and what causes them will be carried out in section 6.

In summary, PBUT adsorption capacities could be determined for the protein-based adsorption material. The ferritin cage without further modifications was already capable of adsorbing

PBUTs. First hints towards the importance of the higher-order assembly were found, but contradicting trends were observed. Modification with the investigated ligands Phe and C10 showed a small trend to higher adsorption capacity, but these are only minor and not significant with respect to the unmodified cages. As described before other ligands bearing positively charged and hydrophobic features could improve the binding capacity. The ferritin cysteine variants are variable in terms of possible cargo. This is demonstrated in the following section.

5.1.7 Encapsulation of fluorophore molecules for bioimaging and biosensing applications

Due to the variability of the ferritin-cysteine system, further applications depending on the selected cargo are possible. The incorporation of normal and pH-sensitive organic dyes opens routes toward applications in bioimaging and biosensing and are tested for their feasibility. After internalization (endocytosis) by the cell, colloidal drug carrier systems are in general located in endosomes/lysosomes and need to release their cargo in the cytosol to accomplish their therapeutic function.^[251] Since the pH varies between the cytosol and different intracellular organelles like the endosomes/lysosomes,^[252,253] pH sensing probes allow the monitoring of physiological changes inside cells and are crucial for drug carrier development. Dye-loaded ferritin is an attractive candidate for such a probe material due to its small size and biocompatibility. Especially for pH-sensitive dyes, encapsulation in the inner cavity represents a great advantage. Close to negatively charged surfaces like cell membranes or proteins, an altered pH value is present due to the attraction of H⁺ ions. Fluorophore molecules in close vicinity will be influenced by this local pH leading to misinterpretation of the experiment.^[254] Ferritin encapsulated dye would be not influenced by this effect, since the cage guarantees a defined uniform chemical environment around the dye.

The dye conjugation was first tested with the dyes Alexa Fluor 488 and Rhodamine 6G linked to a cysteine reactive maleimide group. The protein cage was disassembled under acidic conditions and upon restoring to neutral conditions the respective dye derivatives were added. The mixture was kept at rest to allow reassembly of the container. The derivatized proteins were purified with SEC. In the resulting chromatograms shown in Figure 5.24a and b, an additional absorbance at a dye-specific wavelength was observed.

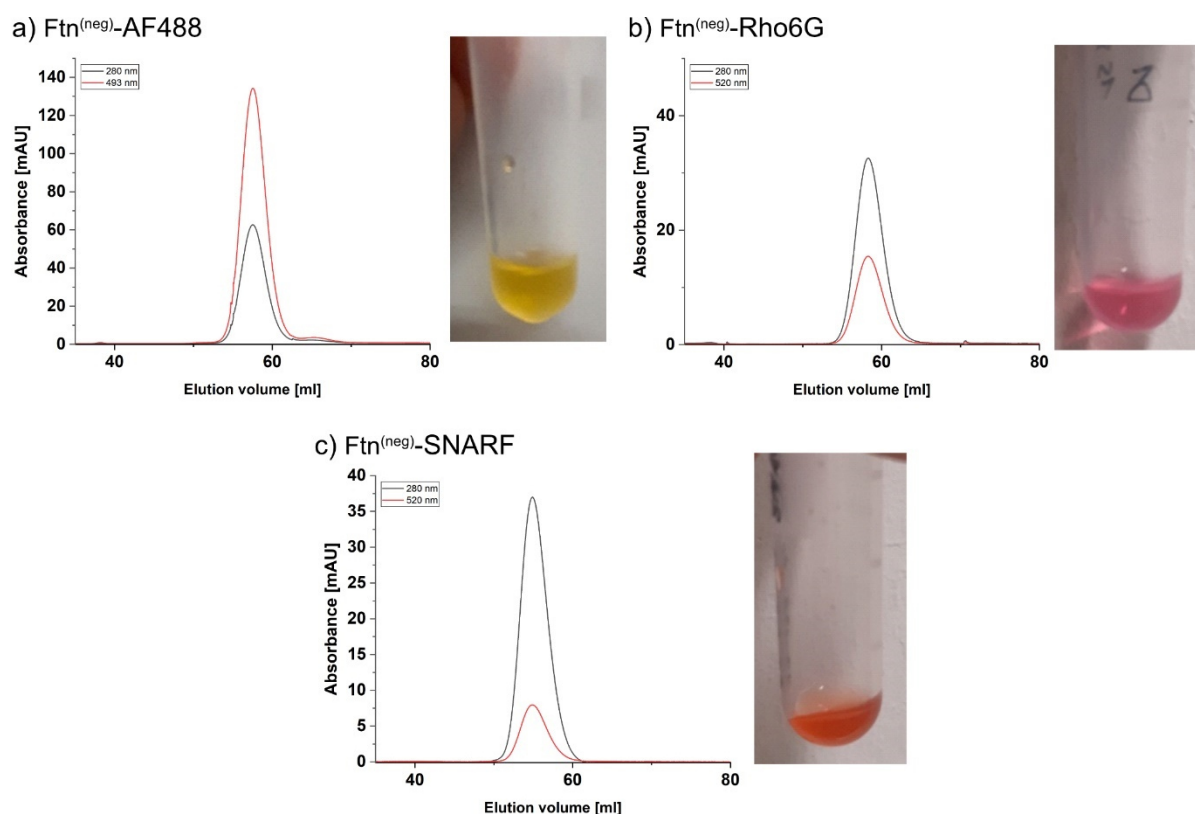


Figure 5.24: Fluorophore-loaded ferritin variants. Loaded $\text{Ftn}^{(\text{neg})}$ variants show additional absorbance at a dye-specific wavelength during SEC indicating successful encapsulation of the dye. After encapsulation a color change of the samples is visible.

In combination with the change in color of the resulting sample, the loading of the dyes could be verified. Experiments were repeated with the pH-sensitive seminaphtharhodafuor (SNARF) dye. Successful encapsulation could be verified by the additional absorbance during SEC (Figure 5.24c). The photoluminescence (PL) spectrum of the $\text{Ftn}^{(\text{neg})}$ -SNARF variant is shown in Figure 5.25a. A change in the region between 550 nm and 600 nm at different pH values was observed, indicating that the dye still responds to a pH change of the surrounding medium. The pH-dependent change will be used in future experiments to determine the pH value of the intercellular components. A pH calibration curve was prepared by determining the ratio of luminescence at 580 nm to 640 nm (I_y/I_r) at different pH values. The plotted results illustrated in Figure 5.25b can be fitted by a sigmoidal function with the inflection point between a pH of 9 and 10. Apparently, the dye was still able to react to pH changes. However, the inflection point was expected to reside between 6 and 8. This is potentially caused by the negative charged inner surface of the ferritin cage leading to adsorption of H^+ counter ions and a lower local pH value.

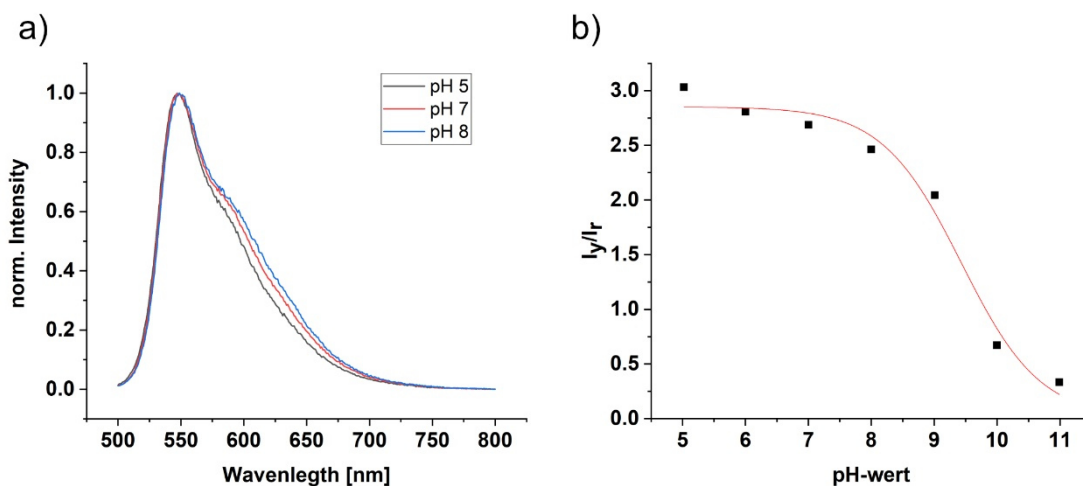


Figure 5.25: Ftn^(neg)-SNARF pH sensing capability. a) Photoluminescence spectrum of Ftn^(neg)-SNARF (ϵ_{ex} : 488 nm) values show deviation in a region between 550 and 600 nm at different pH values. b) Ratio yellow (580 nm) to red (640 nm) luminescence plotted against pH value fitted by a sigmoidal function.

Since the SNARF dye was coupled in the vicinity to the surface, it reacts to the lower local pH value and not to the pH value of the bulk solvent, leading to the shift in the calibration curve.^[254] The system right now is not suitable to detect the interesting changes between the cytosol (pH 7.4) and the endosomes/lysosomes (pH 4-6). Sensitivity in the desired pH regime could be facilitated by encapsulation of pH-sensitive dyes showing a drop in their I_y/I_r value at lower pH values, which is shifted to higher pH values in the ferritin cavity. Oregon Green, which I_y/I_r value drops between 4 to 6 in solution and close a negatively charged surface it is shifted to 6-8,^[255] is a possible candidate. Since the shift is caused by the potential of the surface and the dye position is precisely defined by the anchor site, exchanging charged amino acids around the anchor site could open a route to further fine-tune the sensitivity of the dye. Techniques for redesigning the ferritin's inner cavity are extensively investigated in the next chapter and could be also applied to this system.

5.2 Designing toxin binding Ftn^(neg) variants

The ferritin cage has evolved to facilitate the storage of iron ions. As a result, its inner cavity possesses a high density of negatively charged amino acids. The resulting negative surface potential could hinder the adsorption of the negatively charged PBUTs. Efforts were done to redesign the inner cavity to be more beneficial for the PBUTs adsorption. The applied design strategies were based on changing the polarity of the inner surface or designing PBUT binding sites. Protein models were generated *in silico* with the Rosetta molecular modeling suite using fixed-backbone and ligand docking protocols. Selected designs were produced and characterized in terms of yielding soluble assembled protein cages and their PBUT adsorption capacity.

5.2.1 Computational design

In the following section, different design strategies are discussed, and models based on these strategies are generated.

5.2.1.1 Increasing the density of amino acids with hydrophobic side-chains

Increasing the density of amino acids with hydrophobic site chains could be a viable strategy to increase the affinity to the partial hydrophobic PBUTs. Since mutations at critical positions could lead to improperly folded and insoluble proteins, possible mutation sites were identified using the fixed-backbone (fixbb) protocol of the Rosetta software suite. Details on applications, scripts and used protocols can be found in section 8.8.2 and 10.5.4. The fixed back-bone (fixbb) protocol can exchange amino acids at defined positions and determine respective Rosetta energy score. Mutations were limited to amino acids placed in the inner cavity. The list with all considered positions can be found in the appendix Table 10.1. The fixbb protocol was allowed to introduce amino acids with hydrophobic side chains at these positions. The full residue file can be found in the appendix in section 10.5.4.3. From the first results, 16 possible mutation sites were found. The total ferritin subunit is composed of 183 amino acids. Counting in the four mutations introduced for the surface supercharging, this would result in exchanging around 10% of the total protein. The number of variants that can be tested in *in vitro* experiments is limited. The design decision was made to test variants with increasing number of mutations, to gather information about the stability of the protein. The lowest number of introduced hydrophobic residues were chosen to be four, allowing removal of all negatively charged amino acids at solvent exposed positions. This variant was termed Ftn^(neg)-Ap4. Additional mutation sites were ranked by their accessible surface area and added incrementally to the protein. A variant with three more mutations termed Ftn^(neg)-Ap7 was designed. Finally, to test the limit of the ferritin backbone flexibility, a variant with all 16

introduced mutations (Ftn^(neg)-Ap16) was selected. The sequence of these variants aligned with the Ftn^(neg) sequence is shown in Figure 5.26. Used residue files can be found in the appendix in section 10.5.4.3. No significant difference in Rosetta scores of the hydrophobic variants and the Ftn^(neg) was found as seen in Table 5.4. Therefore, even Ftn^(neg)-Ap16 should yield soluble protein.

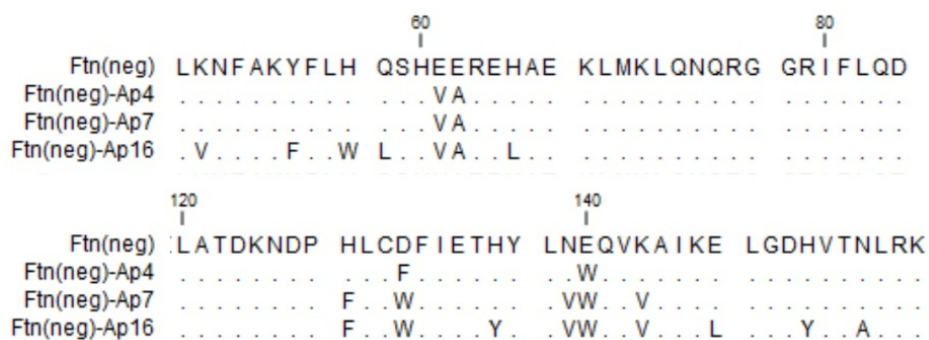


Figure 5.26: Sequence alignment for Ftn^(neg)-Ap variants. For the variant Ap4, mutations were selected to replace charged residues at the inner surface of Ftn^(neg). Variants Ap7 and Ap16 are expanded variants of Ap4 with additional three and twelve mutations at the most accessible positions.

Table 5.4 Rosetta score of Ftn^(neg) variants with additional hydrophobic residues. Best scoring model from 5000 trajectories with increased density of hydrophobic groups in the inner cavity.

Protein	Rosetta Score
Ftn ^(neg)	-1394,847
Ftn ^(neg) -Ap4	-1412,467
Ftn ^(neg) -Ap7	-1425,678
Ftn ^(neg) -Ap16	-1449,891

The expected surface charge of the variants was determined with the APBS tool^[256] in *PyMol* and shown in Figure 5.27. The results revealed a significant reduction of negative surface charge between the non-mutated Ftn^(neg) and the structure of Ftn^(neg)-Ap4 in the upper left region of the subunit. The decrease was expectable since the mutations were aimed to replace the charged amino acids in this region. The mutations were located near the 3-fold channel and the decrease in negative charge could improve uptake of the toxins. However, a significant area of negative charge remained in the cavity, visible on the right side of the protein subunits in Figure 5.27. These patches were caused by negatively charged amino acids located deeper

in the cavity. Mutations were omitted at these sites to maintain the integrity of the protein. The introduction of amino acids with hydrophobic sidechains in the Ap7 and Ap16 variants lead to slight reduction of the negatively charged patch, but this is most likely just because of the sterically more demanding groups shielding the negatively charged amino acids.

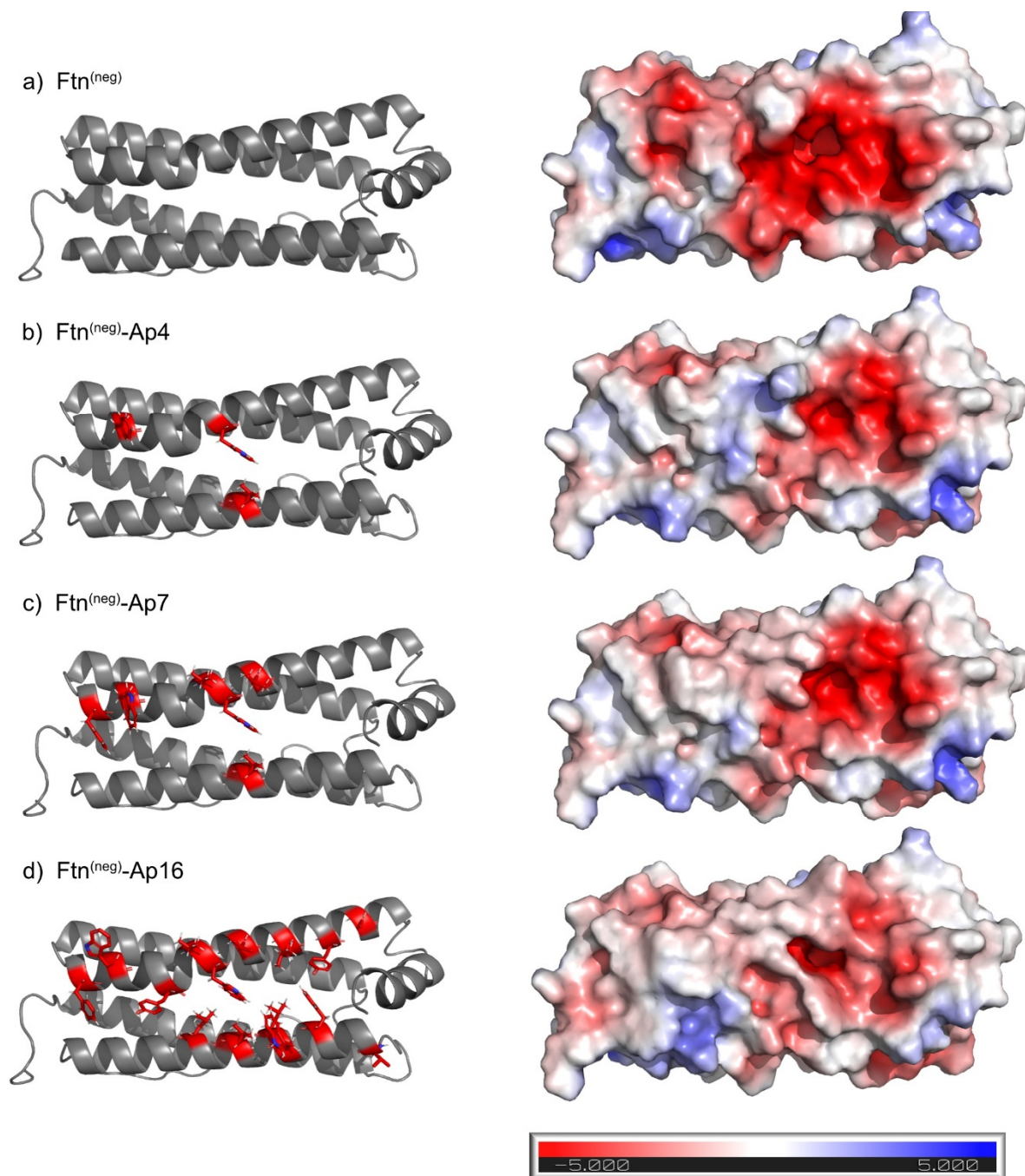


Figure 5.27: Structure and surface potential of Ftn^(neg) mutations on the inner surface. Structure of designed ferritin variants (Mutations are highlighted in red) The surface potential of the mutated variants Ftn^(neg) variants (electrostatic potential is given from -5 to 5 kT e⁻¹). Images were generated using the *PyMol* software and the *APBS* tool.^[256]

Protein production and characterization explained in detail in the section 5.2.2.1. In summary, all variants yield correctly folded soluble and assembled protein cages. It appears that ferritin has a high tolerance to mutations. This finding enables to apply more elaborate design techniques like described in the following sections.

The performed Rosetta protocols only give information about the stability of the protein. In an effort to determine how the redesigned variants perform in binding of the PBUTs, ligand-docking protocols were used to evaluate the ligand binding affinity. IS was selected as a ligand for the simulation. The docking protocol first performed a low-resolution docking step in which the sidechains were abstracted as single atoms (centroids) to find an initial binding site. Subsequently, high-resolution docking with full-atom side chains was performed. Finally, the model was energetically minimized and the binding energy was determined by subtracting the energy of the protein without ligand from the energy of the protein-ligand complex. The methods and scripts for ligand-docking can be found in section 8.8.3.2 and 10.5.6. The specific script can be found in section 10.5.6.1. The energy difference is called $dG_{\text{separated}}$. The protocol was performed with multiple ligand starting positions from where the molecule can find an optimal binding environment in a radius of 15 Å. Results and initial starting points are shown in in the appendix in Figure 10.15. The results indicate that $\text{Ftn}^{(\text{neg})}\text{-Ap4}$ has the most favorable binding sites, while the binding strength of the other variants decrease with increasing number of mutations. In the Ap7 and Ap16 variants, mutation sites were selected by the SASA. As a consequence, also polar positive amino acids like histidine or lysine were exchanged. Rosetta's energy function appears to favor interactions between the toxin and these residues probably due to the formation of hydrogen bonds between the charged groups. As visible in the heatmap generated for the $\text{Ftn}^{(\text{neg})}\text{-Ap4}$ (Figure 10.15b), positions with the highest $dG_{\text{separated}}$ values were found near to each other (vertical Positions 3, horizontal position 2,3,4). These positions were located between the two helices allowing interactions with amino acids from both sides. In the next section, the binding energy at these sites was tried to be improved by stochastic exchange of amino acids with the goal to maximize the $dG_{\text{separated}}$ value.

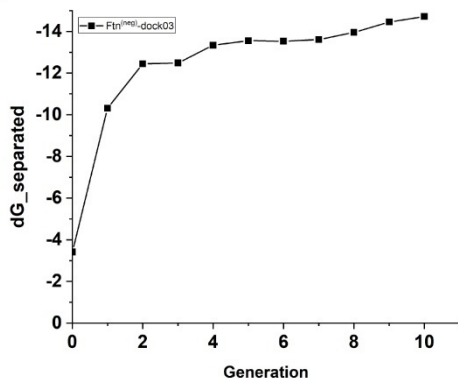
5.2.1.2 Designing binding sites based on ligand docking protocols

According to equation (8) on page 174 in the appendix, already one bound toxin per subunit would result in an adsorption capacity of around 10 mg g^{-1} , surpassing the capacities found for many conventional adsorbents like carbon based adsorbents^[50] or zeolites.^[49] Based on that consideration, the design of strong binding sites for the $\text{Ftn}^{(\text{neg})}$ subunit is a possible strategy to increase PBUT adsorption. Results for chemical modified ferritin indicate that a mixture of hydrophobic and positively charged groups are essential for adsorption. To find suitable

combination of mutations, ligand-docking protocols were applied to the energetically minimized Ftn^(neg) model. Details on the options and scripts used can be found in section 10.5.6. Detailed description of the ligand-docking protocol can be found in section 2.5.2.3. The variants were named after the start position of the toxin with respect to the same grid used for determining the binding energy of the Ap variants (Figure 10.15a). Since the highest values were reached at a location between the two helices at the inner surface, all starting positions were located there. One is near the 3-fold channel (Ftn^(neg)-dock03), one in the middle of the subunit (Ftn^(neg)-dock23) and the last near the small E-helix (Ftn^(neg)-dock43). 5000 independent trajectories were calculated. Resulting output was filtered in terms of sufficiently high Rosetta score, as a measure of protein stability, and first metrics about the ligand binding. More precise metrics for the filtered models were calculated with the Rosetta InterfaceAnalyzer mover. The output was once again filtered to exclude models with insufficient ligand geometry and sorted after their dG_{separated} used as a measure of binding strength. This value was derived from the energy difference between the score of the protein-ligand complex and the protein conformation without the ligand. Early tests revealed that the dG_{separated} values could be further increased by applying the same protocol again on the best structure produced by a previous simulation. This procedure was automated with a bash-script (section 10.5.6.6), which automatically passed the best model from a previous run to the next one to allow a way of evolving the protein to reach a higher binding affinity.

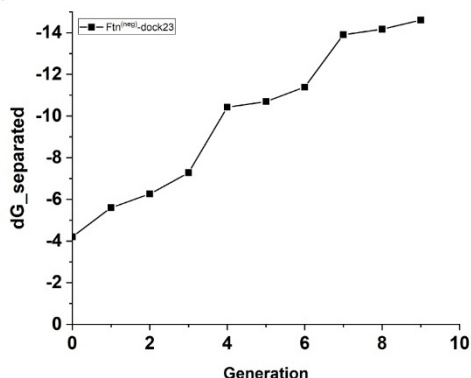
Ten consecutive generations with 5000 independent trajectories were generated for the three starting positions. Resulting dG_{separated} values over all iterations (in the following described as generations) are plotted in Figure 5.28 together with sequence alignment of the best model per generation. All structures show sufficient stability as determined by the total score of the Rosetta simulation (Ftn^(neg): -1394; dock03: -1385; dock23: -1390; dock43: -1398). For the dock23 variant (Figure 5.28b), a steady increase in the dG_{separated} value linked to changes in amino acid composition could be observed. The introduction of amino acids that stabilize the ligands could be observed by steps in the dG value, for example, in generation 4 and 7. For the other variants, in later generations only small improvements to the dG_{separated} linked to changes in the amino acid composition could be found. This is strongly pronounced in the dock43 variant (Figure 5.28c) where after one generation the amino acid composition remained nearly unchanged for all further generations. This is combined with a stagnating dG_{separated} score. Similar results could be seen for dock03 (Figure 5.28a), but here at least in the 8th generation further mutations were found leading to a small increase in the dG_{separated} value.

a) Ftn^(neg)-dock03



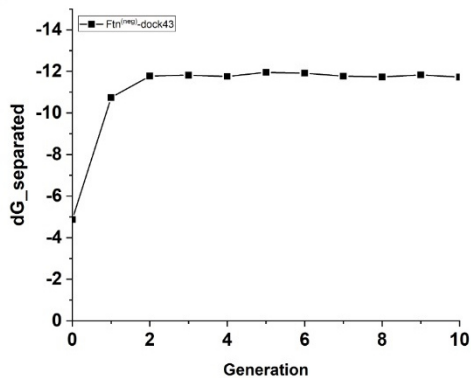
	60		80		140
Ftn(neg)	HEEREHAEKL	MKLQNRGGR	I	TDKNDPHL	CDFIETHYLN EQ
Gen1	...RK...E...	...E...G...	...EH...KRH.R.R.
Gen2	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen3	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen4	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen5	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen6	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen7	...RK...E...	...E...E...	...EH...KRH.R.R.
Gen8	...RK...E...	...R...DK...E...	...EH...KRH.R.R.
Gen9	...RK...E...	...R...DK...E...	...EH...KRH.R.R.
Gen10	...RK...E...	...R...DK...E...	...EH...KRH.R.R.

b) Ftn^(neg)-dock23



		60		140
Ftn(neg)	NFAKYFLHQS	HEEREHAEKL		CDFIETHYLN EQVKA I KELG
Gen1RK...E...DRH.R.R...
Gen2RK...E...DRH.R.R...
Gen3RK...A...DKH.E.K...
Gen4RA...E...DKH.E.K...
Gen5R...A...DKH.E.K...
Gen6	...R...A...	...R...A...DKH.E.K...N...
Gen7	...ER...IG...R...D...	...R...D...DKH.E.K...N...
Gen8	...ER...IG...G...D...	...G...D...DKH.E.K...N...
Gen9	...ER...IG...G...D...	...G...D...DKH.E.K...N...

c) Ftn^(neg)-dock43



		60		140
Ftn(neg)	NFAKYFLHQS	HEEREHAEKL		CDFIETHYLN EQVKA I KELG
Gen1	...A...ER...	...RK...E...DRH.R.R...Y...
Gen2	...ER...	...RK...E...DRH.R.R...Y...
Gen3	...ER...	...RK...E...DRH.R.R...Y...
Gen4	...ER...	...RK...E...DRH.R.R...Y...
Gen5	...ER...	...RK...E...DRH.R.R...Y...
Gen6	...ER...	...RK...E...DRH.R.R...Y...
Gen7	...ER...	...RK...E...DRH.R.R...Y...
Gen8	...ER...	...RK...E...DRH.R.R...Y...
Gen9	...ER...	...RK...E...DRH.R.R...Y...
Gen10	...ER...	...RK...E...DRH.R.R...Y...

Figure 5.28: Variations in toxin binding affinity in amino acid composition during modeling generations. The dG_{separated} value is plotted for each generation and the changes in amino acid composition are highlighted by an alignment to the parental Ftn^(neg) structure.

The observations highlight the strong bias of this method, linked to the very primitive selection algorithm applied to determine the starting structure. Only the best model per generation is selected, resulting in a strong bias. However, compared to the starting point, the affinity of all models could be increased by a two- or three-fold. This strategy was sufficient for the early test done in this thesis. For future investigation, more elaborated evolution algorithm could be introduced by allowing more variability for choosing the model that is passed to the next

generation. This could be achieved by choosing the “successful” model based on a probability increasing with increasing $dG_{\text{seperated}}$. However, especially in later generations, when the input already possesses a high affinity, many models with the same composition as the input model will be generated, and due to its already high $dG_{\text{seperated}}$ the probability to choose one of the many unchanged models instead of one model with changed sequence will be high. By basing the heredity on separate amino acids instead of the whole model this problem could be circumvented. To deduce the impact on the binding affinity of one amino acid at a specific position in the protein a python script (see section 10.5.8), was written, which calculated the average $dG_{\text{seperated}}$ of all models with a specific amino acid at each position and compares it to the mean $dG_{\text{seperated}}$ of all investigated models. If the mean $dG_{\text{seperated}}$ for the chosen mutation is higher, the mutation has a positive effect on the binding affinity. The script was originally written for the analysis of the docking models. In first test runs it was able to identify mutations leading to an increased $dG_{\text{seperated}}$ as seen in Figure 10.16, but it has to be tested on larger data sets. The probability of one amino acid to appear at a specific position in the sequence passed to the next generation could be based on the results of this script.

The calculated $dG_{\text{seperated}}$ values were given in Rosetta energy units and it is difficult to estimate how strong the affinity will be in reality. As a reference, the ligand docking algorithm was performed for an IS molecule located at the Sudlow site II of BSA, which is known to bind IS in the human body. The resulting $dG_{\text{seperated}}$ value was found to be -12.7. This is comparable to the values calculated for the designed binding sites. Ligand position found in the model with the best binding energy together with the introduced mutations are illustrated in Figure 5.29. For Ftn^(neg)-dock03 and dock23 Figure 5.29b and c several mutations were introduced around the ligands. Multiple positively charged amino acids were introduced leading at least for the dock03 and 43 variants to the formation of a positively charged patch. In dock03, the sulfate group of the toxin was located near the positively charged patch while the ring structure was located in more negatively charged area. Some mutations are quite far away from the toxin structure. In which way they influence the toxin binding could not be determined completely, but removing them from the model will decrease respective $dG_{\text{seperated}}$ values. Despite the fact that also positively charged amino acids were introduced, the surface potential of dock23 did not changed significantly. The sulfate group was located in an area of strong negative surface potential. The ring rests directly upon the protein backbone where the residues were replaced with glycine amino acids.

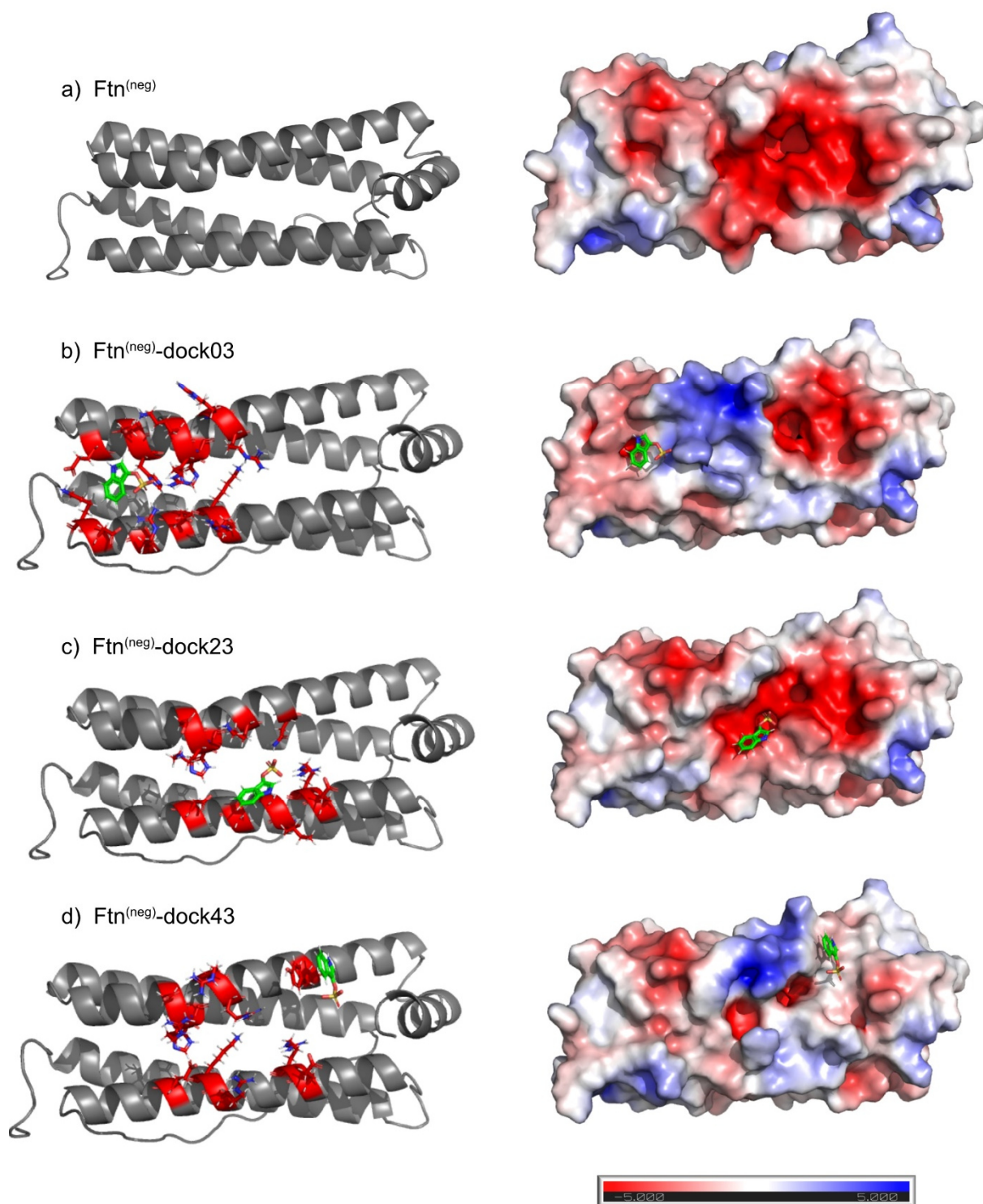


Figure 5.29: Positions of ligands, mutations and surface potential of the designed $\text{Ftn}^{(\text{neg})}$ -docking variants. Cartoon representation of one subunit is shown for all three redesigned variants and the parental structure. For redesigned structures, introduced residues are highlighted in red and the position of the ligand is shown. Additionally, the accessible surface is and their potential is shown for all variants (electrostatic potential is given from -5 to 5 kT e^{-1}). Images were generated using the *PyMol* software and the *APBS* tool.^[256]

The structure of dock43 in Figure 5.29d only showed one introduced tyrosine near the toxin and all other mutations were far away and appear to not have a direct influence on the toxin binding. The strongly negatively charged patch located near the E-helix was either buried by sterically demanding groups or overruled by positively charged amino acids. As a result, the toxin was located at a rather uncharged region, with only a small negative charged patch located near the ring-nitrogen atom.

To further evaluate the proposed binding sites, the chemical environment of the toxin in the different structures was investigated. Respective illustrations are given in Figure 5.30.

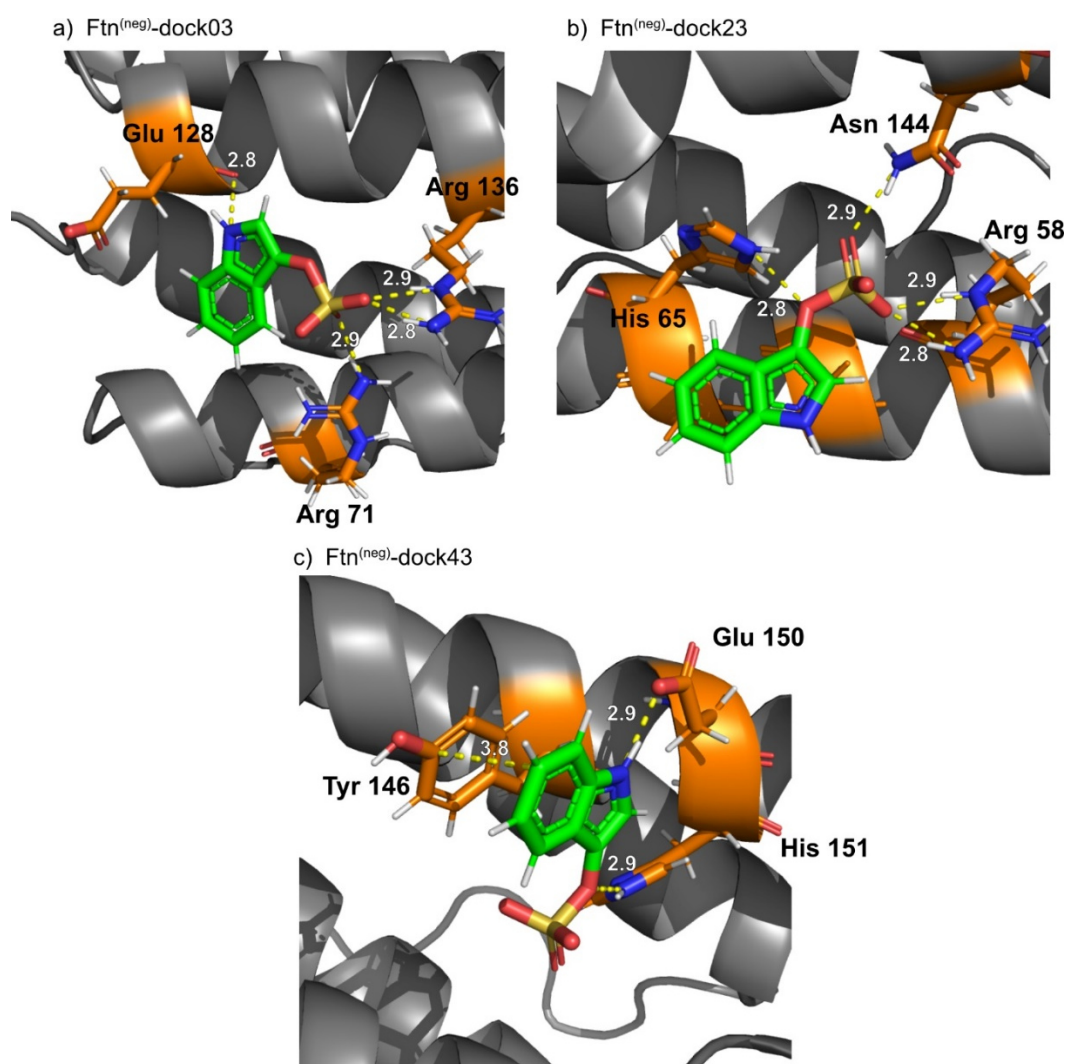


Figure 5.30: Chemical environment of toxins in docking models. The toxin-protein complex was further inspected using the *PyMol* and *Coot* software. Functional groups and residues in vicinity to the toxin were identified and distances between functional groups of the protein and the toxin were measured as indicated by the yellow dotted lines.

In the model dock03 (Figure 5.30a), the toxins seem to interact with three residues (Glu128, Arg71, Arg136). The sulfate group points towards the functional groups of two arginine residues with distances between 2.8 Å and 2.9 Å, which are typical for hydrogen bonds.^[257] It appears that the sulfate group is stabilized by up to three hydrogen bonds. An additional hydrogen bond is possible between the nitrogen atom in the ring structure of the toxin and the carbonyl group in the backbone of Glu128. Despite the hydrogen bonds, no interactions between any residues and the hydrophobic ring can be observed. For the dock23 model, a similar stabilization of the sulfate group by formation of up to four hydrogen bonds with the residues Arg58, His65 and Asn144 can be observed in Figure 5.30b. The ring structure of the toxin is located directly at the backbone, where only two glycine residues are present, but similar to dock03 no further stabilization of the ring is observed. In contrast to that, the model dock43 (Figure 5.30c) shows interactions between the toxins indole-ring and the aromatic system of Tyr146. The distance between them is around 3.8 to 4.0 Å, which is expected for π - π interactions. The nitrogen atom of the indole-ring is further stabilized by hydrogen bonds between the carboxylic acid group of Glu150. The sulfate group is stabilized by one additional hydrogen bond to His151. The additional stabilization of the hydrophobic ring of the ligand is unique to this structure, but the sulfate group is only stabilized by one hydrogen bond to a histidine. This could explain why the dG_separated value of this structure is lower than the others, where three hydrogen bonds are present to stabilize the sulfate group.

To maximize the adsorption capacity, it is interesting to combine multiple binding sites on one subunit. Consequently, a variant combining all three designed binding sites was modelled. Respective alignment and the structure with the highlighted mutations can be seen in Figure 5.31.

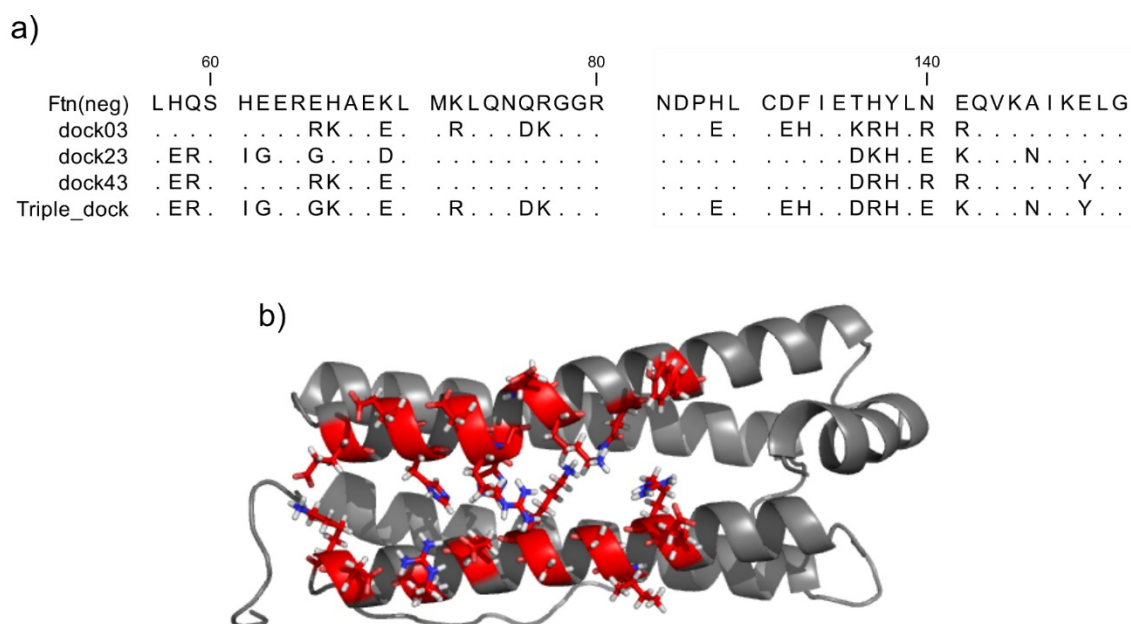


Figure 5.31: Ftn^(neg) with all binding sites combined. a) Alignment of variants with the single binding sites and the variant with the combined binding sites. B) structure of variant with the combined binding sites.

In many natural examples like the Sudlow II site of HAS, small molecules like the PBUTs are bound in specific binding pockets and are partly buried between the protein chains.^[30] In the surface potential images in Figure 5.29 it is visible that two backbone helices are diverging apart and opening a small space near the E-helix. In preparation to the design of a dedicated binding pocket, the residues in this small pocket were exchanged to alanine in order to expand the size. The ligand was placed in the pocket and 30 generations of the ligand-docking protocol were performed. Multiple mutations located around the ligand were introduced (Figure 5.32a). In total, 12 amino acids were exchanged. Inspection of the accessible surface area and their potential in Figure 5.32b revealed that the placement in the cavity hides the hydrophobic indole-ring from the bulk solvent, while the hydrophilic sulfate groups pointed outwards the pocket. The pocket still possessed a strong negative charge probably due to the high density of negatively charged residues in the area. Residues involved in stabilization of the toxin are highlighted in Figure 5.32c. The sulfate group is stabilized by three hydrogen bonds with amino groups from Gln54, Arg143 and Arg147. Favorable orientation and distance between the indole-ring and the aromatic ring of Tyr144 indicating π - π interactions stabilizing the hydrophobic part of the toxin. Additionally, a hydrogen bond between the partial positive charged nitrogen atom in the toxin's indole-ring with the carboxylic acid group of Asp61 is observed. Due to the many favorable interactions, the resulting $dG_{\text{separated}}$ value is with a value of -27 the highest achieved in all simulations. Despite the mutations introduced deep in

the backbone the protein should still be stable according to the total_score of -1319, which is in the same range as unmodified ferritin.

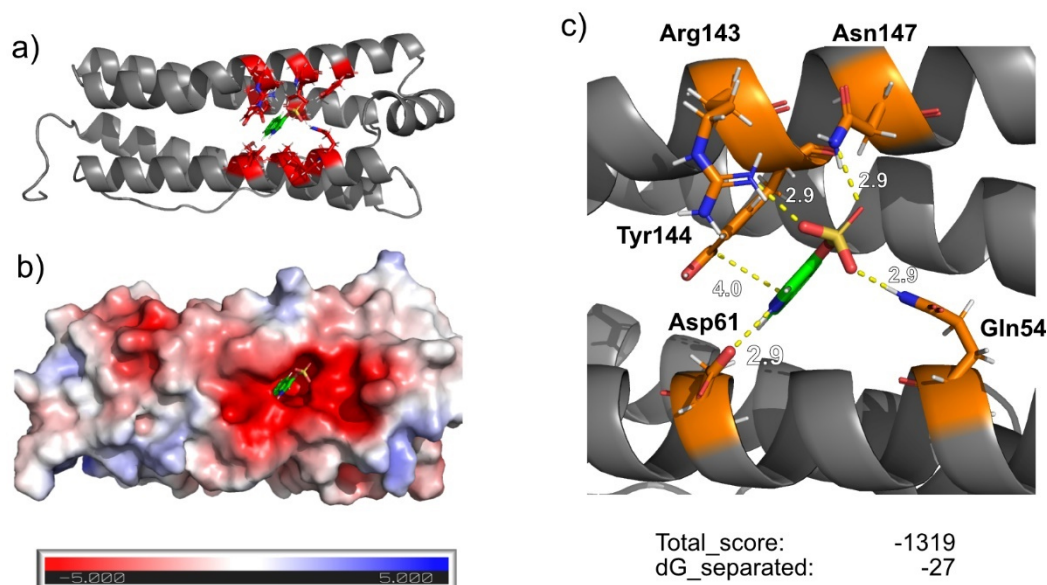


Figure 5.32: Design of a binding pocket. a) Mutations introduced to the Ftn^(neg) subunit to generate a PBUT binding pocket. b) Solvent accessible surface area and surface potential. The toxin is placed in a buried pocket with a strong negative potential (electrostatic potential is given from -5 to 5 kT e⁻¹). c) Chemical environment of toxin in binding pocket. Stabilizing interactions for toxin sulfate group and indole-ring can be observed. Images were generated using the *PyMol* software and the *APBS* tool.^[256]

In summary, it can be said that the ligand docking protocols could be applied successfully to the ferritin structure. In this work, the ligand docking was restricted to the IS, but the method can also be adapted for other toxins. Four variants with promising binding sites were selected for *in vitro* experiments to investigate protein stability and PBUT binding capacity. However, in order to reach the site, the PBUT needs to pass the pores of the protein cage and enter the inner cavity. Modifications to increase the mass transport are explored in the following part.

5.2.1.3 Pore modification

Increasing the diameter of the pores is useful for the loading of small molecules to the ferritin cavity. In most cases the expansion is reached by changing external conditions like increasing the temperature or moderate chaotropic conditions. This is not suitable for the removal of PBUTs. Expanding the pores by genetic modifications is reported for several protein cages and is often achieved by deletion of amino acids in the area of the pore. This was demonstrated for *Thermotoga maritima* encapsulin where up to 9 amino acids lining the channels were either

deleted or replaced with glycine leading to an 11 Å increase in pore diameter and 7-fold increase in mass transfer.^[258] It was further demonstrated that increasing the size and changing the polarity of the pores could increase the flux of artificial cargo in the protein cage.^[259] Enlargement of the 4-fold channel of ferritin is reported by truncating the E-Helix leading to a significantly expansion of the 4-fold channel.^[260] Similar enlargement of the 3-fold channel is not reported, but detailed investigations about the influence of amino acids near the 3-fold channel on the iron uptake are available. Hydrophilic amino acids (His-118, Asp-131, Glu-134, His-134, His-136 and Asp-139) are lining the 3-fold channel with Asp-131 located at the narrowest part of the pore. These negatively charged amino acids are reported to be responsible for the selectivity of the ferritin towards divalent negatively charged ions as determined by electrostatic calculations.^[150] Their influence on the mass transfer was also experimentally verified by exchanging negative charged amino acids like His-118, Asp-131 or Glu-134 to alanine or isoleucine leading to a decrease of ferritins catalytic activity.^[154,261,262] While in ferritins found in vertebrates iron ions mainly enter through the 3-fold channel,^[263] in ferritins originating from prokaryotes and plants ions are also transported through the 4-fold channel.^[264,265] It was demonstrated that suitable mutations to vertebrates ferritin allow iron entry through the 4-fold channel.^[266]

The reported findings suggest that exchanging the amino acids lining the 3-fold channel to non-polar amino acids could increase the uptake of the anionic PBUTs by weakening the negative charge density and increasing the pore diameter. The established modeling protocols allow to create respective designs and evaluate their approximate stability. Residues near the 3-fold channel were exchanged for alanine residues. Resulting models exhibit similar Rosetta scores as the unmodified protein as seen in Figure 5.33. The surface potential of the models was determined with the *PyMol APBS* tool^[256] and is illustrated in Figure 5.33.

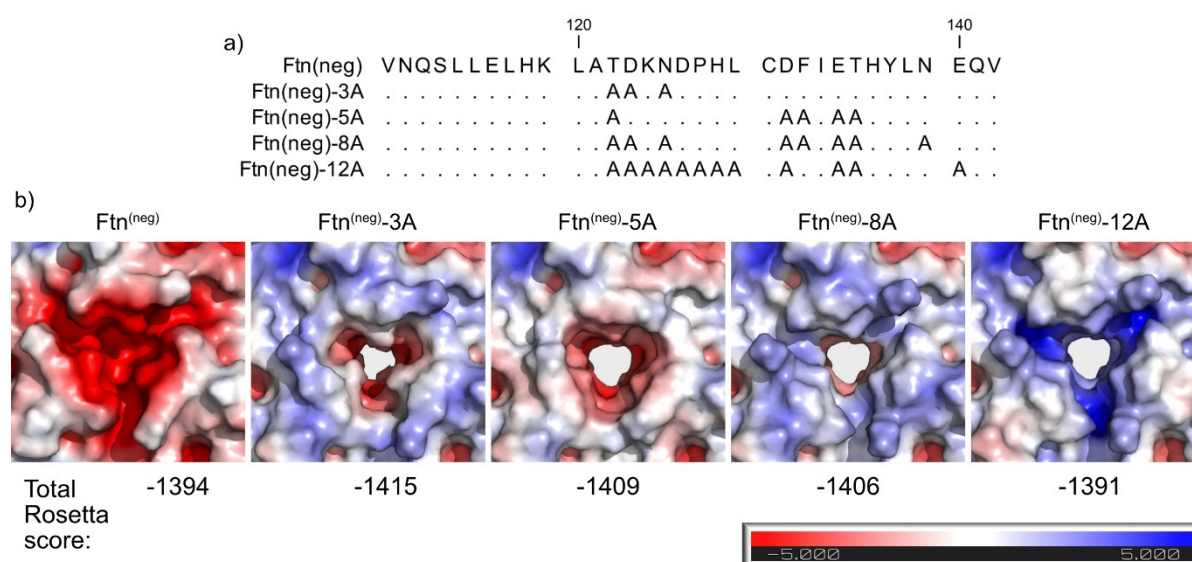


Figure 5.33: 3-fold channel redesign. a) Sequence alignment of Ftn^(neg) and redesigned variants. b) Surface area and charge for Ftn^(neg) variants with increasing numbers of residues located at the 3-fold channel replaced with alanine residues, resulting in increase in pore size. By targeting negatively charged residue the polarity can be inverted (electrostatic potential is given from -5 to 5 kT e⁻¹). Images were generated using the *PyMol* software and the *APBS* tool.^[256]

Significant increase in pore size can be observed with increasing number of introduced alanine residues. Additionally, by exchanging negatively charged amino acids, the polarity at the pore entrance changed to positive values. Even after introduction of eight alanine amino acids in Ftn^(neg)-8A, the pores have still some negative surface potential left as seen in Figure 5.33. As an early test, the variant Ftn^(neg)-3A was combined with the Ftn^(neg)-Ap4 variant and tested in terms of stability and PBUT adsorption (see section 5.2.2.4 and 5.2.5.2). In native cages, the high density of negatively charged amino acids near the pore transport iron ions to the ferroxidase site. Design ideas concerning the transport way and the ferroxidase site are discussed in the next section.

5.2.1.4 Removing the ferroxidase site and iron transport pathways

Ftn^(neg) is a supercharged version of the human heavy chain ferritin. The H chain variant contains the ferroxidase site, which is able to oxidize iron ions and store the resulting iron oxide. This reaction could lead to a depletion of iron in the blood in later dialysis applications. To prevent this, mutations to disrupt the ferroxidase site and the iron transport were introduced. The active site of the ferroxidase site is built up from the amino acids E27, Y34, E62, H65, E107 and Q141.^[267,268] The amino acids at this position were allowed to be exchanged to the amino acids bearing either a hydrophobic or positive charged sidechain during a Rosetta fixbb

protocol. Respective mutations targeting the ferroxidase site directly are highlighted in red in the sequence alignment and in the structure in Figure 5.34a and b.

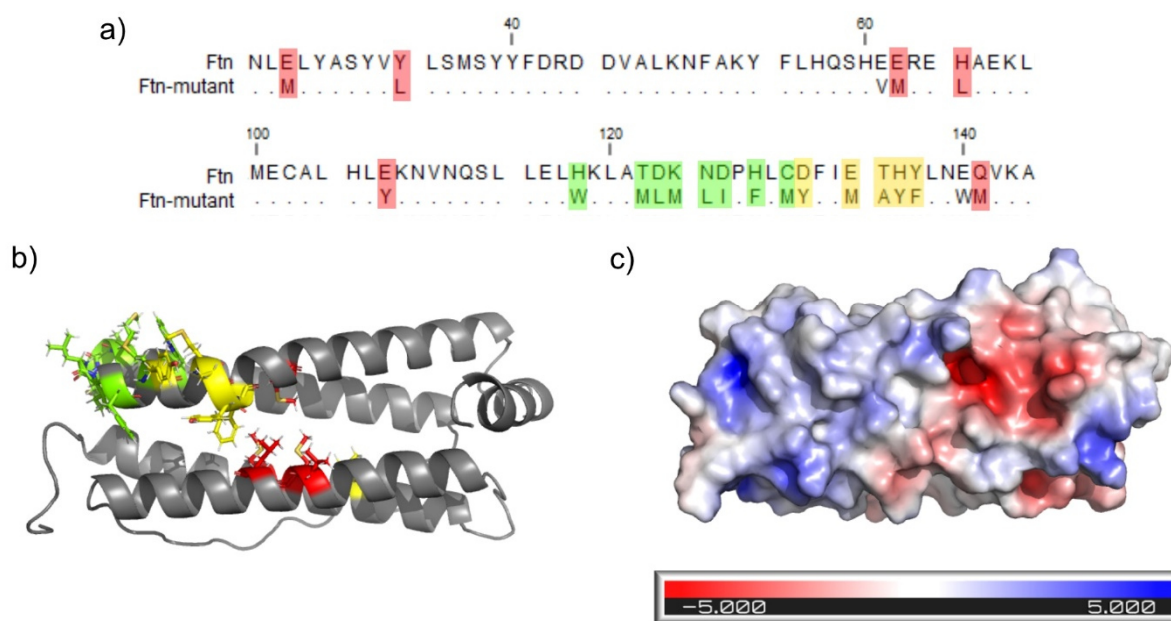


Figure 5.34: Ferritin model with removed ferroxidase site and iron transport pathways. Rosetta fixbb protocols targeting amino acids present in the ferroxidase site (red), in the iron transport pathway (yellow) and in the 3-fold channel (green). A) Introduced mutations are shown in a sequence alignment and b) in the protein structure. C) The surface charge has shifted to positive or uncharged values in regions where the mutations are present (electrostatic potential is given from -5 till 5 kT e⁻¹). Images were generated using the *PyMol* software and the *APBS* tool.^[256]

Additionally, negatively charged amino acids transport iron ions from their entry point at the 3-fold channel to the ferroxidase site.^[268] To prevent the general entry of ions, which could possibly hinder the PBUT adsorption, these mutations were also targeted for redesign. Respective mutations are highlighted in green if present in the 3-fold channel or in yellow if part of the transport pathway in Figure 5.34a and b. In the regions of introduced mutations, the surface potential changed to apolar or positive values as seen in Figure 5.34c. The introduced amino acids at the 3-fold channel (green in Figure 5.34b) have all bulky side chains which could hinder overall PBUT uptake. Probably they could be better exchanged to alanine residues as suggested in the models in section 5.2.2.4. This design showed that it is possible to remove the active site from ferritin, which should be included in future designs.

In this work, multiple protein variants were designed *in silico* guided by the Rosetta molecular modeling software. In the following sections, the expression and purification of selected

variants will be investigated, solid adsorbent material will be fabricated and finally the adsorption capacity towards the PBUTs will be determined for selected variants.

5.2.2 Expression, purification and characterization of redesigned ferritin variants

In this section, the expression and purification of the *in silico* designed ferritin variants is described. The resulting proteins are characterized to verify the correct assembly and the presence of the introduced mutations.

The genetic constructs for all designed variants were produced by the *GenScript Biotech Corp.* The pET22b(+) vector was used. The received plasmids were introduced in calcium competent *E. coli* production strains using a heat-shock method. Experimental details for the performed protocols as well as on protein production and purification can be found in the methods section 8.4. Protein expression and purification was performed using established protocols for Ftn^(neg). In short, *E. coli* production strains bearing the respective plasmids were grown to an OD₆₀₀ of 0.6 in TB medium at 37°C. Protein production was induced by addition of IPTG and protein production was performed at 18°C for 48 hours. The cells were harvested and lysed using sonication techniques. The majority of *E. coli* proteins were separated using heat precipitation. Protein cages were further purified with ammonium sulfate precipitation, ion-exchange and size-exclusion chromatography. Complete experimental details can be found in section 8.4.2. First, the variants with the introduced hydrophobic amino acids are described, followed by variants derived from ligand-docking protocols. At the end, variants with changes to the pore and introduced binding pockets are evaluated.

5.2.2.1 Ferritin variants with hydrophobic sidechains

Ferritin variants with an increased density of amino acids with hydrophobic side chains were designed in chapter 5.2.1. Ion-exchange chromatography revealed an altered elution behavior of the mutated variants. The conductivity of the buffer mixture can act as a measure of overall protein charge. Since the cages should be already assembled in this step, only the charge of the outer surface can be observed. Resulting IEC chromatograms are shown in the appendix Figure 10.17. In comparison of the unmodified Ftn^(neg) eluting at a conductivity around 47 mS cm⁻¹, the mutated variants eluted at lower values around 42 mS cm⁻¹. This was unexpected, since all mutations were limited to the inner surface of the protein and should have no influence on the overall charge of the outer surface. Since the shift was observed for all variants, the four mutations they all have in common must be responsible for this phenomenon. Respective mutations are E61V, E62A, D131F and E140W. All replace negatively charged residues, so it is plausible that they could lead to a decrease in charge. And apparently due to the long-range of electrostatic charges, this also affected the outer

surface. This finding should be considered in future research since it was believed that changes to the inner surface have no influence on the outer surface. The potential of the outer surface was determined using the *PyMol APBS* tool^[256] and are illustrated in the appendix in Figure 10.18. Here, only a minor difference in the outer surface potential can be observed. The region is far away from the mutation sites, therefore it could be also an artifact because of slightly different side chain conformation from the different energy-minimized structures. Indications on the correct assembly of the protein cage could be derived from the elution volume of the size-exclusion chromatography. An overlay of the SEC chromatogram of the ferritin variants is depicted in Figure 5.35a.

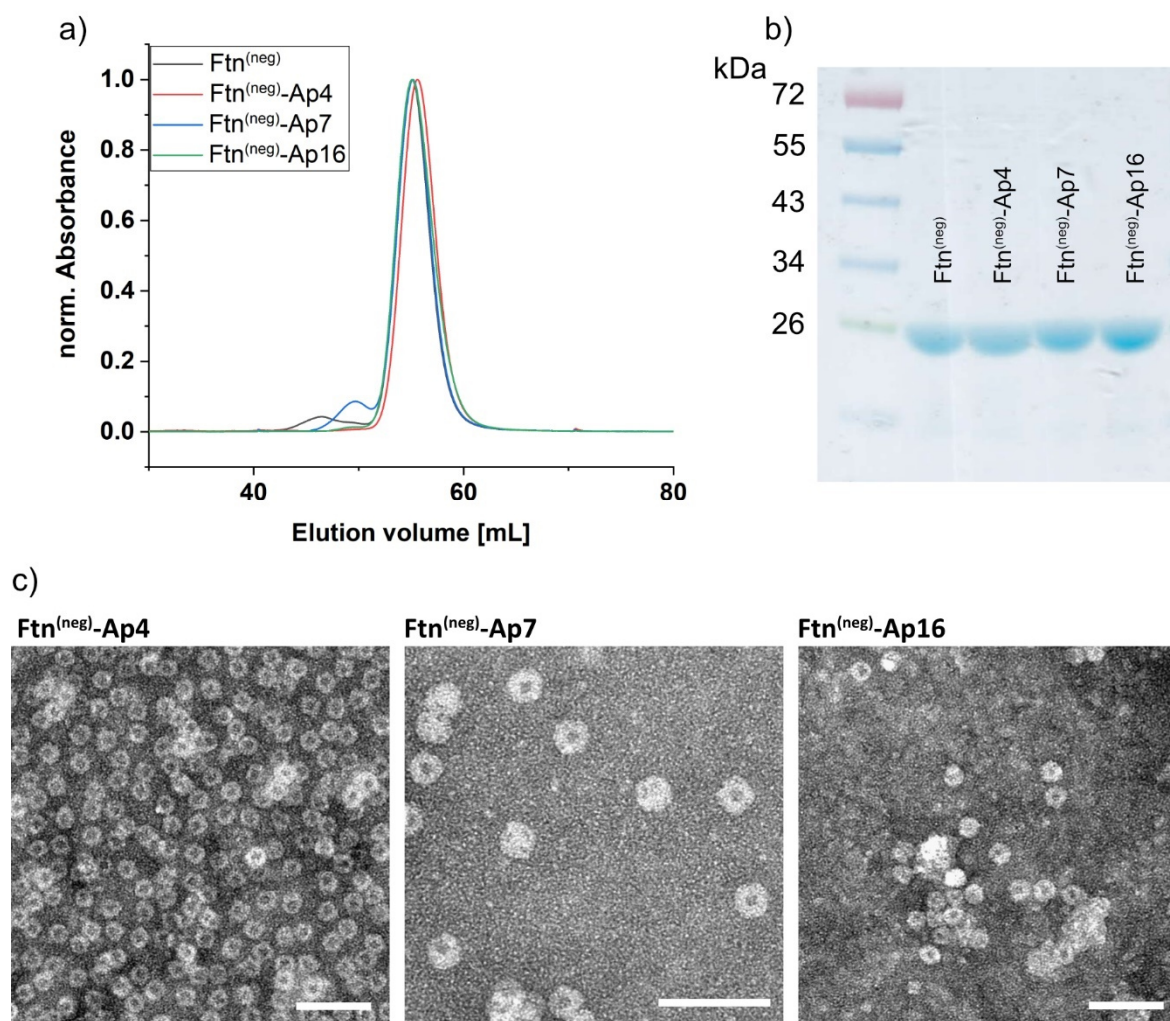


Figure 5.35: SEC chromatogram, SDS-PAGE and TEM images for Ftn^(neg)-Ap variants. a) SEC chromatogram showed elution for all variants at the same volume. b) Presence of the ferritin variants in the eluted fractions is verified by SDS-PAGE c). The correct assembly of the protein cages were confirmed by negative-stain TEM images. Scale bar 50 nm.

The redesigned variants showed the same elution volume as the unmodified ferritin, indicating a similar size and a correct assembly. Protein-containing fraction were collected and tested for the presence of the ferritin variants with SDS-PAGE. The respective gel is shown in Figure 5.35b. Bands are found slightly below the mass of 26 kDa, which is in good agreement with the mass of the ferritin subunit. No further bands were observed indicating that the ferritin variants could be separated from all other proteins present in the host organism. The SEC results already indicate that the new ferritin variants were still able to assemble to full cages. This was further confirmed by negative-stain TEM images shown in Figure 5.35c. In all images, the characteristic spherical cage structure can be observed. To verify that the correct mutations are present in the ferritin variants, ESI-MS measurements were performed and the measured mass were compared to the mass theoretical expected for the given amino acid composition. The results are summarized in Table 5.5 and the full spectrum for each variant can be found in Figure 10.20 in the appendix. For all variants, the measured mass exactly fit the expected mass, confirming that all mutations are present in the given variants. However, for the Ftn^(neg)-Ap4 variant, the final mass only differs by 1 Da from the mass of the parental Ftn^(neg) structure. Due to this overlap the question if the mutated residues are indeed present in this variant cannot ultimately answered by this result. Investigations of the respective crystal structure in the following will confirm the presence of the mutations.

Table 5.5: Molecular mass of Ftn^(neg)-Ap variants. Molecular mass for the Ftn^(neg)-Ap variants as determined by ESI-MS measurements in comparison to theoretical mass for the given amino acid sequence.

Protein	MW _{Theo.} ^[a]	MW _{Meas.}
	[kDa]	[kDa]
Ftn ^(neg)	21.196	21.196
Ftn ^(neg) -Ap4	21.197	21.197
Ftn ^(neg) -Ap7	21.202	21.202
Ftn ^(neg) -Ap16	21.160	21.160

[a] Determined with ProtParam tool^[242]

Even though the surface charge is slightly decreased for the redesigned variants, hanging-drop crystallization techniques under the same conditions used for Ftn^(neg) yielded crystals with the expected morphology, as seen in the Figure 5.36. However, for the variant Ftn^(neg)-Ap7 some precipitant could be observed during crystallization. Despite the fact that the outer

surface is not changed, the mutations seems to have some influence on the higher-order assembly.

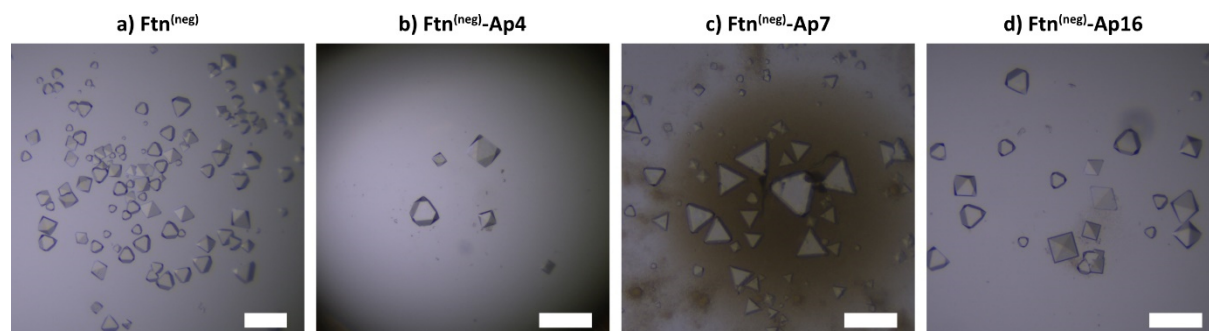


Figure 5.36: Crystals of Ftn^(neg) variants with increased density of hydrophobic amino acids. Crystallization under conditions optimized for Ftn^(neg) yield crystals for (b-c) all variants with increased density of hydrophobic side chains similar in size and morphology to a) unmodified Ftn^(neg) crystals. Scalebar 400 μm .

Diffraction data for the respective crystals were collected at the *Deutsche Elektronen-Synchrotron DESY* and the crystal structures were determined. Data statistics and refinement details for the respective crystal structures are summarized in Table 10.8. To verify the presence of the introduced mutations, electron density omit maps were fabricated for each variant. The maps were calculated using a model where all introduced mutations were replaced by the amino acid glycine. In the omit map for the Ftn^(neg)-Ap4 variant shown in Figure 10.22 the introduced mutations fit very well to the $F_{\text{O}}-F_{\text{C}}$ difference map (in green). This validates the presence of the mutations, which was questionable due to the low difference in mass of this variant and its parental structure. The omit maps for Ftn^(neg)-Ap7 (Figure 10.23) and Ftn^(neg)-Ap16 (Figure 10.24) also show a good fit of the introduced mutations with the free electron density validating the incorporated mutations

In summary, the designed ferritin variants could be expressed and purified with already established protocols. Correctly assembled containers were obtained as confirmed by SEC and TEM results. The correct introduced mutations could be verified by ESI-MS measurement and can be identified in the respective crystal structure. Despite the fact that the mutations were introduced inside the cavity, elution behavior in ion-exchange chromatography indicated small changes to the potential of the outer surface. Nevertheless, crystallization conditions optimized for Ftn^(neg) still yielded crystals with expected morphology. It is interesting that even the Ftn^(neg)-Ap16 mutant with 16 mutations yielded stable protein cages. This indicates the high plasticity of the ferritin backbone towards mutations. Further it is proven that the Rosetta molecular modeling suite is a suitable tool to determine mutation sites, which will lead to correctly folded and soluble proteins. In the next section, it will be investigated if this also holds

true for the variants generated by ligand docking protocols where the stability of the cages was only a secondary factor during the modeling.

5.2.2.2 Ferritin variants with toxin binding sites

Ferritin variants bearing a toxin binding site were designed in section 5.2.1.2. The proteins were purified with the already described precipitation methods, the IEC chromatogram shown in Figure 10.19 again yielded information about the surface charge. Comparing the conductivity for all variants reveals that the Ftn^(neg)-dock03 and dock23 (Figure 10.19b and c) eluted at roughly the same buffer conductivity as the Ftn^(neg) (Figure 10.19 a). On the other hand, the variant dock43 (Figure 10.19d) eluted at reduced conductivity values similar to the Ftn^(neg)-Ap variants (Figure 10.17b, c and d). This further strengthens the hypothesis that modifications to charged amino acids in the cavity can have an influence on the outer surface. In order to better understand this phenomenon, the sequences of the unchanged Ftn^(neg), the Ftn^(neg)-Ap4 and the docking variants were overlaid and mutations involving charged amino acids were examined. All redesigned variants have a mutation at position E140 to either an uncharged or positive charged residue in common. Despite the similar elution behavior in the IEC, the dock43 and Ap4 variant do not share any further mutations. On the other hand, the variants dock03 and dock23 share two mutations with the Ap4 variants (dock03: D131, E140; dock23: E61, E140). The numbers of removed and added negative charged amino acids are summarized in Table 10.6. The variant dock03 in sum has two negative charged residues more than the Ftn^(neg). For this variant, the conductivity during elution is found to be the highest with a value of 49 mS cm⁻¹. The conductivity at which dock23 is eluting is slightly lower with 48 mS cm⁻¹. Both values are higher than for unmodified Ftn^(neg), which elutes at a value of around 47 mS cm⁻¹. However, the difference between this value is small and could be just a coincidence. More suitable analytics such as zeta potential measurements could be performed to verify the results. If they hold true these results indicate that the surface charge of the outer surface can be altered by charged groups on the inner surface. This observation can probably be explained by the long-ranged nature of electrostatic forces. Interestingly this effect is also and possibly even greater, depending on the position of the mutation in the inner cavity, since dock43 shares the same net change in negative charged amino acids as dock23. Despite that the conductivity is significantly reduced to a value of 43 mS cm⁻¹. Since the described changes to the outer surface charge had no great influence on the higher-order assembly of the cages, this phenomenon was not further investigated.

The SEC chromatograms shown in Figure 5.37a reveals elution of all docking variants at a similar volume as the Ftn^(neg) variant. To verify that the respective protein was indeed present in the eluted fractions, SDS-PAGE was performed as shown in Figure 5.37b

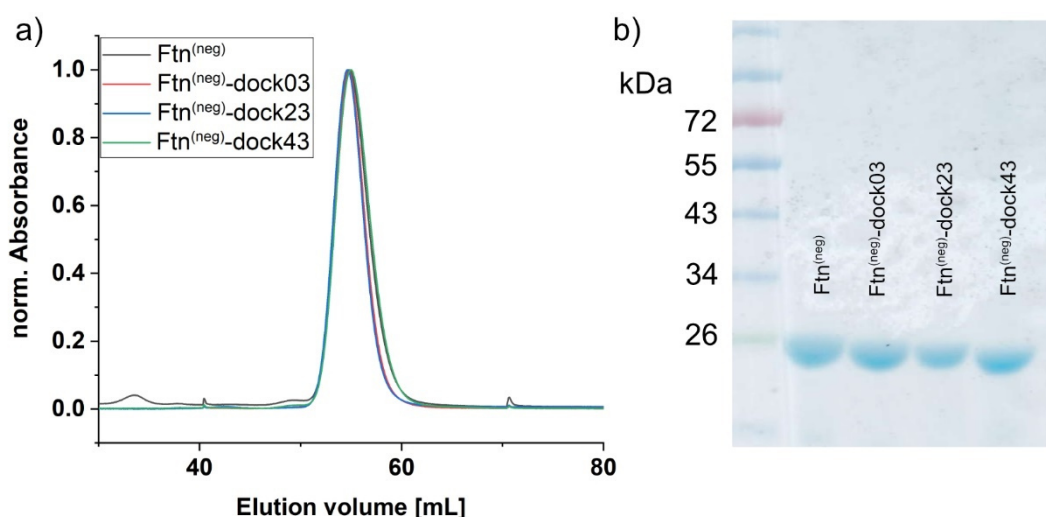


Figure 5.37: SEC and SDS-PAGE of Ftn^(neg)-docking variants. a) SEC chromatograms show elution for all variants at the same volume. Presence of the ferritin variants in the eluted fractions are verified by b) SDS-PAGE.

In the SDS-PAGE, bands for all variants were found slightly below the marker for 26 kDa, which fits to the mass of the ferritin subunit. No further bands were detected suggesting that the protein cages could be successfully separated from all protein contaminations. The SEC elution behavior indicates that the cages are completely assembled. To verify the presence of the introduced mutations, again ESI-MS measurements were performed. The results are summarized in Table 5.6 and respective spectra can be found in Figure 10.21 in the appendix. The determined mass fitted perfectly to the expected mass revealing that the produced variants have the same amino acid composition as the designed variants.

Table 5.6: Molecular mass of Ftn^(neg)-docking variants. Molecular mass for the Ftn^(neg)-dock variants as determined by ESI-MS measurements in comparison to theoretical mass for the given amino acid sequence.

Protein	MW _{Theo.} ^[a]	MW _{Meas.}
	[kDa]	[kDa]
Ftn ^(neg)	21.196	21.196
Ftn ^(neg) -dock03	21.287	21.287
Ftn ^(neg) -dock23	21.071	21.071
Ftn ^(neg) -dock43	21.345	21.345

[a] Determined with ProtParam tool^[242]

The ability of the variants to form crystals is investigated in hang-drop crystallization set-up using the same conditions optimized for Ftn^(neg). As seen in Figure 5.38, variants yielded crystals with expected morphology and size. The changes to the surface charge observed in the IEC seems to not drastically influence the higher-order assembly of the cages.

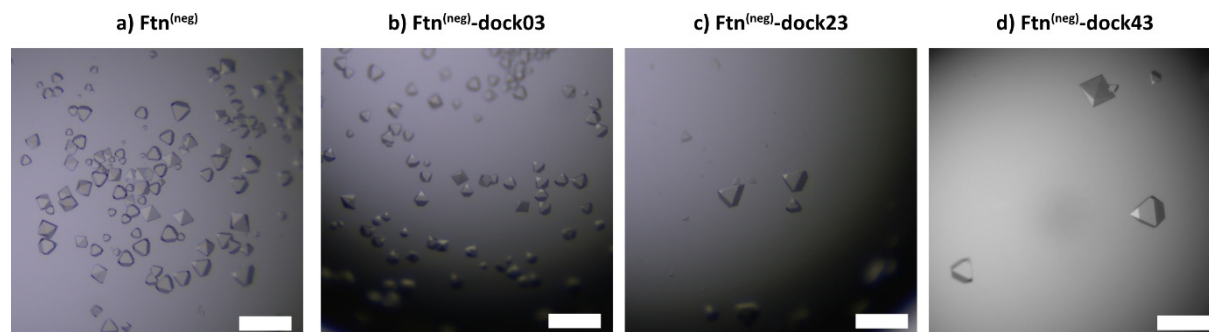


Figure 5.38: Crystals of Ftn^(neg)-dock variants. Crystallization under conditions optimized for Ftn^(neg) yield crystals for (b-c) all docking variants similar in size and morphology to a) unmodified Ftn^(neg) crystals. Scalebar 500 μm .

Diffraction data for the respective crystals were collected at the *Deutsche Elektronen-Synchrotron DESY* and the crystal structures were determined. Data statistics and refinement details for the respective crystal structures are summarized in Table 10.9. To validate the presence of the desired mutations again electron density omit maps were prepared using models, where the introduced mutations were replaced by the amino acid glycine. However, in the variant Ftn^(neg)-dock23 two of the desired mutations actually are glycine (E61G and E64G). Consequently, instead of glycine alanine amino acids were introduced. In the respective omit map shown in Figure 10.26 the red electron density $F_{\text{O}}-F_{\text{C}}$ difference map indicates that indeed no side chain is present validating the presence of the amino acid glycine. In general the overlay between the introduced mutations and the free electron density in the omit maps of dock03 (Figure 10.25), dock23 (Figure 10.26) and dock43 (Figure 10.27) is not ideal. Many of the introduced amino acids are glutamic acid, lysine or arginine possessing long side chains. Due to their flexibility and the various conformations they can adopt in the crystal corresponding electron density could not resolved well for these amino acids. At least, the diffuse electron density indicates the presence of a flexible amino acid expected at this position. The identity of the desired amino acid cannot be verified from the crystal structure alone. However, in combination with the results gathered from ESI-MS measurement, the presence of all mutations located at the desired positions can be verified.

In summary, also the docking variants could be produced and purified with the already established purification protocols for Ftn^(neg). The desired mutations are present and at the desired position as confirmed by ESI-MS measurements and investigation of respective crystal

structures. Despite the high number of mutations, the subunits are still soluble and can assemble to the full protein cage as verified by their elution behavior in SEC. Since there are no mutations present at the outer surface, the cages still crystallize under the known conditions. If the more challenging variant bearing a binding site hidden in a pocket between the back bone or combining the three designed sites also yields soluble proteins is investigated in the following section.

5.2.2.3 Ferritin variants with toxin binding pocket and multiple binding sites

Expression and purification of the ferritin variant with a dedicated binding pocket for IS designed at the end of section 5.2.1.2 were also performed with the established protocols (section 8.4.2). In contrast to the other variants, no protein cages were observed after IEC and SEC. SDS-PAGE after cell lysis shown in Figure 5.39a revealed that the main part of the ferritin visible as a band around 26 kDa remain insoluble in the cell pellet. Similar results could be observed for the variant with a combination of multiple binding sites (Data not shown). This was the only time in this work that a protein variant with favorable Rosetta score does not form a soluble protein.

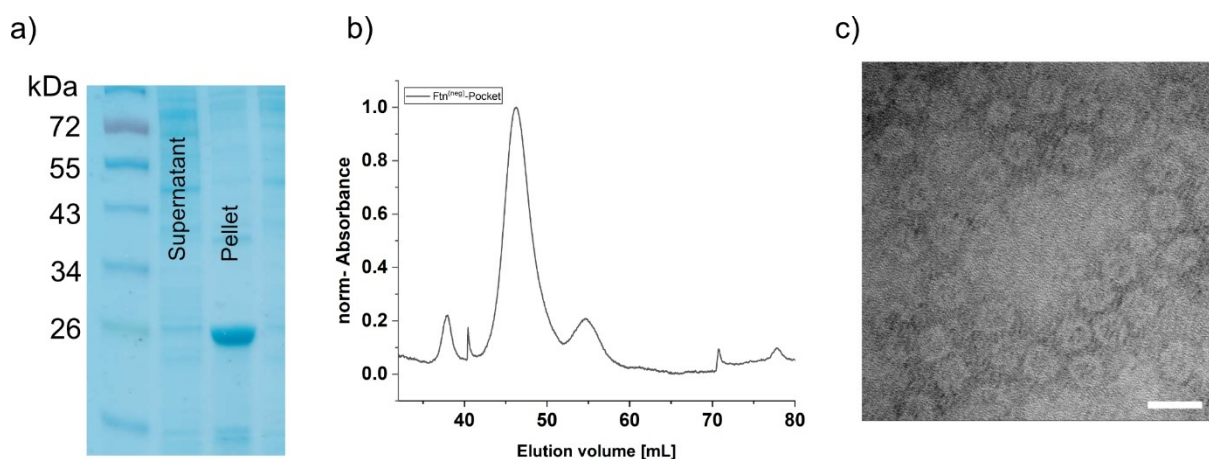


Figure 5.39: Expression and purification of Ftn^(neg) with toxin binding pocket. a) SDS-PAGE after lysis of the production host. The main part of the ferritin could be located as insoluble inclusion bodies in the pellet. b) SEC after solubilization and refolding induced by dilution. The small peak at an elution volume of 55 mL is expected to contain the protein cage. c) Negative-stain TEM images reveal correctly folded cages in the small SEC peak. (Scale bar 20 nm).

Screening of production and purification parameters (cell density during induction, inducer concentration and buffer composition) led to no improvement. Techniques for the purification of protein from insoluble inclusion bodies were tested. In short, first the lysis pellet was washed to remove membrane and cell debris, followed by solubilization of the pellet by treatment with a buffer containing high detergent concentrations. Complete experimental details can be found

in section 8.4.3. The pellet could be solubilized nearly completely. Refolding was induced by reversing the denaturing conditions. For that the protein solution was diluted and allowed to reassemble overnight. SEC of the reassembled solution (Figure 5.39b) showed a large aggregate peak (elution volume of 40 to 50 mL) but also a small peak at an elution volume of 55 mL, where properly folded cages are expected. Correctly folded protein cages could be found in negative-stain TEM images (Figure 5.39c). This result confirms that it is possible to purify and resolubilize the $\text{Ftn}^{(\text{neg})}$ -pocket variant. However, protein yield is very low, and the majority was lost to aggregates. Extensive optimization in terms of refolding concentration, temperature and buffer composition are needed. Also, other techniques for more gentle dilution such as dialysis or even chromatographic methods could be applied. Successful purification in sizable amounts needed for further experiments were not possible in the time frame of this work.

5.2.2.4 Ferritin variants with expanded pores

Ferritin variants with expanded pores due to the incorporation of alanine amino acids at the 3-fold channels were designed in section 5.2.1.3. Since changes at the interface between subunits are delicate, instead of ordering the whole gene construct with up to 12 mutations, the changes were introduced iterative by QuikChange mutagenesis. An intermediate with up to 3 mutations was generated. The genetic construct bears similar mutation to the inner cavity as $\text{Ftn}^{(\text{neg})}$ -Ap4 but three additional mutations at the 3-fold channel. Therefore, it is named $\text{Ftn}^{(\text{neg})}$ -Ap4-3A. Despite the mutations at a sensible area, the variant could be successfully expressed and purified.

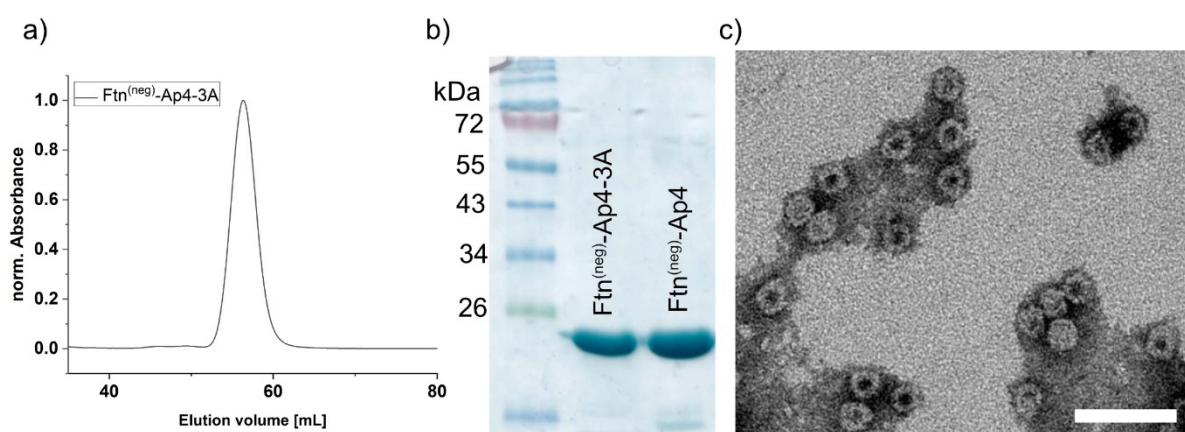


Figure 5.40: Characterization of ferritin with expanded pores. a) Size-exclusion chromatogram of $\text{Ftn}^{(\text{neg})}$ -Ap4-3A revealed elution of intact assembled protein cages. b) Presence of the ferritin in the SEC fractions confirmed by SDS-PAGE. A strong band is observed at the expected mass. c) Negative-stain TEM images of reassembled ferritin cages further confirm correct cage assembly.

Correctly assembled cages were eluted during SEC at expected volume as seen in Figure 5.40a. The presence of ferritin protein in the eluted fractions were verified by SDS-PAGE, revealing a strong band at the expected position (Figure 5.40b). The correctly assembled cages were finally found in negative-stain TEM images (Figure 5.40c). In hanging-drop experiments, it was not possible to produce Ftn^(neg)-Ap4-3A with the established conditions for Ftn^(neg) as shown in Figure 5.41.

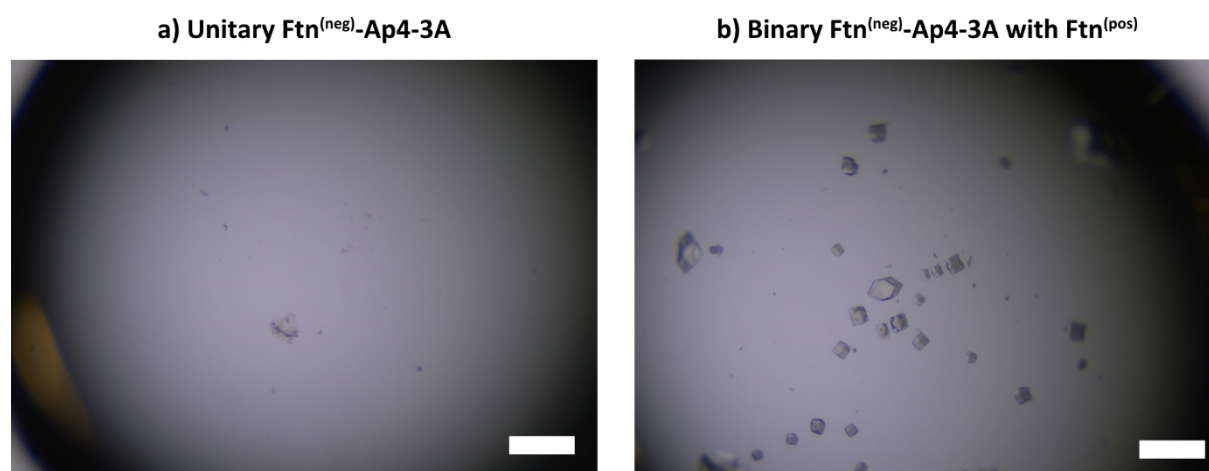


Figure 5.41: Ftn^(neg)-Ap4-3A crystals. a) Crystallization under unitary Ftn^(neg) condition showed no visible crystals. b) Crystallization with Ftn^(pos) produced binary crystals. (Scalebar is equal to 500 μ m)

The mutated positions (T122A, D123A, N125A) are located directly at the opening of the channel and therefore located at the outer surface. In the crystal structure of Ftn^(neg) (PDB-ID: 5JJK) the cages are contacting each other with their 3-fold channel. The introduced mutations are located directly in this contact area and explain why the unitary crystals are not forming anymore. In the binary structure, the interface between cages is located in the middle of a ferritin subunit and the mutations have no effect on the higher-order assembly of the cages. As a result, crystallization under conditions with the positive supercharged ferritin produced binary crystals. According to the results, the exchange of three amino acids to alanine at the 3-fold channel does not interfere with the correct assembly of the protein cages. This opens a route to more expanded and even positively charged pores as suggested by the protein design in section 5.2.1.3. Further expansion by the introduction of more mutations were not possible in the timeframe of this work.

Up to seven Ftn^(neg) variants designed with the Rosetta Software suite could be expressed and purified during *in vitro* experiments. In the following sections, their ability to form a heterogeneous material, the biocompatibility of the resulting material and their adsorption capacity towards the selected PBUTs are investigated.

5.2.3 Fabrication of heterogeneous adsorbent materials

All variants except Ftn^(neg)-Ap4-3A due to its mutations at the crystal interface showed to form crystals by self-assembly in hanging-drop crystallization. These variants were subjected to the batch crystallization techniques previously used for the functionalized variants in section 5.1.4. The surface charge of the Ftn^(neg)-Ap variants was slightly lower than for the negative supercharged protein cages (Figure 10.17). Additionally, in the first experiments with the established crystallization condition, the formation of visible crystals of the Ftn^(neg)-Ap variants took 3 days longer than for the unmodified Ftn^(neg). Therefore, optimal crystallization parameters in terms of precipitant concentration were screened and results are shown Figure 5.42.

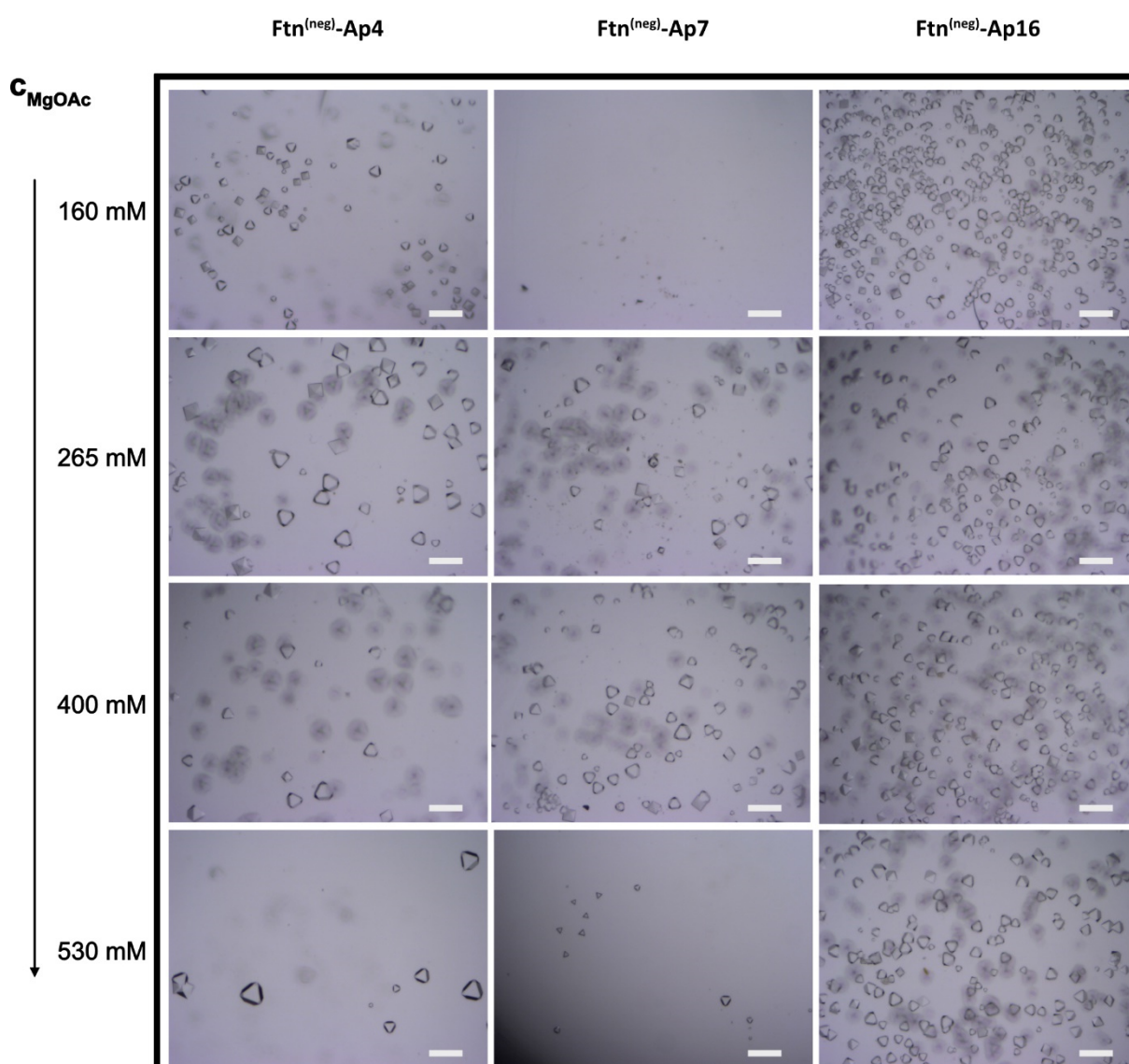


Figure 5.42: Batch crystallization of Ftn^(neg)-Ap variants. Precipitant concentration is screened for the Ftn^(neg)-Ap variants. The protein concentration is kept fixed at 3 mg mL⁻¹. The crystallization behavior of Ftn^(neg)-Ap7 is deviating from the other variants but formed crystals at 265 and 400 mM. Scalebar 250 μ m.

Variants Ftn^(neg)-Ap4 and Ap16 showed crystal formation at all precipitant concentration. Ap4 showed similar behavior like Ftn^(neg) where the crystal size increased with increasing precipitant concentration. This trend is also observed in Ap16 but not so strongly pronounced which could be due the fact that many more crystals were formed. In contrast, the variant Ap7 showed only sizable crystallization at a precipitant concentration of 265 and 400 mM. At the highest concentration, only a few small crystals could be observed and at the lowest no crystal was visible. Probably crystal formation was not finished at the high concentration. The findings for Ap7 agree with the hanging drop crystallization techniques were also only a few small crystals could be found after a week. Why Ap7 showed such a deviation in its crystallization cannot be fully explained from the present data. The IEC elution already showed that mutations inside the cavity can still have an effect at least on the electrostatics of the outer surface, but the deviation between Ap7 and the other variants was small (Figure 10.17). Also, from the set of exchanged mutation Ap7 shares four out of seven mutations with Ap4 and all of its mutations are present in Ap16. This phenomenon could be further investigated by screening different crystallization conditions and investigation of the crystal structure of the achieved crystals. Since the focus of this work is the fabrication of a protein-based adsorbent, this was not further investigated. For batch crystallization in later experiments, a precipitant concentration of 265 mM was selected. For the docking variants, crystals could be achieved in similar amount and size as for the unmodified cage. Fabrication of non-crystallin material yield similar results for all variants. Even the Ap4-3A variant which is not crystallizing under standard conditions can turned into a solid non-crystalline protein material.

Solid adsorbent could be fabricated for all designed variant by techniques already established for the chemical modified protein cages with slightly optimized conditions. In the following section, the biocompatibility of the respective material is investigated.

5.2.4 Biocompatibility

Similar to the functionalized Ftn^(neg) variants in section 5.1.5, the biocompatibility of the redesigned ferritin variants was evaluated in terms of presence of endotoxins and platelet activation. As a representative, the variant Ftn^(neg)-dock43 was selected. The expression of TNF- α mRNA of endothelial cells upon incubation with crystalline and non-crystalline adsorbent fabricated from the Ftn^(neg)-dock43 variant is shown in Figure 5.43a.

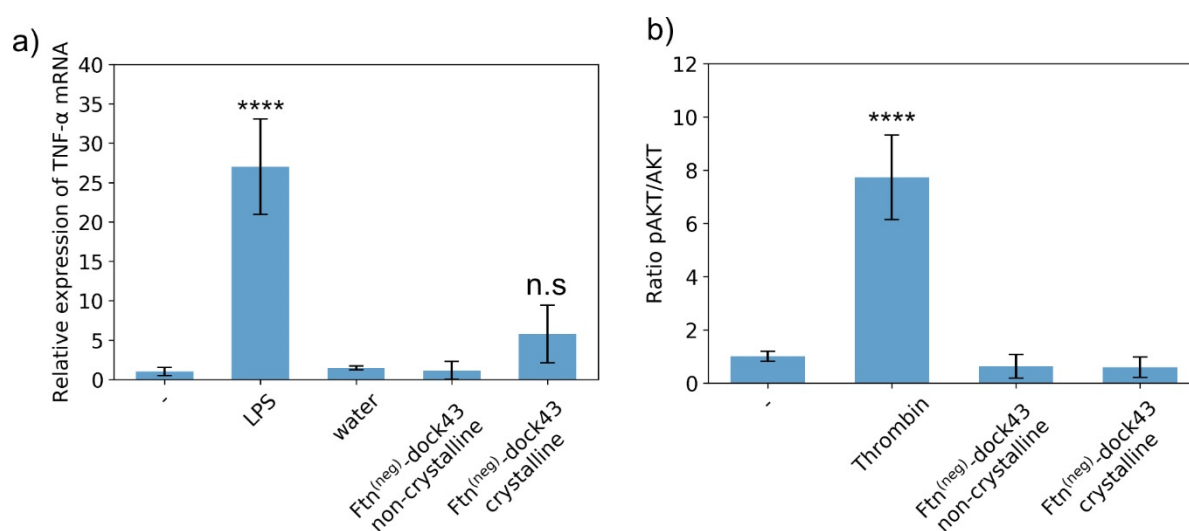


Figure 5.43: Biocompatibility assay for redesigned Ftn^(neg)-dock43. a) Relative expression of TNF- α as a measure for endotoxin contaminations for crystalline and non-crystalline adsorbent. b) Ratio of phosphorylated to non-phosphorylated AKT as a measure for platelet activation. The crystals are fixated with sulfo-SMCC crosslinker. For both assays, no significant deviation from the negative control is observed. **** $P \leq 0.0001$ compared to negative control.

The non-crystalline adsorbent showed a similar level in TNF- α mRNA expression as the negative control sample, indicating absence of endotoxins. The crystalline material showed slightly increased values. Endotoxins belong to the group of lipopolysaccharides bearing multiple negative charged phosphate groups at neutral pH. It could be possible that due to the increased density of positive charged amino acids at the inner surface, binding of this molecules is preferred. However, due to its high molecular weight, it is unlikely that the LPS can reach the inner surface. Also, the increase in TNF- α mRNA expression is not significant. Additionally, the non-crystalline material was composed from exactly the same protein cage, which was expressed and purified in the same way, and showed no increased endotoxin level. All this indicates that the increased value was caused by inaccuracies during the measurement and no contamination with endotoxins was present in the samples. To study platelet activation, in Figure 5.43b, the ratio of phosphorylated to unphosphorylated AKT is plotted for the

investigated materials. For all materials, no elevated pAKT concentration could be measured indicating that the materials did not induce the activation of blood platelets.

Amino acid exchange on the inner surface showed no changes in biocompatibility of the material. These results are in agreement with prior experiments conducted for the chemical modified and unmodified cages in section 5.1.5. In the following section the capability of these variants to adsorb PBUTs is investigated

5.2.5 Adsorption capacity of redesigned Ftn^(neg) variants

The same assay established in section 5.1.6.1 was applied to determine the adsorption capacity of the redesigned ferritin variants. First crystalline and non-crystalline materials based on ferritin variants with an increased density of hydrophobic sidechains in the inner cavity and the pore region were tested, followed by materials build from ferritin variants designed with ligand-docking protocols. Finally, the recyclability of a selected variant was investigated.

5.2.5.1 Ferritin variants with hydrophobic sidechains

The effect of the inserted mutations for the protein variants designed in section 5.2.1.1 on the adsorption of IS are summarized in Figure 5.44a. Focusing on the crystalline material (blue), slightly increased capacities could be observed for the Ap4 and Ap7 variant in terms of IS. The Ap16 variant with the most redesigned amino acids showed no increase in comparison to the unmodified Ftn^(neg). On the other hand, in terms of the non-crystalline material, a significant increase in adsorption capacity could be observed for the Ap16 variant. Also, the Ap4 variant reached similar values. Determined capacities for IS were 672 ± 39 for Ap7 and $686 \pm 2 \mu\text{g g}^{-1}$ for Ap16 and therefore twice as high as for unmodified Ftn^(neg), which only reached $327 \pm 55 \mu\text{g g}^{-1}$. In the light of this results, it seems obvious that the amino acid exchange had a significant effect on the PBTs adsorption. This strong effect was only visible for the non-crystalline material. For Ap7, no difference in adsorption capacity between its crystalline and non-crystalline form could be measured. This is in contrast to the results found for the two other variants and cannot be explained. Even for Ap7, the adsorption capacity was increased in comparison to the unmodified cage. It appears that the mutations have a positive effect on adsorption capacity since crystalline and non-crystalline material of each variant showed increased adsorption capacity. Because the determined values for Ap4 and Ap7 in crystalline form and Ap4 and Ap16 in non-crystalline form were quite similar, it can be assumed that no significant difference is present between the Ap variants. The design of Ap7 and Ap16 is based on Ap4 and only expanded by the introduction of more hydrophobic amino acids at surface accessible positions as discussed in section 5.2.1.1 and seen in the sequence alignment in Figure 5.26. This led to the assumption that the mutations of Ap4 had the greatest influence on the increased IS binding capacity. These mutations specifically targeted negative charged amino acids and lead to a decrease in negative surface potential as shown in Figure 5.27. This mutation seems to be very beneficial for the overall adsorption capacity, while the further introduction of amino acids with hydrophobic sidechains have only a minimal effect.

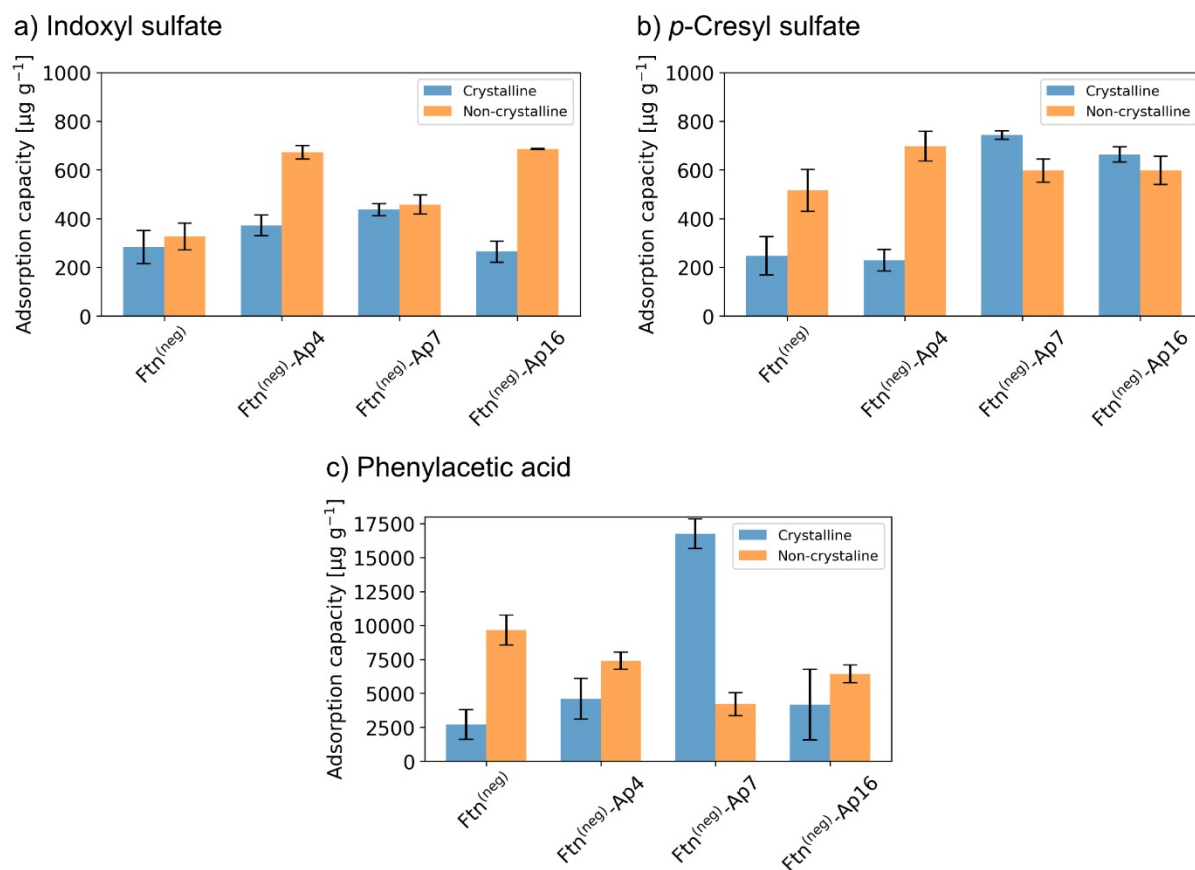


Figure 5.44: Adsorption capacity for crystalline and non-crystalline materials based on Ftn^(neg)-Ap variants.

The adsorption capacity for crystalline and non-crystalline adsorbent based on ferritin variants with increased density of hydrophobic side chains is shown for a) IS at 44 mg L^{-1} , b) *p*CS at 41 mg mL^{-1} and c) PAA at 474 mg mL^{-1} .

The adsorption capacity of the non-crystalline materials towards *p*CS shows a similar trend as seen in Figure 5.44b. This finding indicates a similar effect of the mutations. The difference between crystalline and non-crystalline adsorbent was with only 15% not so well pronounced as for IS. Interestingly for the crystalline material, high adsorption capacities were found for Ftn^(neg)-Ap7 and Ap16 reaching values of up to $743 \mu\text{g g}^{-1}$, which is nearly 3 times as high as the capacity of Ftn^(neg). Additionally, in the case of these two variants, the capacity of the crystalline material was higher than the non-crystalline material, which is in contrast to the results of Ap4 and Ftn^(neg) and to the results for IS adsorption. At least the difference to the IS adsorption could be probably explained with the smaller size of the *p*CS molecule, which could possibly diffuse better through the pores of the crystal than the larger IS molecule. However, this doesn't not explain the difference to Ap4 and the unmodified cage. Similar trends could be already observed for the adsorption of *p*CS to the chemical modified cages in Figure 5.22b.

In contrast to the results concerning IS and *p*CS adsorption, for PAA, the non-crystalline adsorbent showed decreased adsorption capacity in comparison to Ftn^(neg) as seen in Figure

5.44c. Crystals of Ap7 showed extremely high adsorption capacity. However, because the attributed measurement was performed with a very low amount of sample and there is a large difference to all other values, it is most likely that this is based on a measurement error. The highest determined adsorption capacity for PAA could be found for the non-crystalline material from the unmodified cage with a value of $9659 \pm 1100 \mu\text{g g}^{-1}$. Why the negatively charged PAA would prefer the more negatively charged surface of Ftn^(neg) is not clear, but since the values of the crystalline materials showed the trend to improved values, a similar trend could be expected here. The flaw of the non-crystalline material is that its properties such as surface area and pore volume could not be characterized during the time of this thesis. Suggestions for future research in this direction are summarized in section 7. It is possible that some unknown factors led to a different material structure in the Ftn^(neg) non-crystalline adsorbent. For comparison between variants, the highly-ordered and defined material could be more suitable. The crystalline material showed slight improvement of adsorption in case of Ap4 and Ap16, but the attributed error is large and the result is not very significant. It can be concluded that at least for the crystalline material some slight but not very significant improvement in adsorption capacity could be observed. However, the influence of the mutations is small and further characterization with respect to the non-crystalline materials is needed.

The designed variants showed in most cases an increase in adsorption. Interestingly no great difference in capacity between the Ap variants could be found indicating that the four mutations they have in common were mainly responsible for the improvement. In conclusion, amino acid exchange targeting the negatively charged amino acids and decreasing the negative surface potential of the inner cavity showed an improvement in adsorption capacity. Additional mutations increasing the density of hydrophobic side chains showed no further improvements, which agrees with results of chemical modification in section 5.1.6.2. As a conclusion on general design criteria, it can be said that the decrease in negative charge density has a greater influence on the capacity than the further addition of hydrophobic amino acids. The four mutations targeting the negatively charged amino acids mainly change the surface potential near the 3-fold channel (Figure 5.27), it can be hypothesized that this benefits the overall mass transport in the inner cavity and is responsible for the main improvements. Further improvement of the mass transport achievable by expanding the size of the 3-fold channel is investigated in the following section.

5.2.5.2 Ferritin variants with expanded pores

The mass transport inside the inner cavity of the protein is governed by the 3-fold channel. Strategies to change its size and polarity by exchanging sterically demanding and negatively charged amino acids were explored in section 5.2.1.3. Adsorption capacities for respective materials are shown in Figure 5.45a.

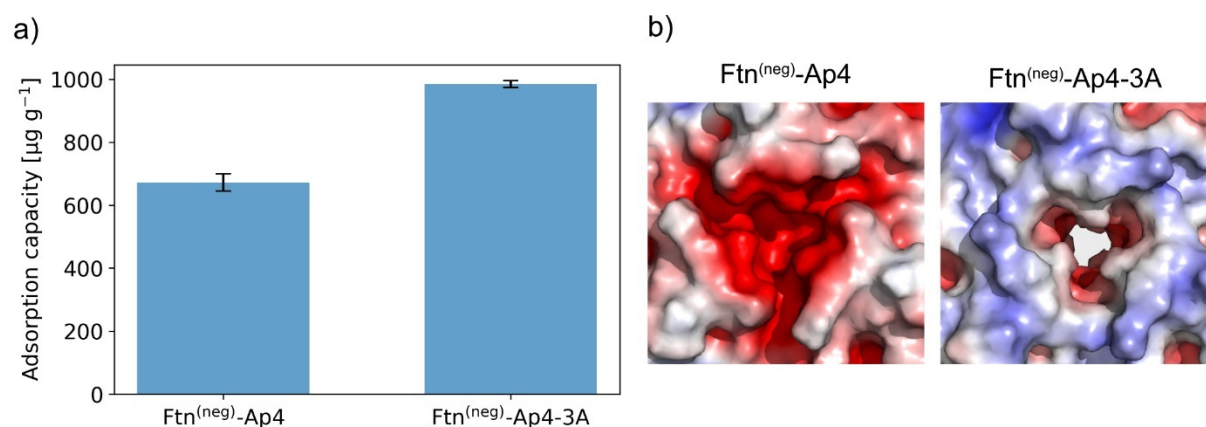


Figure 5.45: Adsorption capacity of $\text{Ftn}^{(\text{neg})}\text{-Ap4}$ variant with expanded pore. a) Indoxyl sulfate adsorption capacity at a concentration of 44 mg L^{-1} of an $\text{Ftn}^{(\text{neg})}\text{-Ap4}$ variant with 3 amino acids exchanged to alanine at the 3-fold channel is compared to similar with native pore. b) SASA and surface potential at 3-fold channel for both variants.

The mutations at the pore led to a significant improvement in adsorption capacity of 46% in comparison to the variant with native pores. The variant reached an adsorption capacity of $985 \mu\text{g g}^{-1}$. This highlights the importance of mass transport towards the inner cavity. Due to the symmetry at the channel, three mutations per subunit are adding up to nine mutations at the pore. The expansion and change in surface charge can be clearly seen in Figure 5.45b. The mutations are mainly located at the outer part of the pore, and it is visible that the inner part still bears a sizable negative charge. This could be completely overridden by introducing more mutations to the 3-fold channel, leading to a complete positively charged pore like seen in Figure 5.33 in section 5.2.1.3. However, it should be kept in mind that the strong charge could bind the toxins in the pore blocking the entrance for further toxins.

The pore mutations led to a sizable increase in adsorption capacity. This result shows the importance of entry in the inner cavity for the overall adsorption. Combination of mutations expanding the pore and decreasing the surface charge in its proximity are promising design features to increase the mass transport inside the cavity. However, after entering the cage the proteins still need to be bound by some sort of affinity. In the following part it will be investigated if designed toxin binding sites are a possible strategy to increase the affinity.

5.2.5.3 Ferritin variants with toxin binding sites

In section 5.2.2.2, ferritin variants with binding sites for the toxin IS were designed using a ligand-docking algorithm. The adsorption capacity of these variants in crystalline form towards IS as well as for the other toxins are summarized in Figure 5.46. The crystalline material is chosen since the highly ordered defined material is better suited for comparing changes between mutated protein. For IS adsorption shown in Figure 5.46a, a strong increase in binding affinity could be observed for the redesigned variants. The highest value was reached by the variant dock03. The capacity reached by this variant was with $521 \mu\text{g g}^{-1}$ twice as high as the unmodified cage ($247 \mu\text{g g}^{-1}$). However, variant dock23 and dock43 also showed sizable improvement in adsorption capacity with an increase of 47% or 86%. This result show that the implementation of ligand-docking algorithm increasing the PBUT affinity of the protein is a promising strategy.

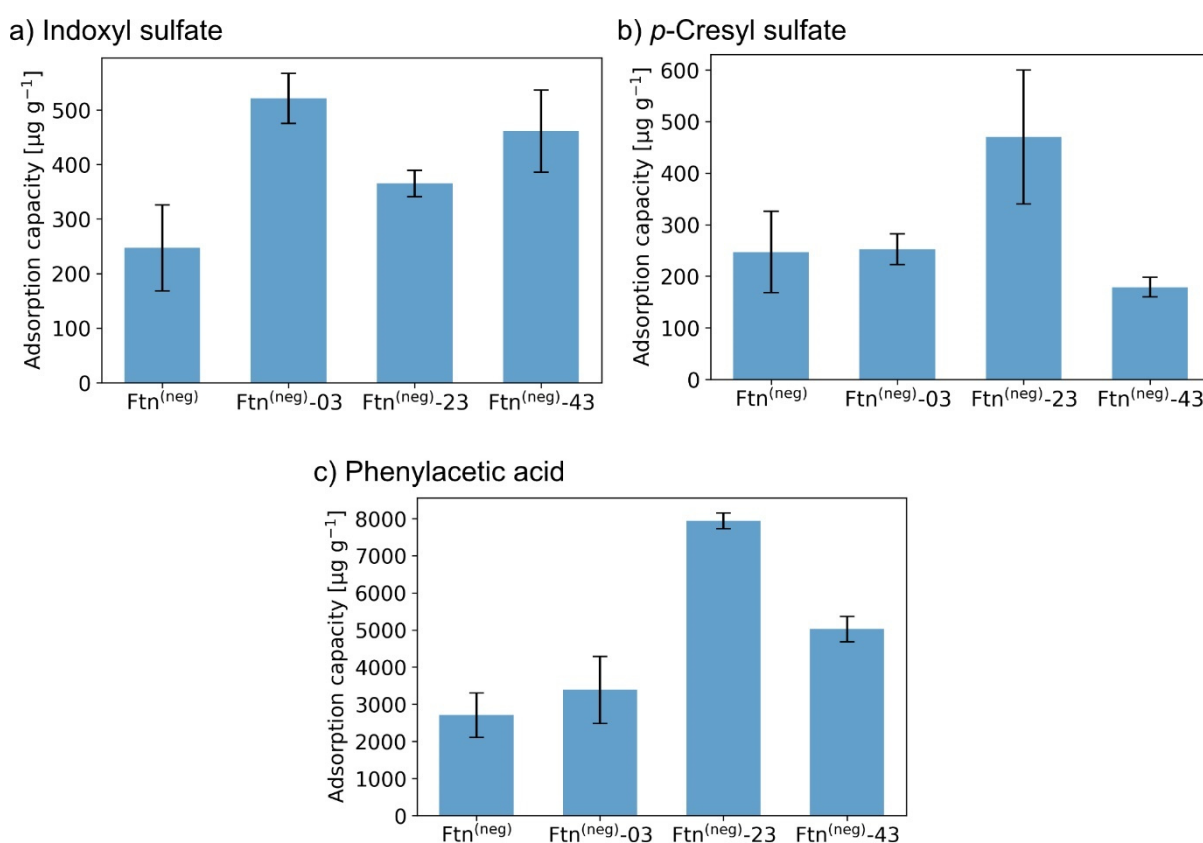


Figure 5.46: Adsorption capacity of crystalline material based on $\text{Ftn}^{(\text{neg})}$ variants with toxin binding sites. Adsorption capacity for crystalline material based on ferritin designed with ligand-docking algorithm for a) IS at 44 mg L^{-1} , b) *p*CS at 41 mg L^{-1} and c) 474 mg L^{-1} .

The $dG_{\text{separated}}$ from the models suggest that the binding of dock03 and dock23 with values around -14 are stronger, while dock-43 with a value around -11 is slightly weaker (Figure 5.29). While this holds true for the dock03, variant the dock43 variant shows higher capacities than the dock23. This showed that the $dG_{\text{separated}}$ value needs to be more interpreted as an

estimate and chemical intuition can still play an important role: The more positive surface potential of the dock03 and dock43 variants (Figure 5.29) and the toxins chemical environment in the binding site (Figure 5.30) looks much more promising for the dock03 and dock43 variants. The variants were originally designed for the PBUT indoxyl sulfate but due to the chemical similarities between the PBUTs, the adsorption capacity towards the other toxins was further investigated. For the PBUT *pCS*, only the variant dock23 showed a significant improvement in adsorption capacity while the other variants showed no improvement in comparison to unmodified Ftn^(neg) as seen in Figure 5.46b. This is quite interesting since variant dock03 and dock43 performed very well for IS. Application of the ligand-docking protocols seems cause some sort of selectivity towards the specific toxin. The reason to this could lie in the partial positively charged nitrogen atom present in the ring of the IS which is not present in *pCS*. The variants dock03 and dock43 both are stabilizing this atom by a hydrogen bond with a carbonyl or carboxylic acid group of Glu128 or Glu150 (Figure 5.30). If the negatively charged sulfate group of the *pCS* binds at the same amino acids as the one of the IS, respective polar functional groups would interfere with the nonpolar ring system of the *pCS* leading to an unfavorable binding. This stabilization of the partial positive nitrogen atom is lacking at the dock23 binding site, explaining why it performs better for *pCS* bearing a phenyl ring instead of an indole ring. The same explanation can also be applied on the adsorption capacity of the docking variants towards PAA shown in Figure 5.46c. The capacity of dock23 reached a value of 7941 $\mu\text{g g}^{-1}$ and is nearly three times higher than the capacity of the unmodified cage with 2710 $\mu\text{g g}^{-1}$. The dock43 variant also showed an improvement of 85% towards PAA. The PAA adsorption was performed at higher concentration so probably the presence of the positive polar patches (Figure 5.29) benefits the adsorption at this concentration. However, in this case also the dock03 variant should show increased values which is not the case. The increased capacity of dock43 cannot be completely explained. This degree of selectivity is an indirect hint that binding of the toxins indeed proceed at the binding site and also in the geometries suggested by the simulation. Otherwise the selectivity for IS over *pCS* and PAA would not be so pronounced.

In conclusion, the design suggested by ligand-docking algorithm has improved the overall adsorption capacity. For the initial designed PBUT IS, a significant increase in binding affinity for all three designs could be found. Interestingly, dock23 performed the worst for IS but outperform the other variants for the other toxins. This results hint that the toxin is indeed bound at the designed site, since in dock03 and dock43 the positively charged nitrogen lacking in the other toxins is stabilized.

In prior experiments, the non-crystalline form of the adsorbent often outperformed the crystalline form. The adsorption capacity of selected docking variants in non-crystalline form were determined and respective results are summarized in Figure 5.47. For all PBUTs, an increase in adsorption capacity of the non-crystalline material with respect to the crystalline material was observed. For IS (Figure 5.47a), the increase was between 30% and 60% with the highest value for dock03 with $685 \mu\text{g g}^{-1}$. This capacity is linked to higher error but also the other docking variants reached high values around $600 \mu\text{g g}^{-1}$. Due to time reasons, not all variants could be measured for the other toxins. For *p*CS (Figure 5.47b) only dock23 and dock43 were measured in non-crystalline form. For dock43, the adsorption capacity increased 2-fold but reaches similar values as the unmodified cage. However, the adsorption capacity of the dock23 variant increased by 20% and reach a value of $570 \mu\text{g g}^{-1}$. For PAA (Figure 5.47c), the adsorption capacity of dock43 increased by 250% reaching a high value of $13225 \mu\text{g g}^{-1}$. The variant showed also an increased capacity in comparison to the $\text{Ftn}^{(\text{neg})}$. In the case of PAA non-crystalline material outperformed the crystalline material for every PBUT and variant.

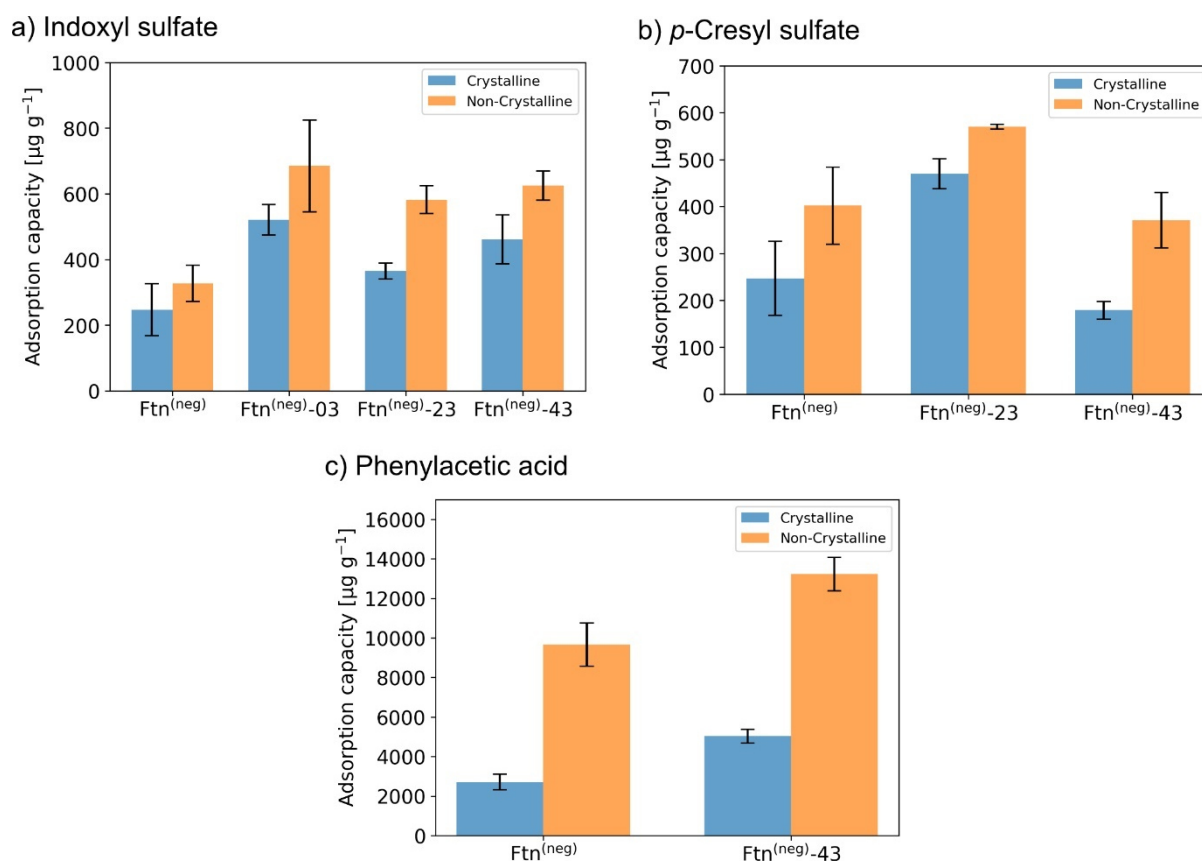


Figure 5.47: Comparison of adsorption capacity of crystalline and non-crystalline material based on $\text{Ftn}^{(\text{neg})}$ variants with toxin binding sites. Adsorption capacity for crystalline and non-crystalline material based on ferritin designed with ligand-docking algorithm for a) IS at 44 mg L^{-1} , b) *p*CS at 41 mg L^{-1} and c) 474 mg L^{-1} .

Also, the trends observed for the crystals in terms of adsorption capacity between the variants remained for all tested materials. If this hold true, the capacity of non-crystalline dock23 for PAA will be even higher than for dock43, but the required measurement was not possible in the timeframe of this work.

In conclusion, it was demonstrated that ligand-docking algorithm can introduce binding sites targeting specific PBUTs and increase the overall binding affinity for the toxins. The modular character of the protein redesign approach allows to combine the design of binding sites located in the cavity with the mutations increasing the mass transport to further optimize the adsorption. Results concerning the selectivity of dock03 and dock43 suggest that targeting unique features of the toxins enabling some sort of control over the binding strength. This may prove helpful to maximize the binding strength. However, complete dissociation of the binding sites under conditions the adsorbent tolerates would allow reuse of the material. This feature would drastically improve the cost-effectiveness of the protein-based adsorbent. Initial experiments investigating the recyclability of the protein-based adsorbent are summarized in the following section.

5.2.5.4 Recyclability

To reuse the protein-based adsorbent, conditions leading to a significant dissociation of toxin needs to be found. Studies towards displacement of PBUTs from proteins suggest that increasing ionic strength and pH weakens the binding to the protein.^[269] Another possibility is to treat the adsorbent with increased temperature. The higher temperature favor dissociation of the toxin due to the increase in entropy. In an initial test, non-crystalline adsorbent synthesized from Ftn^(neg)-dock43 was submitted to a standard IS adsorption assay followed by incubated in water at 60°C or buffer (50 mM Tris, 1 M NaCl, pH 9) over night. Afterwards, the toxin concentration in the supernatant was determined and compared to the adsorbed masses. The results are summarized in Figure 5.48a.

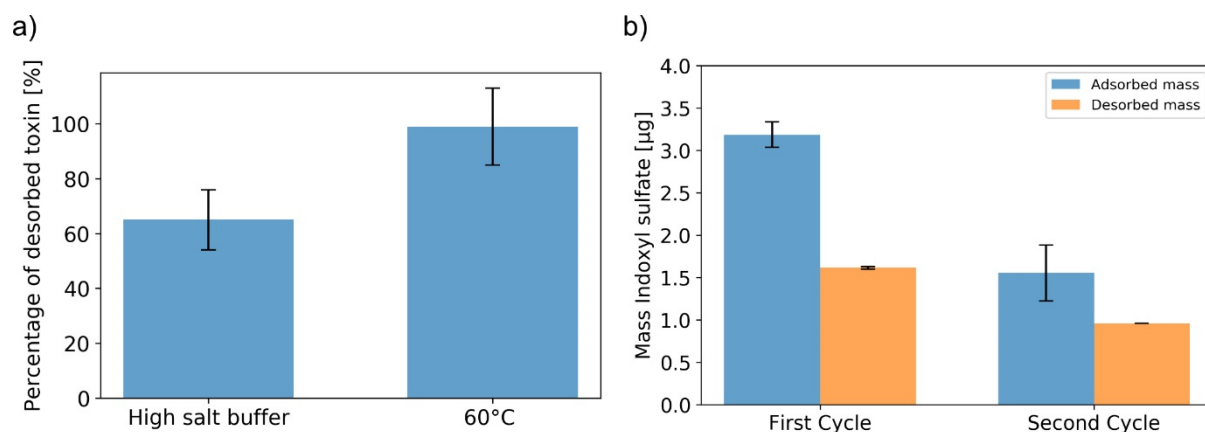


Figure 5.48: Recyclability of protein-based adsorbents. First indications for possible recyclability of the adsorbent. a) Desorbed indoxyl sulfate mass through treatment with high salt buffer (50 mM Tris; 1 M NaCl; pH 9) or heat. b) Two consecutive adsorption cycles with the same material and wash steps in between. IS concentration was set to 44 mg mL⁻¹

Both conditions indicate that it is possible to dissociate the binding between the toxin and the protein. The high pH and salt buffer removed around 65% of the toxin and simple treatment with 60°C hot water removed nearly 100% of the toxins. However, especially after the heat treatment, a color change of the material from white to yellowish could be observed. Probably the conditions were too harsh and the material decomposes. For first recycling test, mild washing condition by incubation in high salt buffer at 60°C for 20 minutes followed by two additional wash steps with the hot buffer were selected. Adsorbed masses for the first and second adsorption cycle as well as the toxin mass in the wash water after the first incubation step are shown in Figure 5.48b. Comparing the adsorbed masses between the first and second cycle revealed a great loss in adsorption capacity. However, the first wash step roughly dissociated half of the adsorbed toxin mass. Interestingly, the amount of removed toxin in the first cycle and adsorbed toxin in the second cycle is roughly equal. The adsorption seems to be reversible, but the washing was not sufficient to displace all bound toxin. In combination with the results for the longer washing steps in Figure 5.48, the results indicate that a complete recyclability could be possible. The main challenge to achieve this is finding a condition capable of removing 100% of the toxin without damaging the protein-based material. Further experiments in this direction should be made, because recyclability will drastically reduce the attributed cost in clinical applications.

The results gathered in the aforementioned sections highlight the advantage computational methods offer to the development of protein-based materials. Understanding interactions of protein cages with themselves could help to facilitate more open morphologies, beneficial for the PBUT adsorption. Understanding interactions with the cage and other proteins might also

prove helpful in decreasing the BSA adsorption on the material. Out of this reason, protein-protein docking protocols are investigated in the following section.

5.3 Protein-Protein docking for interfaces in binary protein crystals

In earlier works in the *Beck group*, a binary protein crystal formed by a negatively and positively supercharged protein cage was developed.^[89] Respective protein cages were designed by the exchange of surface amino acids to charged residues. Both variants contain several mutations, but only one directly located at the contact area of the cages. The question arises if the main driving forces for formation of the binary crystal arises from the resulting charge complementary or from the new interaction at the interfaces. In unpublished work by *Laurin Lang*, mutations at the surface of the supercharged cages were reverted and the effect on the crystallization is evaluated. To support this investigation, protein-protein docking protocols implemented in the Rosetta software suite were applied to this system. The aim is to quantify the interaction energy for different mutated ferritin variants at the contacts areas in the crystal.

Two contact areas (A1 and A2) exist in the crystal structure between the negatively supercharged variant Ftn^(neg) and the positive supercharged variant Ftn^(pos) (Figure 5.50b and Figure 5.51b). The protein-protein docking algorithm was first established for the interface A1 between the fully supercharged variants Ftn^(pos) and Ftn^(neg). Modeling two complete cages would be computational too expensive. Therefore, only two subunits per cage directly located at the interface were used. Prior to the docking, the energy of the protein structure has been minimized. Three models with the lowest energy were suspected to a protein-protein docking algorithm. 1000 trajectories with the protein-protein docking algorithm followed by the docking-analysis algorithm were performed. The complete procedure is described in detail in section 8.8.3.2. In short, the protocol started with a low-resolution docking step. Here the sidechains were abstracted as centroids and the subunits of one container were translated to find an initial suitable binding area. Only residue-pair and van der Waals interactions were calculated during this step. Next, high resolution docking with complete side chains and the full-atom energy function was performed to further refine the binding of the cages. Finally, a side chain energy minimization was performed. Besides this full docking procedure, 10 additional trajectories were modeled, omitting the relative movement of the cages during low- and high-resolution docking and only optimizing the interface energy by perturbation of side-chain conformation. Respective models are serving as a reference. The output was analyzed with the InterfaceAnalyzer mover previously used for protein-ligand docking. The dG_{separated} value is a measure for binding energy. Additionally, the C_α-RMSD with between the model and the crystal structure of the binary lattice are determined. The total score as a measure of overall protein stability of the minimized structures is plotted together with the ones from the complete docking against the RMSD (Figure 5.49a). Similarly, the dG_{separated} values as a measure of binding energy is plotted (Figure 5.49b).

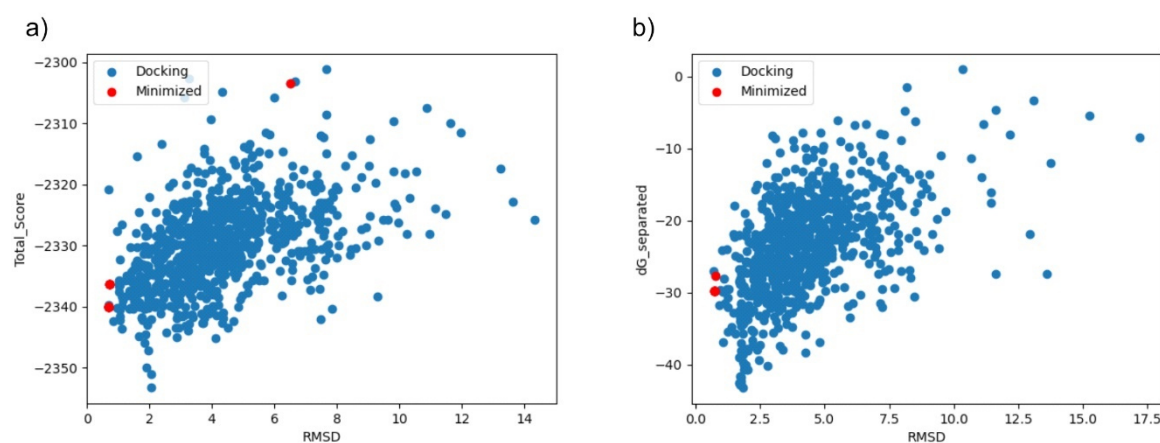


Figure 5.49: Protein-protein docking for interface A1 of Ftn^(neg) and Ftn^(pos). Protein-protein docking was performed and dG_{separated}, total_{score} and RMSD with respect to original crystal structure are determined. As a reference, models omitting the movement of the protein cages relative to each other and only with minimized interface energy are generated (red). a) total_{score} as measure of protein-stability is plotted against the RMSD b) dG_{separated} as measure for binding energy is plotted against RMSD in Å.

The scatter plots show a funnel-like structure, where at the “tip” of the funnel the lowest energies are found at an RMSD around 2 Å. The protein stability and interface energy for the minimized models are in most cases at the lowest RMSDs reaching medium values in dG_{separated}, which is expected since no relative movement between the cage subunits is performed. Therefore, the similarity to the situation in the crystal structure is the highest. The minimized values show some outliers, because of the stochastic nature of the underlying Monte Carlo algorithm. It is hypothesized that the lowest values found in the “binding-funnel” in Figure 5.49b represents the structure the proteins would possess in solution.

To investigate the influence of different mutations on the interface energy, the established docking protocol was applied to various combinations of ferritin variants with different amino acid composition. The investigated variants can be broadly divided in one group with reverted interface mutations and one without. The complete list of the used variants and at which positions they differ from respective supercharged variant is given in Table 10.7. Per interface, 12 different combinations are possible and the docking protocol was performed for each of them. Respective scatter plots are shown in section 10.4 in the appendix. The mean dG_{separated} value for the best 5 structures for each combination was determined and sorted from the highest to the lowest. Except one outlier in the crystal contact A1, the combinations with no interface mutations have the highest dG_{separated} value as seen in Figure 5.50a highlighted in orange. This indicates that the lysine in Ftn^(pos) and glutamic acid in Ftn^(neg) increase the binding affinity between the cages and promote the formation of the binary protein lattice. The differences between the interface energies are quite small. However, in the later

protein crystal contact area A1 is present four times per cage. Since thousands of cages are involved in the crystal formation, even small changes can add up to a great difference in free energy. It is also interesting that the combinations with reverted mutations not located at the interface have more favorable interface energies than the complete supercharged protein cages. Probably these mutations interfere somehow with the interface despite the distance. This could possibly be explained by the long-range character of electrostatic interactions, which could be altered by the removal or addition of the charged amino acids.

Contact area A1

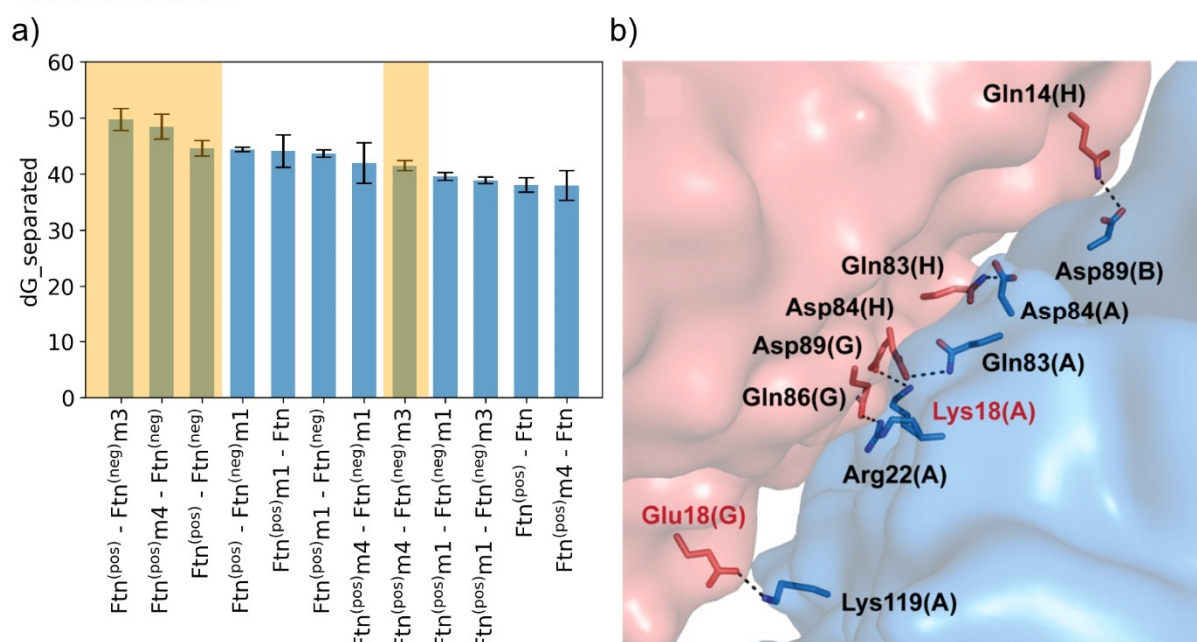


Figure 5.50: Interface energy in binary protein crystal. a) Interface energy as measured by $dG_{\text{separated}}$ from highest to lowest value for the contact area A1. Combinations containing no mutations in the interface are highlighted in orange. b) Residues present in the interface. The mutated residue is highlighted in red. (Images of crystal contacts are adopted from reference 89 with permission from the American Chemical Society)

For contact area A2 (Figure 5.51a) no real trend can be observed. Since only the variant Ftn^(pos)-m1 contains a mutation, which is present in the interface, only 4 of the 12 combinations have interfaces which are affected by the changes. Attributed $dG_{\text{separated}}$ values are either among the highest or lowest values. It appears that the changes to the interface have no great influence on its energy. This appears unreasonable due to the fact that the interface mutation undergoes three interactions with residues from the neighboring cage (Figure 5.51b). However, this could be also a hint that the polar interactions have a much lower influence on the interface energy than expected. The difference in $dG_{\text{separated}}$ is small and the error

attributed to the values large, rendering this finding not very significant. The observations could also be completely random.

Contact area A2

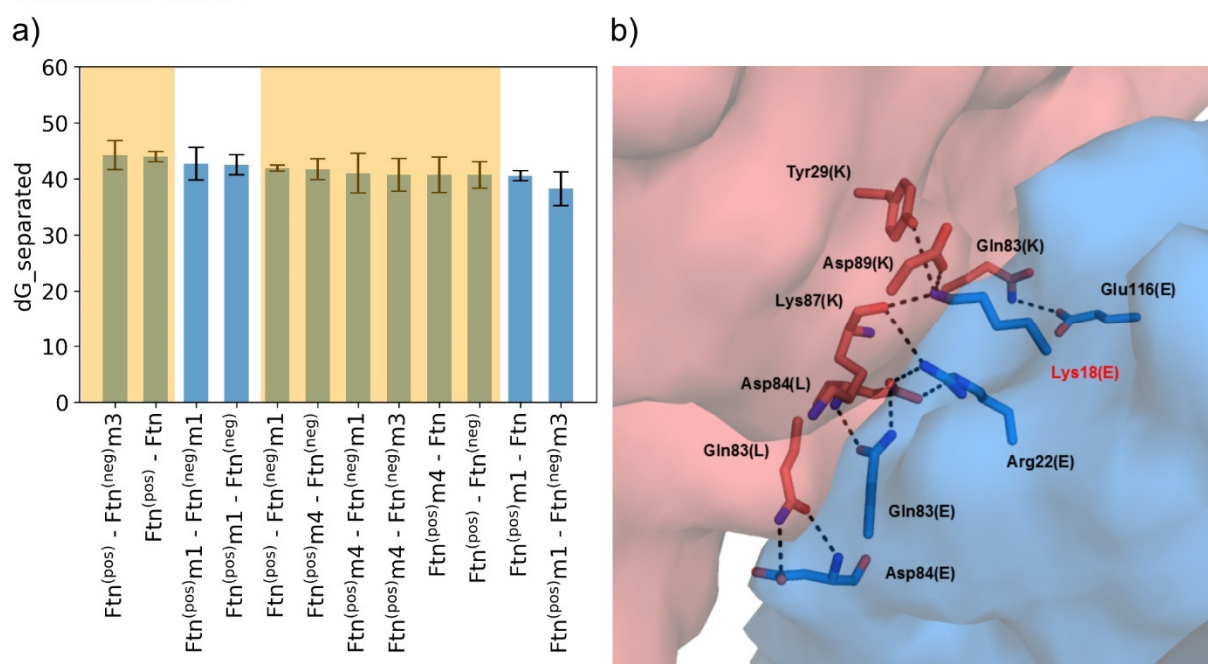


Figure 5.51: Interface energy in binary protein crystal. a) Interface energy as measured by $dG_{\text{separated}}$ from highest to lowest value for the contact area A2. Combinations of variants with no reverted mutation at the interface are highlighted in the bar diagram in orange b) Residues present in the interface are illustrated, where the mutated residue is highlighted in red. (Images of crystal contacts are adopted from reference 89 with permission from the American Chemical Society)

In summary, a protein-protein docking protocol could be established and applied to interfaces in a binary ferritin crystal. First results showed some reasonable trends for contact area A1. Taking the mean value for the best output structures aimed to mitigate the effect of outliers was helpful to get a quick overview. Combinations with a rather large error bar like for example Ftn^(pos)-m1 and Ftn^(neg)-m1 also do not show a pronounced “binding funnel”. Apparently, the docking algorithm failed to find a local minimum in energy. This can either be caused by too less sampling or the fact that no real minima exist. The observed interface is quite large and probably in some cases the total generated 3000 models per combination and contact area are not sufficient for finding a global energy minimum. Scatter plots for all combinations should be reviewed and combinations not showing a clear funnel-like structure should be repeated to get more reliable results. There is also a variety of available protein-protein docking software,^[270] which can be used to verify the results derived by the established method. In some scatter plots the “binding funnel” can be observed at higher RMSDs. Since the docking were only performed with a fraction of the cage and only for one interface at a time, it is possible

that the interfaces adopt unreasonable geometries. For future simulations, constraints should be added, limiting the relative movement of the cages relative to each other. It is advisable that the generated output structures are visual inspected to further evaluate the side chain interactions at the interface. Comparison of the model to the crystal structures derived from *in vitro* experiments produced protein variants could further help to evaluate the model and serve as a benchmark set to further optimize the docking of protein cages. Right now, the protein-protein docking was only used for analyzing the interaction between the cages. Similar to the ligand-docking protocols, the algorithm could be expanded by a design step to alter the interactions between the cages. The two contact areas A1 and A2 have some overlapping residues, but they also have unique residues. By weakening the interactions of amino acids present in both areas and developing two sets of amino acids with high affinity at unique positions, three ferritin variants with different affinity to each other could be fabricated. This could open the route to a ternary crystal.

6 Summary and conclusion

The main objective of this thesis was to investigate the general feasibility and effectivity of a protein-based material for PBUT clearance as full-blood contact adsorbent. Fabrication of a solid material based on the variants of the protein cage ferritin could be demonstrated in either crystalline (Figure 5.11, Figure 5.12) or non-crystalline (Figure 5.15) form. Chemical fixation was used to stabilize the produced materials and increase their stability in conditions similar to the blood (Figure 5.14). The absence of endotoxin contaminations and platelet activation could be confirmed for all investigated material forms (Figure 5.16, Figure 5.17, Figure 5.18 and Figure 5.43). Clinical applications of the material appear to be possible, but more elaborated studies in terms of hemocompatibility should be conducted to confirm the applicability as full-blood contact adsorbent. Also, adsorption of plasma proteins, observed for BSA in initial studies, should be further investigated. To determine the effectivity in terms of PBUT clearance, an assay to determine the adsorption capacity of the materials was established in section 5.1.6.1. Three PBUTs were selected as model systems, namely the indoxyl sulfate (IS), *p*-cresyl sulfate (*p*CS) and phenylacetic acid (PAA). Materials based on unmodified protein cages already showed affinity towards PBUTs (Figure 5.23). The main part of this thesis is focused on increasing the binding affinity of the adsorbents by modifications to the inner cavity of the protein cage. Two strategies were developed and tested, namely the introduction of guest molecules with an affinity towards the PBUTs and protein redesign guided by computational methods.

For the introduction of guest-molecules, ferritin variants containing up to four reactive thiol groups per subunit located exclusively on the inner surface were designed (section 5.1.1), produced and purified (section 5.1.2). Presence of the mutations is confirmed by ESI-MS (Table 10.3). Results from SEC and TEM suggest an intact cage architecture (Figure 5.2). Protocols capable of coupling hydrophobic molecules bearing a phenyl group or a long aliphatic chain were developed (section 5.2.3). Loading of up to 96 molecules per cage (Table 5.2 und Table 5.3) without affecting the cage stability and morphology (Figure 5.6 and Figure 5.7) was demonstrated. Since the modifications are located at the inner cavity, fabrication of solid materials proceeds under the same conditions as the unmodified material (section 5.1.4). Additionally, no change in biocompatibility can be observed (Figure 5.16, Figure 5.17). This result highlights the great advantage of the protein cage. Since the inner and outer surfaces are strictly separated by size-restrictive pores, chemical modifications on the inner surface have little effect on the higher-order assembly of the cages or the interaction with its environment. This allows to optimize these properties independent of each other, which is helpful for further material development. The adsorption capacity of the materials made from

chemical modified cages show for all PBUTs a slight increase in adsorption capacity in comparison to the unmodified cage. However, this change is small and not very significant (Figure 5.22). This result allows the conclusion that the affinity of the PBUTs to purely hydrophobic molecules is, low most likely since the toxins contain a negatively charged polar group. Efficient adsorbents needed to stabilize both characteristics. The problem may be overcome by the introduction of more suitable ligands as discussed in (section 5.1.6.2). Further, the ferritin-cysteine system is extremely variable as proven by the encapsulation of various organic dyes (Figure 5.24).

The other investigated strategy embodied redesign of the protein cages inner cavity to increase its affinity towards PBUTs. Specific properties of the cage could be altered by a combination of rational design and the fixbb protocol of the Rosetta software suite, which found possible mutation sites for the desired amino acid without affecting the proteins stability (section 5.2.1.1). This technique was carried out to design variants with reduced negative surface potential (Ap4), higher density of amino acids with hydrophobic side chains (Ap7 and Ap16) or expanded pores (Ap4-3A). Results from adsorption capacity experiments of chemical modified cages suggest that a combination of polar, positively charged and hydrophobic interactions is needed for effective PBUT binding. To create binding sites to satisfy these requirements, ligand-docking protocols were applied (section 5.2.1.2). With the exclusion of one design bearing a deep pocket for efficient adsorption and one variant with multiple binding sites, all ferritin variants could be produced and purified by established protocols. Presence of mutations were confirmed by ESI-MS and omit maps. Intact cage architecture was verified by SEC and TEM images (section 5.2.2). All mutations were exclusively located on the inner surface. In some case changes to the outer surface charge could be observed in IEC (Figure 10.17, Figure 10.19). However, no significant changes to higher-order assembly were observed. The variants can be assembled to crystalline and non-crystalline materials by similar techniques as the unmodified cage. Furthermore, biocompatibility assays show similar results (section 5.2.4). These results indicate that changes of the inner cavity do not alter properties like higher-order assembly or biocompatibility governed by the outer surface. Some variants show significant higher adsorption capacities compared with the original protein cage. For IS adsorption, the highest capacity of $985 \mu\text{g g}^{-1}$ could be reached for the variant $\text{Ftn}^{(\text{neg})}\text{-Ap4-3A}$. It has to be said that this is the only variant with expanded pores investigated. However, the variant showed 43% better performance compared to controls lacking the channel mutation (Figure 5.45). This is a strong evidence that mass transport through the pores plays a key role. Expansion of the pore appears to be a promising strategy to increase the overall adsorption capacity. For mutations exclusively on the inner surface, the capacities for Ap4, Ap16 and dock03 all in non-crystalline form are nearly on the same level. This is particularly interesting

for Ap4 and Ap16 (discussed in section 5.2.5.1) since Ap4 contains four mutations and Ap16 the same mutations plus 12 more. In comparison to the unchanged cage, these four mutations doubled the adsorption capacity. The 12 additional mutations adding further hydrophobic amino acids at surface exposed positions lead only to a further increase of 2%. Increased density of only hydrophobic amino acids appears to not favor the adsorption of the PBUT which agrees with prior results with the chemically conjugated hydrophobic molecules. The impact of only four mutations could be explained by the decrease in negatively charged surface area near the 3-fold channel. In the native container, this area has a negative surface potential and guides iron ions to the ferroxidase site. It is possible that the negative charge hinders the diffusion of the negatively charged toxins in the cavity. Amino acid exchanges in Ap4 removing this surface charge increased the mass transport (Figure 5.27). Since the Ftn^(neg)-Ap4-3A mutant possessing only mutations in entry area of the cage showed by far the best capacities, changes to this area seems to have a great impact on the overall adsorption. This should be considered for future designs in the simplest way by adding the four Ap4 mutations and the three alanine residues at the threefold channel to other variants. The variants designed with ligand docking-algorithm reach high adsorption capacities around 600 $\mu\text{g mL}^{-1}$. The variants dock03 and dock43 target a unique feature of the IS toxin. Adsorption capacities towards IS of these variants were doubled in comparisons to the unmodified cage, but no changes could be observed for *pCS*. This selectivity hints that the toxin binding really proceeds directly at the binding sites. The variants designed with ligand docking-algorithm still possesses negative potential in the entry area, so significant increase in adsorption capacity could be reached by adding the Ap4-3A mutation to the docking variants.

Conclusion on the impact of the higher-order assembly on the adsorption can be drawn from the experiments using material in crystalline and non-crystalline form. In the majority of the experiments, the non-crystalline material has outperformed respective crystalline material. This is especially true for the adsorption of PAA, where capacities were increased by a 3.5-fold. However, some results especially for *pCS* showed contradictory behavior, but important information about the material are lacking. Basically, the non-crystalline material is precipitated protein, which could not be completely characterized during the time of this thesis. For a more in-depth understanding of the previous results, the surface area and pore-size distribution in the respective material are crucial. Suggestions for suitable experiments will be given in the outlook. However, the results indicate that the higher-order assembly can have a great impact on the PBUT clearance, possibly by improving the accessibility of the protein cages. Since protein cages in the crystal adopt the closest packing, the assumption that the non-crystalline material has a more open structure should be confirmed. With this assumption made, the increased adsorption capacity in terms of PAA could be explained with the easier diffusion of

the toxins through the non-crystalline material. In the crystalline material, the channels between the protein cages are small (around 30 Å, Figure 10.11) and diffusion to cages located in the inside could be limited. The non-crystalline material will most likely have larger pores allowing the access of more protein cages. This effect should become more pronounced with increasing concentration of the toxin, explaining why the improvements were the highest in PAA. This result demonstrates that for further improvement of the adsorbent, diffusion through the material needs to be considered. In terms of the crystalline material, this could be achieved by decreasing the crystal size. In smaller crystals, the distance between the cages inside to the outer solvent is smaller and diffusion to them is easier. Further optimization ideas for the higher order assembly are discussed in the outlook.

In comparison to each other, the protein redesign strategy clearly outperformed the chemical modification route in terms of adsorption capacity. Performance of the chemical modified variant could probably be increased if more suitable ligand systems could be found. However, in terms of future applications, cost-effectiveness is an important factor allowing more patients access to the therapy. Since for the redesigned variants, further functionalization after expression and purification is not needed, they clearly present the more cost-effective option. Out of this reason, at least for clearance of PBUT future research should focus on the redesign strategy. However, cysteine-containing ferritin variants could be applied in other applications. In this thesis, possible applications in the field of bioimaging were already explored.

In terms of PBUT adsorption the conducted research identifies three possible routes for further optimization.

1. Changing the higher-order assembly of the cage to obtain more accessible morphologies could improve diffusion of the PBUTs towards the cage pores.
2. Entry in the cage has a great impact on overall adsorption and can be improved by changing the polarity around the pores or expand their size.
3. The hydrophobic and polar (negatively charged) functional groups of the toxins need to be stabilized to reach high adsorption capacities. Ligand-docking protocols may help designing suitable motifs.

In conclusion, it can be said that the fabrication of a protein-based PBUT is a promising strategy. While it should be possible to make this material from human serum albumin or plasma proteins, which natively bind the toxins, protein cages are more advantageous building blocks. The cage-like architecture results in an inherent high surface area. Since the binding sites are located in the inner cavity, assembly will not lead to blocking of these sides. Modifications to the binding sites will also not affect the biocompatibility or assembly. The

adsorption capacity could be increased to values of nearly 1 mg g^{-1} (Ftn^(neg)-Ap4-3A: 985 mg g^{-1}) which is already in the range of other published adsorbents like zeolites (1 mg g^{-1})^[49] or carbon based adsorbents ($3.2 - 3.7 \text{ mg g}^{-1}$).^[50] Further the results so far indicate that mutations for the binding sites and the increased mass transport can be combined to increase the adsorption capacity. Also, theoretical considerations assume, that high adsorption capacities around 10 mg g^{-1} (Equation (8)) are already achievable, when only one toxin is bound per ferritin subunit. Investigations in terms of hemocompatibility are promising and it can be expected that the adsorbent can be brought in contact with the blood without harmful side effects. Therefore, easy and cost-effective incorporation of the material in conventional dialysis methods is possible, allowing a large-scale treatment. In terms of production, the recombinant proteins can be produced from cheap, abundant and environmentally friendly base materials and the process can be scaled-up effortlessly. The purification of the proteins is a cost-intensive step, but due to its high thermostability, ferritin can be purified by a combination of heat precipitation and affinity chromatography. Also, first results in terms of simple recyclability of the adsorbent are promising. In the light of this results, ferritin-based adsorbents seem to possess great potential to become the cost-effective, hemocompatible adsorbent needed for large-scale treatment of dialysis patients. However, the material is still in an early development phase and more research needs to be conducted. Starting points for future research are presented in the following section.

7 Outlook

The non-crystalline materials performed well in the experiments done in this research project. It was hypothesized that this is due to their open structure but in fact a deeper characterization of the material is lacking. Analytic methods like gas adsorption techniques to determine the surface area and the pore structure of the material should be carried out. In an early experiment with crystalline material, no surface adsorption could be detected. Reasons for that may be sample amount or too harsh drying methods. The non-crystalline material can be produced in great quantities and more gentle drying methods like lyophilization could be applied to overcome this problem. If a suitable analytical method is established, investigations towards the relationship between adsorption capacity, surface area and pore structure can be conducted. Further, the influence on reaction conditions such as protein and crosslinker concentration on the surface area and pore structure could be investigated. The advantage of the crystalline material may be that it can achieve the highest density of protein cages per volume, but it appears diffusion inside the material limits its performance during adsorption. For an optimal adsorbent morphology, a mixture between crystalline and non-crystalline material could be an optimal solution. Experiments with PAA and the non-crystalline material as reference could be performed to find crystal sizes where the diffusion is not limiting the adsorption. It can be expected that the resulting crystals will be very small and not suitable for the later application. These crystallites could be subjected to further crosslinking agent and additional free cages to form a non-crystalline material with larger pores around them and achieve a possible high density of binding sites while enabling the accessibility of the cages by the toxins. For later application in dialysis machines, further properties will play a role. For example, a high back pressure can result if the material is too dense. However, since the material will be assembled from nano-sized building blocks, the variability of the assembly will allow to tailor the adsorbents properties according to the requirements. Further possible assembly strategies could be coating of the cages to a surface. The blood could be pumped across the coating, allowing maximal accessibility of each cage and guaranteeing a low back pressure. Intensive optimization and research in this direction need to be performed before effective application is possible.

Further optimization regarding the outer surface needs to be done concerning the adsorption of BSA to the material. This phenomenon should be further investigated. Even if the adsorption turns out to have no harmful effect on patients in later applications, the BSA can block the entrance to the material and therefore diminishing the overall toxin adsorption. Interaction between positive charged patches of the BSA with the negative charged container surface could be responsible for the binding. Understanding this interaction may be the first step to hinder

them. The established protein-protein docking protocols may prove helpful for finding interactions between the ferritins outer surface and the BSA. The algorithm used for “rastering” the inner surface of the Ap-variants to evaluate their binding affinity could be combined with the protein-protein docking protocol for a first overview.

Aforementioned suggestions focus on modifications to the outer surface. Independently to that, modifications to the inner surface can be performed to increase the adsorption in the cages. The experiments have highlighted the importance of the toxin entry in the cage. First experiments with expanded pores showed a significant increase in adsorption capacities. The results need to be verified. If they hold true, designs with even further expanded pores with positive surface potential were already presented and could be tested. The pore diameter could be further increased by deletion of amino acids near the pore region. It could be also possible to use ferritin variants already possessing large pores like ferritin from *Archaeon Archaeoglobus fulgidus* (PDB ID: 1SQ3). Mutations decreasing the negative surface potential near the pore entry proved to have a great effect on the adsorption capacity. It is also hypothesized that this is due to an improved entry to the cage. This could be tested by adding respective mutations to a variant with designed binding sites (dock variant), which should lead to a significant improved adsorption capacity. If this holds true, mutations in this area should be included in all future variants and possible optimization in this area for example by exchanging more charged amino acids could be investigated. The actual binding sites should be located deeper in the cavity to not block the toxins entry. The results of the research empathize that for efficient binding of the toxins, the hydrophobic and polar negative characteristics need to be targeted. Ligand-docking protocols have proven to be a valuable tool to design respective binding sites. In this experiment, the binding site design was completely done by the computational method and the resulting binding sites are quite large and efforts to combine them in one protein failed. For future designs, a more rational approach could be embodied with the aim to create a binding motif with low space requirement to allow multiple binding sites per subunit. The toxin binding site in the variant dock43 is a promising starting point for this development since it consists only of three amino acids located at the same chain. Experiments aiming to further assess its effectivity should be conducted. For optimization again, the developed ligand docking protocols can help to improve the binding sites. The dock43 subunit were again subjected to the ligand-docking algorithm, but an arginine was inserted manually near the binding site to stabilize the sulfate group by additional hydrogen bridges. The resulting design (Figure 10.28) showed an increase of $\Delta G_{\text{separated}}$ from -11 to -14. Further optimization *in silico* and *in vitro* with the aim to develop a small binding site with suitable affinity, which can be introduced at multiple positions of the subunit, should be conducted. The optimal design of the inner surface should include

expanded pores and a suitable entry area, accelerating the mass transport in the inner cavity. Inside the cavity, multiple binding sites combined from amino acids with hydrophobic and positively charged sidechains should be located.

Thanks to the strictly separated inner and outer surface of the protein cage, optimizations of the inner surface and the outer surface can be independently investigated. Beneficial features can be combined in a final design to yield an overall optimized adsorption material. The modular design of the ferritin-based adsorbent allows a great extent of variability. Studies demonstrated that adsorbents containing illuminated gold nanoparticle were capable of scavenging of free-radical and mitigating dialysis-induced oxidative stress.^[271] Encapsulation of gold nanoparticle is established for ferritin^[118] and respective loaded cages could be combined with empty cages opening the route for a multifunctional adsorbent material.

As a last part of the outlook, a new treatment strategy based on the developed protein cages is described. The idea arises from two facts. The first one is that people with CKD are five to ten times more likely to die from cardiovascular disease than to reach end-stage CKD and qualify for renal replacement therapies.^[3] Blood purification therapies should start in earlier stages of the CKD, which is not possible since the required lifelong treatment repeated multiple times per week is too cost-intensive. One-time treatment therapies are not known for CKD (except probably kidney replacement). The second fact is that a great fraction of many retention solutes linked to cardiovascular complications like *pCS*, IS or PAA are generated by gut bacteria. In fact, hemodialysis patients without colon have the same plasma *pCS* level as healthy patients.^[272] This is further verified by mice experiments.^[273] The toxin adsorbing ferritin cages may provide a possibility to remove the toxins directly in the colon by the introduction of genetic modified bacteria expressing the protein cage. In the best case, the cage gene is directly introduced in bacteria which metabolize the toxins. Produced toxins are sequestered in the cage and excreted with the bacteria. Since the bacteria are reproducing within the colon, no additional treatment is needed. Of course, special attention is needed for the administration of genetic modified organisms and measures should be taken to avoid their release in the environment. Knock-out parts of their metabolism to make them dependent on a special chemical could be a possible way to control the bacteria. This one-time treatment could help to offset the impact of retention solutes generated in the gut and could increase the life quality and expectancy of many patients. Since after initial development it would be a relatively cheap therapy, also the treatment of patients in low-income countries, which have no access to life-saving renal replacement therapies, would be possible.

8 Experimental part

8.1 General

Unless stated otherwise ultra-pure water prepared with a Purelab Flex 2 system (resistivity 18.2 M Ω cm) was used for preparation of buffers for functionalization, crystallization, and adsorption experiments as well as for chromatography. Prior to use the buffers were filtered through 0.22 μ m membrane filters (Merck). 15 mL and 50 mL polystyrene tubes (Carl Roth) or 1.5 – 2 mL reaction tubes (Carl Roth) were used for functionalization and batch crystallization experiments as well as for general sample storage. For centrifugation of larger flask (15-50 mL) an *Eppendorf* 5810R centrifuge and for smaller flask (1.5-2 mL) a *Heraeus Fresco 21* microcentrifuge from *Thermo Scientific* were used. Large volumes above 50 mL for cell harvesting were centrifuged with a *Multifuge X Pro Series* from *Thermo Scientific*. Experiments depended on sterile conditions were performed under a safety cabinet *Maxisafe 2030i* from *Thermo Scientific*. Competent cells and cryo cultures were stored at -80°C. Cell pellets were stored in the freezer at -20°C. Purified protein solutions or solutions of functionalized protein cages were stored at 4°C in the fridge. Crystalline or non-crystalline solid protein materials were stored at 20°C. Protein concentrations were determined with a *NanoDrop One C* spectrophotometer from *Thermo Scientific*. For dialysis (including buffer exchange, washing and concentration) of protein solutions centrifugal filter units Sartorius Vivaspin® Turbo 15 or Amicon® Ultra-0.5 with a molecular weight cut-off (MWCO) of 30 kDa were used.

8.2 Chemicals

All chemicals were purchased from commercial sources and used without further purification (see table of chemicals in section 12).

8.3 Analytic methods

8.3.1 TEM

Uranyl acetate stained protein samples were investigated by transmission electron microscopy. For negative staining copper grids, 200 mesh, covered with Formvar and carbon (Ted Pella, 01810), were incubated face down on a droplet of 10 μ L protein solution (concentration 0.3 mg mL⁻¹). Hereafter, the grid is washed 3 x in ultrapure water, followed by 1 wash and 1 incubation (45 s) step on 2% uranyl acetate drops. The remaining solution is blotted and the grids dried. All analysis was carried out with a Philips CM 300 UT TEM at 100kV.

8.3.2 ESI mass spectrometry

The protein sample were rebuffed to ultrapure water with an Amicon® Ultra 0.5 mL (MWCO 30 kDa) centrifugal filter. The protein sample is filled up with ultrapure water to a volume of 500 μL , concentrated to roughly 20 μL and refilled to 500 μL . The rebuffing is repeated 5 times and the concentration is set between 0.15 and 0.2 mg mL^{-1} . The mass of the proteins is determined with electron-spray ionization time of flight mass spectrometry (Agilent 6224 ESI-TOF-MS). Measurement is done in positive mode with a mass range of m/z 110 – 3200, with a rate of 1.03 spectra/s. The source temperature was set to 325°C, the drying gas flow to 10 L min^{-1} , the nebulizer pressure to 15 psig and the capillary voltage to 4000 V. Data interpretation was performed using the software MestreNova. The resulting spectrum shows multiple signals like observed in the spectra in Figure 10.3. Each signal can be attributed to an ion with the charge z . From the mass to charge ratio (m/z value) of two neighboring signals the charge z can be calculate after equation 1 with m_H as the mass of one proton (1.0072676 Da).

$$z_i = \left(\frac{\frac{m}{z_{i+1}} - m_H}{\frac{m}{z_i} - \frac{m}{z_{i+1}}} \right) \quad (5)$$

When the charge z is known, the weight of the protein can be calculated for one signal from its mass to charge ratio following equation 2.

$$M_{\text{Protein}} = z_i \left(\frac{m}{z_i} - m_H \right) \quad (6)$$

8.3.3 UV/Vis spectroscopy

8.3.4 Bradford assay

Bradford assay was used to determine concentration of proteins loaded with cargo molecules and for determine mass of BSA adsorbed to the protein-based materials.

For calibration Ftn^(neg)-4xCys or BSA solutions at concentrations of 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 mg mL⁻¹ were prepared. Bradford solution (AppliChem) is filtered through a syringe filter (0.22 µm) and 750 µL solution is mixed with 15 µL sample or calibration solution. After 5 min incubation time absorbance at 595 nm is determined with an UV/Vis spectroscope (Agilent Cary 60 UV-Vis). Linear regression is done for the calibration row and the concentration of the unknown sample is calculated.

8.3.5 Toxin assay

For determination of the uremic toxin adsorption capacity of the ferritin variants adsorption experiments were conducted. All solutions and samples were handled in glassware (Macherey-NAGEL Vials N9) since in initial experiments adsorption of the toxins on the polypropylene walls of the reaction tubes was observed.

At first, a stock solution of the desired toxin with a concentration of 50 µg mL⁻¹ for *pCS* and IS and 500 µg mL⁻¹ for PheAc is prepared by solving respective toxin in a buffer (50 mM Tris, 0.13 M NaCl, pH 7.4). From the initial stock solutions, a calibration row is prepared at concentrations of 0.01, 0.05, 0.1, 0.2, 0.5, 0.7, and 1 µg mL⁻¹ for later determination of the absolute toxin concentrations in the samples.

The stock solutions were diluted further to reach final uremic toxin concentrations expected in a stage five CKD patient (41 mg L⁻¹ for *pCS*,^[250] 44 mg L⁻¹ for IS,^[249] and 474 mg L⁻¹ for PheAc^[33]). The adsorbent was centrifuged down at 1500 g for 2 min and the complete supernatant was removed from the sample. 150 µL of the respective toxin solution was added and the crystals were incubated for 3 h at room temperature. Additionally, 150 µL of the toxin solution was incubated as a control. Three aliquots of 10 µL were removed from all samples and diluted 100 times in ultrapure water. Finally, crystals were washed with water, dried under vacuum, and weighed.

The uremic toxin concentration was quantified by an ultrahigh-performance liquid chromatography system (1290 infinity II UHPLC, Agilent) with a reversed phase C18 column (Zorbax Extend-C18, 2.1x50 mm, 1.8 µm particle size, Agilent) coupled to an electron-spray-ionization quadrupole-linear ion trap-mass spectrometer (QTRAP 5500, Sciex). All compounds were measured in positive mode with an Ion spray Voltage of -4500 V, Temperature 500°C, Ion Source Gas 1 and 2 50 psi, Curtain Gas 20 psi. MRM parameter for all compounds are listed in Table 8.1.

As solvent a mixture of HPLC grade water (LiChrosolv® Merck) and acetonitrile (LiChrosolv® Merck) both with addition of 0.1% formic acid (Honeywell Fluka) is used. Specific compositions at each step during the 15 min long chromatography program are summarized in Table 8.2.

Before each sample incubated with the crystals, the control is measured. Chromatograms are evaluated with the Analyst® Instrument control and Data processing Software. Peaks are integrated and the toxin concentration is determined from the calibration row. From the concentration difference between the control and the samples the amount of adsorbed uremic toxin is determined. The adsorption capacity is then determined by dividing the mass of the adsorbed toxins by the mass of the crystals.

Table 8.1. MRM parameters

	Q1 mass [Da]	Q3 mass [Da]	Collision energy [V]
Indoxyl sulfate			
Quantifier	211.884	80.000	-38.000
Qualifier	211.884	80.900	-24.000
p-Cresyl sulfate			
Quantifier	186.852	106.900	-32.000
Qualifier	186.852	79.900	-24.000
Phenylacetic acid			
Quantifier	134.927	90.900	-12.000
Qualifier	134.927	64.900	-66.000

Table 8.2: Sequence of HPLC program

Time [min]	H ₂ O [%]	Acetonitrile [%]	Flow rate [mL min ⁻¹]
0.00	97	3	0.3
8.00	70	30	0.3
9.00	40	60	0.3
10.50	97	3	0.3

8.4 Protein production and purification

8.4.1 QuikChange site-directed mutagenesis

The introduction of cysteine anchor sites has been performed by multiple cycles of QuikChange™ site-directed mutagenesis using a two-step polymerase chain reaction (PCR) protocol.^[274] The used primers for the different mutation sites are shown in Table 8.3. A mixture of 2.9 µL pET-22b(+) plasmid containing the gene of interest (7 ng µL⁻¹), 1 µL of 10 mM dNTP mix, 5 µL reaction buffer 10x (100 mM KCl, 100 mM (NH₄)₂SO₄, 200 mM Tris-HCl, pH 8.8, 20 mM MgSO₄, 1% Triton® X-100, 1 mg mL⁻¹ nuclease-free bovine serum albumin), 1 µL Pfu DNA polymerase (2.5 U µL⁻¹) and 38.1 µL ultrapure water was prepared. The mixture was split in half and 1 µL of forward or reverse primer (10 pmol µL⁻¹) was added to each tube. PCR thermocycler (Eppendorf Mastercycler Nexus PCR Cycler) was prepared by initial heating step for 30 s at 95°C. The first step of the PCR protocol was composed of 3 cycles of 30 s denaturation at 95°C, annealing for 1 min at 61°C followed by elongation for 6 min at 68°C. After the first 3 cycles, the separate mixtures with the forward and reverse primers were combined and the PCR was continued for 16 additional cycles with the same parameters as the first 3. Followed by a final elongation phase for 10 min at 68°C to finish the PCR. Digestion of parental plasmid was done by adding 1 µL DpnI (10 U µL⁻¹) and incubation overnight at 37°C. DpnI was deactivated by 20 min heating at 80°C and the mixture was purified using the NucleoSpin® Gel and PCR clean-up kit according to the manufacturer's instructions. *E. coli* DH5α calcium-competent cells were incubated with 200 ng of the purified plasmid for 30 min on ice, followed by a heat shock for 45 s at 42°C. Next, cells were incubated for 1 h in super optimal broth (SOB) media, centrifuged at 1000 g, resuspended in 100 µL media, plated on an LB agar plate, and incubated for 16 h at 37°C. A single colony was picked and incubated overnight at 37°C and 250 rpm in 5 mL sterile LB medium supplemented with 150 µg mL⁻¹ ampicillin. The next day, plasmids were extracted by NucleoSpin® Plasmid miniprep kit according to the manufacturer's instructions.

The sequence was confirmed by mixing 500 ng plasmid with 25 pmol T7 forward or reverse primer in a 10 µL solution and sent in for DNA sequencing (Eurofins Genomics). Plasmids with the desired mutations inside were then chosen as parental plasmids for further mutagenesis till all 5 mutations were present.

Table 8.3. Sequence of primers used for QuikChange™ PCR protocol.

Mutation	Forward primer
C130A	5'-C AAG AAC GAT CCG CAT CTG GCC GAT TTC ATC GAA ACC CAC-3'
K53C	5'-T GTT GCA CTG AAG AAC TTT GCG TGT TAC TTT CTG CAT CAG TCC CAT G-3'
E64C	5'- TTT CAT CAG TTT CTC GGC ATG ACA GCG TTC TTC ATG GGA CTG ATG-3'
K143C	5'-AT CCC CGA GTT CTT TGA TCG CAC AGA CCT GTT CGT TCA GAT AGT G -3'
S178C	5' - CAC ACT TTG GGT GAT <u>TGC</u> GAT AAT GAA TCG TAA CTC GAG CAC C - 3'

8.4.2 Protein production and purification

Production of Ftn^(neg) and its variants is identical to the production of Ftn^(neg) previously published.^[89]

First calcium-competent *E. coli* BL21-Gold(DE3) cells were thawed on ice for 10 min. Then 1 μ L of 40 ng μ L⁻¹ plasmid solution was added to the cells and the mixture was incubated on ice for 30 min. Heat shock was done by incubating the mixture at 42°C for 45 s followed by 2 min incubation on ice. The cells were suspended in 1 mL SOB media and incubated for 1 h at 37°C. Cells were centrifuged at 1000 g and 1 mL of the mixture is removed. The cell pellet was resuspended in the remaining solution and streaked out on an LB-agar plate suspended with 150 μ g mL⁻¹ ampicillin and incubated at 37°C overnight.

To prepare precultures colonies of transformed *E. coli* BL21-Gold(DE3) cells (Agilent) were incubated overnight in 5 mL sterile LB-Miller medium supplemented with 150 μ g mL⁻¹ sodium ampicillin at 37°C and 180 rpm.

Then 400 mL of Terrific broth (TB) medium supplemented with 150 μ g mL⁻¹ sodium ampicillin was inoculated with 4 mL of the preculture. The cells were grown at 37°C and 180 rpm till an OD₆₀₀ of 0.6 was reached. Protein overexpression was induced by the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.25 mM and the cells were incubated for an additional 48 h at 18°C. The cells were harvested via centrifugation at 4000 g. Pellets were stored at -20°C until further use.

Cells from 400 mL culture were resuspended in 20 mL buffer (50 mM Tris, pH 7.5, 0.3 M NaCl). Cell lysis was achieved by sonicating (60% amplitude) for six times 1 min on ice with 1 min break in between with a Vibra-Cell VCX-130 ultrasonic processor (Sonics). The resulting suspension was centrifuged at 14,000 g for 20 min to separate the cell debris from the soluble proteins. Denaturation of most of the *E. coli* proteins was achieved by heating the supernatant

to 65°C for 10 min in a water bath. The denaturated proteins were separated by centrifugation at 14,000 g for 15 min. Proteins left in the solution were precipitated with ammonium sulfate at a final concentration of 70% of its saturation concentration, followed by centrifugation at 14,000 g for 20 min. After rebuffering the pellet in 10 mL buffer (50 mM Tris, pH 7.5, 0.15 M NaCl) the ammonium sulfated precipitation was repeated. The resulting pellet was dissolved in 50 mL IEC loading buffer (50 mM Tris, pH 7.5, 0.15 M NaCl) and purified by ion-exchange chromatography (IEC) with a linear gradient from 0.15 to 1 M NaCl using a 5 mL HiTrap™ Q HP anion exchange column (Cytiva). All Ftn^(neg)-4xCys containing fractions were collected and concentrated to a final volume of 2 mL with a Sartorius Vivaspin® Turbo 15 (MWCO 30,000) filter unit. Finally, the sample was purified via gel filtration with a HiLoad 16/600 Superdex™ 200 pg column. All chromatography steps were performed on an Äkta pure system from Cytiva. All Ftn^(neg)-4xCys containing fractions were collected and stored at 4°C until further use.

8.4.3 Purification from exclusion bodies.

The proteins were produced analog to the previous section. Cells from 200 mL culture was suspended in 20 mL buffer (50 mM Tris, pH 7.5, 0.3 M NaCl). Cell lysis was achieved by sonicating (60% amplitude) for six times 1 min on ice with 1 min break in between with a Vibra-Cell VCX-130 ultrasonic processor (Sonics). The resulting suspension was centrifuged at 14,000 g for 20 min. The pellet was resuspended in a wash buffer (50 mM Tris, 1 mM EDTA, 1% Tween-20, pH 7.5) to remove cell debris and membrane proteins. After 5 min incubation the suspension is centrifuged again. The step was repeated three times. The washed pellet is incubated in solubilization buffer (50 mM Tris, 0.3 M NaCl, 8 M Gua, pH 8) at a final concentration of roughly 15 mg mL⁻¹ with respect to the wet cell mass. After 2 h the pellet were completely dissolved. For protein refolding 1 mL of the solubilized protein solution is diluted in 49 mL SEC buffer (50 mM Tris, 0.3 M NaCl, pH 7.5). The mixture is kept at rest at RT overnight and next day purified via SEC. Ferritin containing fractions are collected for further analysis.

8.5 Functionalization of cysteines in the inner cavity

8.5.1 Functionalization with 2-iodo-N-phenylacetamide

5 mg Ftn^(neg)-3xCys or Ftn^(neg)-4xCys were incubated for 4 h in disassembly buffer (10 mM phosphate; 50 mM NaCl, pH 2). After 3½ h 10 eq (with respect to each cysteine) Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Iris Biotech GmbH) was added to the solution from a 10 mg mL⁻¹ stock solution. Subsequently, the solution was filled up to 15 mL with reassembly buffer (50 M Tris, 50 mM NaCl, pH 7.6) and concentrated with a membrane filter (Sartorius Vivaspin Turbo 15; 30 kDa MWCO) to a final volume of 200 µL. Again 10 eq. TCEP were added and the solution was filled up to a volume of 2 mL. The pH value was set

to 7.6 with 1 M NaOH or HCl. Afterwards 2 mL Ethanol containing 20 eq. 2-iodo-N-phenylacetamid (abcr GmbH) was added to the solution and the mixture was stirred at 300 rpm in the dark for 1 h. Next, the solution was filled up with reassembly buffer to a total volume of 30 mL. The protein reassembles overnight. Finally, the solution is concentrated to 2 mL and purified via gel filtration on a HiLoad 16/600 Superdex™ 200 pg column. Protein containing fractions were collected and stored at 4°C for further use.

8.5.2 Functionalization with 2-Bromo-N-decylacetamide

The functionalization follows closely the protocol for 2-iodo-N-phenylacetamid. During the functionalization reaction the protein/TCEP solution was not filled up to 2 mL but to 800 µL and then 3.2 mL Ethanol containing 40 eq. 2-bromo-N-decylacetamid (Sigma-Aldrich) was added to the solution. All other steps were performed equal to the protocol of 2-iodo-N-phenylacetamide functionalization.

8.5.3 Functionalization with fluorophores

For loading of organic dyes containing maleimide groups an already published protocol was used following an acidic dis/reassembly approach.^[164]

First, 1 mg mL⁻¹ Ftn^(neg) was disassembled under acidic conditions by incubating for 4 h in disassembly buffer (10 mM phosphate; 50 mM NaCl; pH 2). After complete disassembly the solution was diluted by twice the amount of reassembly buffer (50 mM Tris; 50 mM NaCl; pH 7.5) and the pH was adjusted to 7.4 by addition of 1 M NaOH or HCl. Next, 1 eq. of respective dye from a 200 mM DMSO stock solution was added and the mixtures was incubated overnight at RT in the dark. The solution was concentrated with a centrifugal filter unit (Sartorius Vivaspin Turbo 15; 30 kDa MWCO) and washed two times with SEC buffer (50 mM Tris; 300 mM NaCl, pH 7.5) and purified by SEC on a HiLoad 16/600 Superdex™ 200 pg column. Ferritin containing factions were collected and stored at 4°C till further use. The following dyes coupled to a maleimide group were used Rhodamine 6G (Atto-Rho6G, Atto-Tech ≥99%), Alexa Fluor 488 (Thermofisher, 99%) or carboxy-SNARF (BOC Sciences ≥90%).

8.5.3.1 Determination of fluorophore loading

The amount of AF488 loaded to the Ftn^(neg) cysteine variants was determined by UV/Vis spectroscopy. The protein cage was disassembled at pH 2 prior to measurement to avoid quenching of the dye due to high local concentrations. Since the protein has no absorbance at wavelength of 493 nm, dye concentration was directly determined by preparing standard calibration curves with concentrations ranging from 1 – 10 µM. Protein concentration where also determined by a calibration row in the range of 0.1 – 1.0 mg mL⁻¹. However, since the dye

has also absorbance at 280 nm, the respective value needed to be subtracted prior to the calculation. Determination of 280 nm absorbance of the fluorophore assumes that the ratio of absorbance at 280 nm and 493 nm is constant at any concentration and was calculated from the prior determined dye concentration. Finally, loading was determined by calculation the ratio of dye to protein container.

8.6 Crystallography

8.6.1 Hanging drop crystallization

Crystallization of little amounts of protein or functionalized protein variants were performed via hanging drop vapor diffusion techniques. Reservoir solution (100 mM Tris, 500 mM MgOAc, pH 8.5) was prepared in a 24- well manual plate set. Drops were prepared on siliconized glass cover slides (Jena Bioscience) by mixing 2 μ L reservoir solutions with 1 μ L 50 mM Tris, 1 M NaCl, pH 7.5 buffer and 1 μ L of respective ferritin variant. Plates were incubated at 25°C. After one day first crystals were visible.

8.6.2 Structure determination and refinement

Crystals, soaked for 30 s in cryo buffer containing 25% glycerol and the respective reservoir solution, were vitrified in liquid nitrogen.

Diffraction data were collected at 100 K on P11 at the *Deutsches Elektronen-Synchrotron DESY* (support by the beamline staff at P11 is gratefully acknowledged). For structure determination, the CCP4 suite was used.^[275] Data processing and scaling were done with XDS^[276] and the structures were solved via molecular replacement with Phaser.^[277] As input structure, an adapted version of the Ftn^(neg) crystal structure (PDB ID: 5JKK) with Ftn^(neg)-Cys mutations manually added with *Coot*,^[278] was used. At some cysteine sites, additional electron density was observed. At this positions, a shortened version of the aliphatic ligand with only 3 carbon atoms of the aliphatic chain was introduced and covalently linked to the model using AceDRG.^[279] The model was improved by iterative rounds of refinement and manual rebuilding using Refmac.^[280] Metal ions and water molecules were placed based on the electron density. Fe³⁺ ions were placed at the ferroxidase site. Mg²⁺ ions coordinated by six water molecules were placed in the threefold channel. The model were validated with the Molprobit server.^[281] Crystallographic details are summarized in Table 10.5.

Crystal structure of redesigned protein variants (Ftn^(neg)-Ap and Ftn^(neg)-dock) were prepared in a similar way. For molecular replacement with Phaser the models generated by *Rosetta* were used as input structure. Crystallographic details are summarized in Table 10.8 und Table 10.9.

8.6.3 Batch crystallization

For crystallization of larger amounts of Ftn^(neg) and functionalized variants of Ftn^(neg), a batch crystallization approach was employed based on a protocol from I. Rayment.^[240] For a standard experiment, 250 μL of a 50 mM Tris 1 M NaCl pH 7.5 buffer was carefully mixed with equal amounts of Ftn^(neg) stock solution with a concentration of 12 mg mL⁻¹ in a 50 mM Tris 0.3 M NaCl pH 7.5 buffer. Afterward 500 μL of the precipitant solution (133 mM Tris, 333 mM MgOAc, pH 8.5) was added dropwise under constant shaking. The mixture was kept at rest for 7 days at an ambient temperature of 20°C before crystals are fixated.

8.6.4 Fabrication of non-crystalline adsorbent

25 μL 2.5% glutaraldehyde solution is added to 500 μL of a 6 mg mL⁻¹ protein solution, complete mixing is ensured by carefully pipetting the solution up and down for three times. The mixture is kept at an ambient temperature of 20°C overnight. Next day, a white precipitated formed. The material is washed three times with ultrapure water and stored till further use at 20°C.

8.6.5 Crystal fixation

8.6.5.1 Fixation with Sulfo-SMCC crosslinker

Increasing the stability of the crystals was needed for the adsorption experiments. For that purpose, they were fixated with a Sulfo-SMCC (sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate, Thermo Scientific) crosslinker. The crystals were centrifuged down at 1000 g for 2 min. In a standard experiment with 3 mg crystals the crystallization solution was removed till 246 μL were left and 64 μL of a freshly prepared 4.8 mg mL⁻¹ aqueous Sulfo-SMCC solution was added resulting in an end concentration of 1 mg mL⁻¹. The mixture was kept at rest for 4 h at room temperature. And subsequently filled up to 1 mL with ultrapure water, the crystals were spun down via centrifugation at 1500 g for 2 min. The supernatant was removed and the crystals were resuspended in ultrapure water afterwards. The procedure was repeated up to three times to wash the crystals from the residual cross-linking agent. The material was stored at an ambient temperature of 20°C until further use. Crystals were photographed under a Leica S9D microscope with a FlexaCamC1.

8.6.5.2 Fixation with glutaraldehyde

For fixation with glutaraldehyde (Merck) to 1 mL of crystal solution containing 3 mg crystals, 50 μL of a 2.5% aqueous glutaraldehyde solution is added, resulting in an end concentration

of 0,00119%. Crystals were incubated for 4 h and subsequently, three times washed with ultrapure water. The crystals were stored at 20°C till further use.

After the glutaraldehyde crosslinking, the crystals still dissolved in a 60 mg mL⁻¹ BSA solution. The stability could be increased by an additional fixation step. The procedure was repeated, but the crystals were only incubated for 10 min before washing. However, toxins adsorption assays revealed a significantly reduced adsorption capacity. We hypothesize that polymerization of the glutaraldehyde leads to blocking of the pores.

8.7 Biocompatibility assays

8.7.1 Quantitative polymerase chain reaction (qPCR) analysis of mRNA expression in human aortic endothelial cells

The experiments were performed by *Setareh Orth-Alampour* from the *Jankowski group* at the *University Hospital Aachen*.

Human aortic endothelial cells (hAoECs) (Promocell) were cultivated in a Endothelial Cell Growth Medium MV (Promocell). Cells were seeded in 24 well plates (15 x 10⁴ cells/well) at 80% confluence and were incubated for 6 h with 100 ng mL⁻¹ lipopolysaccharides (LPS) or functionalized or non-functionalized protein crystals. After incubation time the total RNA was extracted using RNAeasy mini kit (Qiagen). Reverse transcription was performed using 1 µg total RNA (600 ng), random hexamers, and Verso reverse transcriptase (Thermo Scientific) as per the manufacturer's instructions. For real-time PCR, gene expression levels were quantified using SYBR Green I dye chemistry on a LightCycler 480 system (Roche Applied Sciences). The following primers were used for relative quantification of targeted gene expression - for human TNF-alpha: forward primer 5'-GCC CAG GCA GTC AGA TCA TCT-3', reverse primer 5'-TTG AGG GTT TGC TAC AAC ATG G-3' and for human beta-actin: forward primer 5'-CAA CCG CGA GAA GAT GAC-3', reverse primer 5'-GTC CAT CAC GAT GCC AGT-3'. Data were represented as the mean level of gene expression relative to the expression of the reference gene (β-Actin).

8.7.2 Platelet activation assay

The experiments were performed by *Setareh Orth-Alampour* from the *Jankowski group* at the *University Hospital Aachen*.

The platelets of three donors were isolated by centrifugation at 260 g for 15 min. After a second centrifugation step, platelets were resuspended in Hepes buffer pH 6.6 (10 mmol L⁻¹ Hepes, 136 mmol L⁻¹ NaCl, 2.7 mmol L⁻¹ KCl, 2 mmol L⁻¹ MgCl₂ and 5 mmol L⁻¹ glucose). Platelet suspensions were re-centrifuged in the presence of 1:15 acid citrate dextrose (ACD) and 1 U mL⁻¹ apyrase and subsequently resuspended into Hepes buffer pH 7.45 (10 mmol L⁻¹ Hepes, 136 mmol L⁻¹ NaCl, 2.7 mmol L⁻¹ KCl, 2 mmol L⁻¹ MgCl₂, 5 mmol L⁻¹ glucose and 0.1% BSA). 15 x 10⁶ platelets were incubated for 15 min with 4 nmol L⁻¹ Thrombin in presence of 2 mmol L⁻¹ CaCl₂ or different protein crystals. The platelets were lysed with 4% SDS lysis buffer (200 mmol L⁻¹ Tris, 600 mmol L⁻¹ NaCl, 4% SDS) including EDTA-free Halt Protease Inhibitor Cocktail (1:10; Sigma-Aldrich) and Halt Phosphatase Inhibitor Cocktail (1:10; Sigma-Aldrich). Protein amount was quantified following the protocol for DC protein assay (Bio-Rad). An equal amount of protein from each sample was resolved by 10% SDS–polyacrylamide gel electrophoresis, transferred to nitrocellulose membranes, and blocked with 5% bovine serum albumin (BSA) for 1 h at room temperature. Anti-phospho-Akt antibody (1:1000; Cell signaling Technology) and anti-tubulin (1:1000; Cell signaling Technology) were used as primary antibodies. The blots were incubated overnight at 4°C. A second anti-rabbit antibody (1:1000; Cell signaling Technology) was used and incubated for 1 h at room temperature. Immunoreactive bands were visualized via enhanced chemiluminescence, and densitometry was performed using ‘Quantity One software’ (Bio-Rad Laboratories).

8.7.3 BSA adsorption assay

BSA stock solution was created by solving 60 mg BSA in 1 mL buffer (50 mM Tris, 0.13 M NaCl pH 7.5). 3 mg of cross-linked crystalline adsorbent was incubated in 300 µL BSA solution. BSA solution without sample and crystalline material in water and incubated likewise to serve as control samples. After 3 h 10µL sample was removed and diluted in 990 µL ultrapure water. BSA concentration was determined by Bradford assay as described in section 8.3.4 for each sample a triplicate was measured and the adsorbed mass of BSA was derived from the difference of the control sample and the sample with crystalline material.

8.8 Computational methods

All simulations were done on the high-performance-computing (HPC) cluster *Hummel* (HPC-cluster 2015) of the *University Hamburg*. To connect with the HPC-cluster the software *Putty* were used and data up- and download were done using the *FileZilla* software.

8.8.1 Preparation

8.8.1.1 HPC-cluster and batch scripts

To organize simultaneous work of many users, a batch job system was used. A job was queued by executing a shell-script (sh-file) on one of the front-end nodes, where also basic file navigation was done. A commented minima example for a script is shown below.

```
#!/bin/bash                #shebang

#Options of the batch job

#SBATCH --job-name=Name
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

#System specific initialization

set -e                    #stops process when error is recognized
source /sw/batch/init.sh  #initilize source system

# expands source system (allows parallel Rosetta runs)
module switch env env/gcc-8.2.0_openmpi-3.1.3

#The actual code starts here
```

All scripts were stored as a sh-file and were started from the front-end node with the following command.

```
sbatch script_name.sh
```

All following scripts were started by this sequence, which is not further described. The complete code for each script can be found in the appendix section 10.5.

8.8.1.2 Compilation of Rosetta

The latest version of the Rosetta software (version 3.19) was downloaded from www.rosettacommons.org and uploaded to the cluster. To allow parallelization of independent

trajectories Rosetta was compiled with the MPI (message passing interface) option. The full script is shown in the appendix.

To successful run Rosetta in the MPI modus, shared libraries needed to be defined. This was done by the export `LD_LIBRARY_PATH` command, which needed to be executed before every Rosetta run and was also standardly added to all scripts. Also, the commands were executed with the `mpirun` command to run in MPI modus the. For the sake of shorter example scripts these necessary commands are excluded in the following, but they can be found in the fulltext scripts in the appendix section 10.5.

8.8.1.3 Preparing the input structure

High resolution crystal structure of human heavy chain ferritin (PDB ID: 2CEI) was used as an input structure. For some simulations mutations A18E, C90E, C102E, H105E were added using *Coot*, to account for the changes from the native structure to $Ftn^{(neg)}$.

Every entry in the PDB-file except the lines starting with ATOM or CRYST were deleted. The PDB 2CEI only contained the structure of one single subunit. In the full cage each subunit is surrounded by six other subunits leading to an energetic stabilization. The neighboring subunits can be generated using a symmetry-definition file based on the crystallographic symmetry.

8.8.1.4 Generate a symmetry definition file

The symmetry definition file was created using the *make_symmdef_file.pl* script, which is part of the Rosetta software suit.

1. The input PDB file was copied in a new directory
2. The following command was run to generate the symmetry definition file:

```
make_symmdef_file.pl -m CRYST -r 12.0 -s P 4 3 2 -p 2CEI.pdb  
> symmetry_definition-file.symm
```

This command generated a symmetry definition file for the ferritin container. The full script along with needed input file is shown in the appendix in section 10.5.2 . If not stated otherwise the symmetry definition file was included to every following Rosetta protocol to generate six additional surrounding subunits. Therefore, the original subunit in the middle had the same chemical environment (neighboring protein chains) as in a full container, making sure that the simulations were close to the true structure but saving resources because the full cage didn't need to be calculated.

8.8.1.5 Energy minimization of the input structure

The side chain conformation present in the X-ray crystal structure was most likely not in its energetic minima. Therefore, the Rosetta molecular modeling program offers a predefined protocol called *relax*, allowing to minimize the energy of the input structure. How the *relax* protocol was started is described in the following.

An option file containing the following information were created.

```
# Path to the Rosetta databases
-database $your_Path_to_Rosetta/main/database
# Define the input structure for the relax run
-s Cleaned-PDB.pdb
# Rotamer packing options
-ex1
-ex2
-use_input_sc

# Add coordinate constraints to the backbone based on the
# input structure
-constrain_relax_to_start_coords

# Sets the number of output structures
-nstruct 500

# Define symmetry definition file
-symmetry::symmetry_definition symmetry_definition-file.symm
```

In the following, this text file will be named *options_relax*. After everything was prepared the energy minimization was started by performing the following steps.

1. The prepared input file, the options file and the symmetry definition file were added in a new directory.
2. The *relax* application was started with the following command

```
$your_Path_to_Rosetta/main/source/bin/relax.linuxgccrelease
@options_relax
```

3. After the application was finished, new PDB files named like the input PDB with a number at the end ranging from 1 to the number given with the *nstruct* option are generated. Additionally, a file called *score.sc* containing the total Rosetta score and all energy terms for each outputted protein structure was generated. This file was sorted from the lowest to the highest total score with the following command.

```
grep -H 'SCORE' score.sc |sort -k2 -n|head -n20|nl -w 3 -s ''
>Output-file
```

500 independent trajectories were calculated. The first one possesses nearly identical Rosetta scores indicating that the structure was indeed in a global energetic minimum. The PDB file containing the structure with the lowest score was used as an input for all further modeling tasks. Since the symmetry definition file was applied, the PDB contains 7 subunits. All subunits except chain A were removed from the file when used as input for further applications. The complete scripts are shown in the appendix in section 10.5.3.

8.8.2 Design with fixbb protocols

To identify possible mutation site the predefined fixed-backbone (fixbb) protocol was used. Like the name suggest, in this modeling task the backbone of the protein structure is kept fixed, and the side chains and their conformation were sampled.

For this run new specific input files were created. A new options file (fixbb_options) was created.

```
# Path to the Rosetta databases
-database $your_Path_to_Rosetta/main/database
# Define the input structure for the fixbb run
-s energy-minimized.pdb
# Define the residue file
-resfile name_of_resfile
# Define the weights file
-score:weights weights.wts
# Rotamer packing options
-ex1
--ex2
-use_input_sc
# Sets the number of output structures
-nstruct 500
# Define symmetry definition file
-symmetry::symmetry_definition symmetry_definition-file.symm
```

To specify which amino acids were allowed at which positions a residue file was prepared. A standard example is shown below, for alle simulations performed in this work the full content of the resfile, will be shown for each design in the appendix in section 10.5.4.3.

```
#default command for every position not included in the
#resfile
NATAA # allow only the native amino acid (NATive Amino Acid)
# list of positions in the protein which derivates from the
# default behavior
start
19 A PIKAA AKR
...
```

After the line with *start*, all positions in the input structure where a mutation is allowed were listed. Besides the position number also the chain identifier (in this case A) is needed. There

were different ways to specify the possible mutations. In this case, the PIKAA (pick amino acid) command was used followed by a list of amino acids (in one letter code), which were allowed to introduce. Additionally, a weights file, containing the weights for tuning the energy terms of the score was added. In all simulations done in this work the most recent standard weights file suggested by Rosetta commons was used. In the following scripts this file is called weights.wts and the full content is shown in the appendix in section 10.5.4.2.

The fixbb protocol is stated as follows.

1. The PDB file containing the relaxed protein structure, the options_fixbb, weights.wts, residue and symmetry definition file were copied in a new directory.
2. The fixbb protocol was started by the following command. The complete batch script can be found in the appendix in section 10.5.4.1.

```
$Your_path_to_Rosetta/main/source/bin/fixbb.linuxgccrelease  
@options_fixbb
```

3. The results of the fixbb run were new protein structure files and their Rosetta score stored in the score.sc file. Similar to the relax run, the score.sc file was sorted according to the total score using the grep-command already introduced for the relax run. The amino acid sequence for the best structures was extracted from the PDB files with the following command

```
perl $your_Path_to_Rosetta/main/source/  
scripts/python/public/pdb2fasta.py List_of_PDBs >  
AA_sequence.fasta
```

7. The sequences were downloaded and compared by performing sequence alignments using the *CLC sequence viewer* software. Interesting structures were further investigated.

8.8.3 Design of binding sites

For protein ligand docking Rosetta protocols specified with flexible xml-script were used. The modeling done in this work follows closely a published protocol.^[235] Again, the energy-minimized PDB structure and the symmetry definition file were used. Additionally, the structure of the ligand was prepared for modeling.

8.8.3.1 Preparing the ligand structure

By default, Rosetta only “knows” canonical amino acids and some more molecules. The structure of the ligand and its conformers need to be added to the software suite. This was done by creating a so-called params file, allowing Rosetta to handle the ligands.

1. The ligands structures were prepared with the program *Avogadro* and stored in a data file (sdf file).
2. The web-based server *Frog2*^[282] were used to create a file containing the most likely conformers of the ligand.
3. The structure file and the conformers file were uploaded to the HPC-cluster. The params file was generated by executing the following script. The full script is shown in the appendix in section 10.5.6.1.

```
python ../molfile_to_params.py -n 3IDS -p 3IDS --conformers-in-one-  
file 3IDS_conformer.sdf
```

4. The resulting params file was used for all following Rosetta runs, where the respective ligand is used.
5. Additionally, a file containing the structure of the ligand in a PDB format was generated. This file was added to the end of the PDB file containing the energy-minimized input structure.

8.8.3.2 Ligand-docking protocol

Like mentioned before, instead of a predefined protocol a flexible xml script was used as a protocol for the ligand docking. The script was to a large contend copied from the published protocol^[235] and was shown in full in the appendix in section 10.5.6.2.1. The script file is called dock.xml in the following text.

Similar to the fixxbb or relax protocol an options file called options_dock.txt was created with the following content.

```
#Specifiying the PDB and ligand input file  
-in  
  -file  
    -s '../input/Ap4_3IDS.pdb'  
    -extra_res_fa ../input/3IDS.params  
  
-packing  
  -ex1 #rotamer library  
  -ex2 # rotamer library
```


Experimental part

```
-no_optH false           # optimize hydrogen placements
-flip_HNQ true           # consider HIS,ASN,GLN hydrogen flips
-ignore_ligand_chi true  # don't add additional ligand rotamer

-parser                  # locates the XML file
-protocol ../dock.xml

-mistakes                # restores certain parameters to          #
                        # previously published values

-restore_pre_talaris_2013_behavior true

-nstruct 500            # number of generated output structures
```

For modeling were also mutations should be introduced to the protein backbone a new residue file was created.

```
ALLAA
AUTO
start
1 X NATAA
```

This short file allowed in combination with the `design_interface` mover, which is included in the xml script to detect residues in vicinity to the ligand and their mutation to any other canonical amino acid (`NATAA`).

To start the docking run, following steps were performed.

1. The PDB file containing the relaxed protein and ligand structure, the params-file for the respective ligand, the `options_dock`, residue and symmetry definition file were copied in a new directory.
2. The ligand docking script was started with the command

```
/Path_to_Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease  
@options_design.txt -nstruct 500 > log.txt
```
3. Again, the demanded number of output structures and a score file were generated by the program.
4. To decrease the number of structures a filtering step was done. A file called `metric_treshold.txt` was prepared, which sets the options of the following filtering step.

```
req total_score value < -800 # measure of protein stability
```

```
req if_X_fa_rep value < 1.0 # measure of ligand clashes
req ligand_is_touching_X value > 0.5
output sortmin interface_delta_X # binding energy
```

Only structures were accepted which fulfill the requirements (req) of a sufficient protein stability and certain metrics for a favorable binding site. Ultimately the output was sorted by the interface_delta_X term, which is a measure for the binding energy of the ligand. Filtering based on the metric thresholds defined earlier was done by a built-in Rosetta script DesignSelect.pl. The output of the script was stored in a new file called filtered_pdb.txt. The full script is shown in the appendix in section 10.5.6.3.

5. Additional metrics were calculated using Rosettas InterfaceAnalyzer protocol. First the names of the PDB files were extracted to a single list and expanded by their full path. The full script is shown in the appendix in section 10.5.6.4. The InterfaceAnalyzer script was performed for every PDB in the list, calculating additional metrics for the protein-ligand interface. The InterfaceAnalyzer produced a design_interface.sc file, where all additional metrics were stored.
6. The design_interface.sc file was used to filter the output structures a second time. A file called metric_treshold_2.txt was prepared with the following options.

```
req packstat value > 0.55 # packing metric
req sc_value value > 0.45 # shape complementarity
req delta_unsatHbonds value < 1.5 # unsatisfied hydrogen bonds
req dG_seperated/dSASAx100 value < -0.5 # binding energy per
                                     contact area
output sortmin dG_seperated # measure for binding energy
```

This time the output was sorted after the dG_seperated value, which was used as a measure of binding energy. This value represents the difference in energy of the protein ligand complex compared to the unbound state. Filtering was again done by the DesignSelect.pl script started by a shell script shown in the appendix. A final output file was generated ranking the PDB structures which show sufficient protein stability and good binding metrics ranked by the dG_seperated value.

This was the generic protocol for one ligand docking run. It has been shown that subsequent modeling tasks using the respective output structure as new input can improve the binding energy. Therefore, all prior steps were combined in a script and nested in a for-loop. This allows to perform multiple consecutive docking tasks. And stores the results. The full script is shown in the appendix in section 10.5.6.6.

8.8.4 Protein-protein docking

The method was adapted from an example protocol given in a hands-on workshop on Rosetta kindly made publicly-accessible by the *Meiler-Lab*.¹

Protein-protein docking protocols were performed using the structure of a binary ferritin crystal, composed from positively and negatively surface charged variants (PDB ID: 5JKL). Two contact areas between the cages are present, each of these interfaces was investigated separately and all subunits not present at the investigated interface were deleted from the structure. In the final structure 2 subunits per ferritin variant were retained. As additional preparations all metal ions and water molecules were deleted, and the energy of the remaining protein structure was minimized. Respective mutations were manually added using the *Coot* software. For later evaluation of the results root-mean-square deviation (RMSD) of atomic positions in the protein is determined, here only the residues directly located at the interface were included, to allow precise specification the amino acids in the input structure were renumbered.

Prepared structures were uploaded to the HPC-cluster. The protein-protein docking script was stored in an xml file and also uploaded to the cluster together with an options file specifying the general Rosetta options similar to prior described option files and specific docking options. An important option for the application with the protein container is the `partner` option. This was used to prevent the separation of subunits belonging to the same container and to exclude the energy from their interface for later evaluation. With the option `dock_mcm_trans_magnitude` and `rot_magnitude` the translational and rotational perturbation during the low-resolution docking step were set. Increasing them will increase the search space but limit its accuracy. The full content of the option files is shown in the appendix in section 10.5.7.4.1.

The full content of the protein-protein docking script is also shown in the appendix in section 10.5.7.2. Respective script contains the steps needed for the docking namely the low-resolution docking step, the high-resolution docking step and a final energy minimization of the interface. This script, together with the input pdb-file and the options file was saved in the same directory and the docking simulation was started with a batch script containing the following command. The full script is given in the appendix in section 10.5.7.1.

¹ Tutorial 3: Protein-Protein Docking; <https://meilerlab.org/rosetta-workshop-2020/>

Experimental part

```
/Path_to_Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@docking.options -parser:protocol docking_full.xml -out:suffix _full -
nstruct 1000 >& docking_full.log
```

Additional to the full docking script a version omitting the low- and high-resolution docking step was performed. Respective script only performed the final energy minimization of the sidechains at the interface. The resulting models were used as a reference to the actual docked models for further evaluation. Respective modeling was performed similar to the previous one, but with an xml script where the high- and low- resolution docking steps were removed. The procedure was started by a shell-script containing the following command line. The full script and the bash-script to stat it can be found in the appendix in section 10.5.7.3 and 10.5.7.4.

```
/Path_to_Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@docking.options -parser:protocol docking_minimize.xml -out:suffix
_minimize -nstruct 10 >& docking_minimize.log
```

For the full docking step 1000 trajectories were performed. Some of them had not found a suitable solution during the low-resolution docking and won't produce any output. The output was further evaluated in terms of interface energy and RMSD in comparison to the input structure. The interface energy was calculated similar to the ligand-docking protocols by the InterfaceAnalyzerMover the RMSD was calculated by Rosetta using the RMSD filter. For RMSD determination only residues present in the subunit chains directly located to the interface were included, specified by their residue number in the PDB input file. Respective RosettaScript file is shown in full in the appendix in section 10.5.7.6. Also, additional options file specifying the input structure for the RMSD calculation and further options were prepared (appendix section 10.5.7.6.1). All files were copied in the directory with the docking output structures and started by a bash script by the following line (Complete script in appendix section 10.5.7.5).

```
/Path_to_Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@docking_analysis.options -in:file:s *full*pdb *minimize*pdb >&
docking_analysis.log
```

All calculated metrics were stored in an CSV file. Results were evaluated by plotting the total rosetta score (as a measure for protein stability) and the dG_seperated (as a measure for binding energy) against the RMSD. This resulted in funnel like plots, where the lowest energy can be found by low RMSD values like seen in the two examples shown in Figure 5.49.

For initial runs the structure with the lowest energy during the initial energy- was chosen. However, it was observed that the results of the docking runs divert quite strongly from each other depending on which input structure was selected, even though all share the same amino acid compositions. To account for this the docking were performed for the three best input structures and the best values were selected for comparison with other variants. After the protocol was established, it was applied to all combinations of different protein variants at both interfaces summing up to a total of 24 combinations.

9 Bibliography

- [1] S. Lekawanvijit, A. R. Kompa, H. Krum, *Am. J. Physiol. Renal. Physiol.* **2016**, *311*, F52-62
- [2] B. Bikbov, et al., *The Lancet* **2020**, *395*, 709-733
- [3] A. C. Webster, E. V. Nagler, R. L. Morton, P. Masson, *The Lancet* **2017**, *389*, 1238-1252
- [4] J. Wang, L. Zhang, S. C.-w. Tang, N. Kashihara, Y.-S. Kim, A. Togtokh, C.-w. Yang, M.-h. Zhao, *Kidney int.* **2018**, *94*, 22-25
- [5] T. Liyanage, et al., *The Lancet* **2015**, *385*, 1975-1982
- [6] R. T. Gansevoort, R. Correa-Rotter, B. R. Hemmelgarn, T. H. Jafar, H. J. L. Heerspink, J. F. Mann, K. Matsushita, C. P. Wen, *The Lancet* **2013**, *382*, 339-352
- [7] G. Schlieper, K. Hess, J. Floege, N. Marx, *Nephrol. Dial. Transplant.* **2016**, *31*, 382-390
- [8] S. Ito, M. Yoshida, *Toxins* **2014**, *6*, 665-678
- [9] M. Karbowska, T. W. Kaminski, N. Marcinczyk, T. Misztal, T. Rusak, L. Smyk, D. Pawlak, *Toxins* **2017**, *9*
- [10] L. Dou, E. Bertrand, C. Cerini, V. Faure, J. Sampol, R. Vanholder, Y. Berland, P. Brunet, *Kidney. Int.* **2004**, *65*, 442-451
- [11] M. Yu, Y. J. Kim, D. H. Kang, *Clin. J. Am. Soc. Nephrol.* **2011**, *6*, 30-39
- [12] C. Y. Sun, S. C. Chang, M. S. Wu, *PLoS One* **2012**, *7*, e34026
- [13] J. Holmar, S. de la Puente-Secades, J. Floege, H. Noels, J. Jankowski, S. Orth-Alampour, *Cells* **2020**, *9*
- [14] J. Jansen, J. Jankowski, P. R. Gajjala, J. F. M. Wetzels, R. Masereeuw, *Clin. Sci.* **2017**, *131*, 1631-1647
- [15] R. Vanholder, et al., *Kidney. Int.* **2003**, *63*, 1934-1943
- [16] T. L. Sirich, et al., *Kidney Int.* **2017**, *91*, 1186-1192
- [17] M. Sternkopf, et al., *Toxins* **2019**, *11*, 389
- [18] S. R. Sandeman, et al., *Biomed. Mater.* **2017**, *12*, 035001
- [19] V. Wernert, O. Schäf, H. Ghobarkar, R. Denoyel, *Microporous and Mesoporous Mater.* **2005**, *83*, 101-113
- [20] S. Kato, K. I. Otake, H. Chen, I. Akpınar, C. T. Buru, T. Islamoglu, R. Q. Snurr, O. K. Farha, *J. Am. Chem. Soc.* **2019**, *141*, 2568-2576
- [21] R. Klinke, H. C. Pape, A. Kurtz, S. Silbernagl, *Physiologie*, Georg Thieme Verlag, **2009**.
- [22] A. Levin, et al., *Kidney Int. Suppl.* **2013**, *3*, 1-150
- [23] R. J. Glassock, *J. Ren. Nutr.* **2008**, *18*, 2-6
- [24] F. Durantón, G. Cohen, R. De Smet, M. Rodríguez, J. Jankowski, R. Vanholder, A. Argiles, o. b. o. t. E. U. T. W. Group, *J. Am. Soc. Nephrol.* **2012**, *23*, 1258-1270
- [25] R. Vanholder, A. Pletinck, E. Schepers, G. Glorieux, *Toxins* **2018**, *10*
- [26] V. Saar-Kovrov, W. Zidek, S. Orth-Alampour, D. Fliser, V. Jankowski, E. A. L. Biessen, J. Jankowski, *J. Intern. Med.* **2021**, *290*, 499-526
- [27] E. A. Smith, G. T. Macfarlane, *J. Appl. Microbiol.* **1996**, *81*, 288-302

- [28] T. Gryp, R. Vanholder, M. Vanechoutte, G. Glorieux, *Toxins* **2017**, *9*, 52
- [29] W. R. Clark, N. L. Dehghani, V. Narsimhan, C. Ronco, *Blood Purif.* **2019**, *48*, 299-314
- [30] J. Ludwig, J. Smith, J. Pfaendtner, *J. Phys. Chem. B* **2021**, *125*, 2910-2920
- [31] J. F. Saldanha, D. Yi, M. B. Stockler-Pinto, H. A. Soula, S. Chambert, D. Fouque, D. Mafra, C. O. Soulage, *Biochimie* **2016**, *125*, 53-58
- [32] Y. Itoh, A. Ezawa, K. Kikuchi, Y. Tsuruta, T. Niwa, *Anal. Bioanal. Chem.* **2012**, *403*, 1841-1850
- [33] J. Jankowski, et al., *J. Clin. Invest.* **2003**, *112*, 256-264
- [34] G. Cohen, J. Raupachova, W. H. Hörl, *Nephrol. Dial. Transplant.* **2012**, *28*, 421-429
- [35] S. Schmidt, T. H. Westhoff, P. Krauser, R. Ignatius, J. Jankowski, V. Jankowski, W. Zidek, M. Van der Giet, *Nephrol. Dial. Transplant.* **2008**, *23*, 3485-3493
- [36] R. J. Johnson, J. Feehally, J. Floege, *Comprehensive clinical nephrology*, Elsevier Health Sciences, **2014**.
- [37] G. M. Fleming, *Organogenesis* **2011**, *7*, 2-12
- [38] S. Van Laecke, K. D. Wilde, R. Vanholder, *Artif. Organs* **2006**, *30*, 579-585
- [39] N. Meert, S. Eloot, M. A. Waterloos, M. Van Landschoot, A. Dhondt, G. Glorieux, I. Ledebro, R. Vanholder, *Nephrol. Dial. Transplant.* **2009**, *24*, 562-570
- [40] M. Madero, et al., *Clin. J. Am. Soc. Nephrol.* **2019**, *14*, 394-402
- [41] B. L. Urquhart, D. J. Freeman, J. D. Spence, A. A. House, *Am. J. Kidney. Dis.* **2007**, *49*, 109-117
- [42] A. F. Perna, et al., *J. Ren. Nutr.* **2012**, *22*, 507-514
- [43] S. Yamamoto, et al., *Sci. Rep.* **2015**, *5*, 14381
- [44] M. T. Rocchetti, et al., *Toxins* **2020**, *12*
- [45] F. Brettschneider, et al., *Artif. Organs.* **2013**, *37*, 409-416
- [46] D. Falkenhagen, W. Strobl, G. Vogt, A. Schrefl, I. Linsberger, F. J. Gerner, M. Schoenhofen, *Artif. Organs* **1999**, *23*, 81-86
- [47] K. Rifai, T. Ernst, U. Kretschmer, M. J. Bahr, A. Schneider, C. Hafer, H. Haller, M. P. Manns, D. Fliser, *J. Hepatol.* **2003**, *39*, 984-990
- [48] D. Bergé-Lefranc, C. Vagner, R. Calaf, H. Pizzala, R. Denoyel, P. Brunet, H. Ghobarkar, O. Schäf, *Microporous Mesoporous Mater.* **2012**, *153*, 288-293
- [49] L. Lu, J. T. W. Yeow, *Mater. Des.* **2017**, *120*, 328-335
- [50] D. Pavlenko, D. Giasafaki, G. Charalambopoulou, E. van Geffen, K. G. F. Gerritsen, T. Steriotis, D. Stamatialis, *Sci. Rep.* **2017**, *7*, 14914
- [51] M. S. Tijink, et al., *Biomaterials* **2013**, *34*, 7819-7828
- [52] D. Pavlenko, E. Van Geffen, M. J. Van Steenbergen, G. Glorieux, R. Vanholder, K. G. Gerritsen, D. Stamatialis, *Sci. Rep.* **2016**, *6*, 1-9
- [53] K. A. Cannon, V. N. Nguyen, C. Morgan, T. O. Yeates, *ACS Synth Biol* **2020**, *9*, 517-524
- [54] F. U. Hartl, A. Bracher, M. Hayer-Hartl, *Nature* **2011**, *475*, 324-332
- [55] A. C. MH Glickman *Physiol Rev* **2001**, , 373-428
- [56] C. Chowdhury, S. Sinha, S. Chun, T. O. Yeates, T. A. Bobik, *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 438-468

- [57] T. G. W. Edwardson, M. D. Levasseur, S. Tetter, A. Steinauer, M. Hori, D. Hilvert, *Chem. Rev.* **2022**
- [58] A. Z. Stasiak, E. Larquet, A. Stasiak, S. Müller, A. Engel, E. Van Dyck, S. C. West, E. H. Egelman, *Current Biology* **2000**, *10*, 337-340
- [59] J. G. Heddle, *Nanotechnol Sci Appl* **2008**, *1*, 67
- [60] P. E. Prevelige Jr, J. King, J. L. Silva, *Biophys. Jor.* **1994**, *66*, 1631-1641
- [61] W. M. Aumiller, M. Uchida, T. Douglas, *Chem. Soc. Rev.* **2018**, *47*, 3433-3469
- [62] J. D. Perlmutter, M. F. Hagan, *Annu. Rev. Phys. Chem.* **2015**, *66*, 217-239
- [63] M. F. Hagan, R. Zandi, *Curr Opin Virol* **2016**, *18*, 36-43
- [64] A. Zlotnick, S. Mukhopadhyay, *Trends Microbiol* **2011**, *19*, 14-23
- [65] J. D. Perlmutter, M. R. Perkett, M. F. Hagan, *J. Mol. Biol.* **2014**, *426*, 3148-3165
- [66] M. Uchida, et al., *Advanced Materials* **2007**, *19*, 1025-1042
- [67] M. A. Asensio, N. M. Morella, C. M. Jakobson, E. C. Hartman, J. E. Glasgow, B. Sankaran, P. H. Zwart, D. Tullman-Ercek, *Nano Lett.* **2016**, *16*, 5944-5950
- [68] B. Sana, E. Johnson, P. Le Magueres, A. Criswell, D. Cascio, S. Lim, *J Biol Chem* **2013**, *288*, 32663-32672
- [69] B. Wörsdörfer, K. J. Woycechowsky, D. Hilvert, *Science* **2011**, *331*, 589-592
- [70] E. Sasaki, D. Bohringer, M. van de Waterbeemd, M. Leibundgut, R. Zschoche, A. J. Heck, N. Ban, D. Hilvert, *Nat Commun* **2017**, *8*, 14663
- [71] Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson, T. O. Yeates, *Nat Chem* **2014**, *6*, 1065-1071
- [72] J. E. Padilla, C. Colovos, T. O. Yeates, *PNAS* **2001**, *98*, 2217-2221
- [73] F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. Andre, *PLoS One* **2011**, *6*, e20450
- [74] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates, D. Baker, *Science* **2012**, *336*, 1171-1174
- [75] Y. Hsia, et al., *Nature* **2016**, *535*, 136-139
- [76] J. Lapointe, D. A. Marvin, *Molecular Crystals and Liquid Crystals* **2007**, *19*, 269-278
- [77] A. Korpi, E. Anaya-Plaza, S. Valimaki, M. Kostianen, *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2020**, *12*, e1578
- [78] V. Liljestrom, et al., *Nat. Commun.* **2017**, *8*, 671
- [79] M. A. Kostianen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen, P. Ceci, *Nat Nanotechnol* **2013**, *8*, 52-56
- [80] V. Liljestrom, J. Seitsonen, M. A. Kostianen, *ACS Nano* **2015**, *9*, 11278-11285
- [81] S. Palchoudhury, Z. Zhou, K. Ramasamy, F. Okirie, P. E. Prevelige, A. Gupta, *J. Mater. Res.* **2016**, *32*, 465-472
- [82] N. E. Brunk, M. Uchida, B. Lee, M. Fukuto, L. Yang, T. Douglas, V. Jadhao, *ACS Appl. Bio. Mater.* **2019**, *2*, 2192-2201
- [83] T. Li, R. E. Winans, B. Lee, *Langmuir* **2011**, *27*, 10929-10937
- [84] R. Asor, O. Ben-Nun-Shaul, A. Oppenheim, U. Raviv, *ACS Nano* **2017**, *11*, 9814-9824

- [85] A. Korpi, C. Ma, K. Liu, Nonappa, A. Herrmann, O. Ikkala, M. A. Kostiainen, *ACS Macro Lett* **2018**, *7*, 318-323
- [86] B. D. Tiu, D. L. Kernan, S. B. Tiu, A. M. Wen, Y. Zheng, J. K. Pokorski, R. C. Advincula, N. F. Steinmetz, *Nanoscale* **2017**, *9*, 1580-1590
- [87] M. Uchida, et al., *ACS Nano* **2018**, *12*, 942-953
- [88] T. Beck, S. Tetter, M. Künzle, D. Hilvert, *Angew. Chem., Int. Ed.* **2015**, *127*, 951-954
- [89] M. Künzle, T. Eckert, T. Beck, *J. Am. Chem. Soc.* **2016**, *138*, 12731-12734
- [90] H. Moon, W. G. Kim, S. Lim, Y. J. Kang, H. H. Shin, H. Ko, S. Y. Hong, S. Kang, *J Mater Chem B* **2013**, *1*, 4504-4510
- [91] S. Sim, D. Miyajima, T. Niwa, H. Taguchi, T. Aida, *J. Am. Chem. Soc.* **2015**, *137*, 4658-4661
- [92] S. Sim, T. Niwa, H. Taguchi, T. Aida, *J. Am. Chem. Soc.* **2016**, *138*, 11152-11155
- [93] J. B. Bailey, L. Zhang, J. A. Chiong, S. Ahn, F. A. Tezcan, *J. Am. Chem. Soc.* **2017**, *139*, 8160-8166
- [94] M. G. Joyce, et al., *Cell Rep* **2021**, *37*, 110143
- [95] J. M. Carmen, et al., *NPJ Vaccines* **2021**, *6*, 151
- [96] M. G. Joyce, et al., *Science translational medicine* **2021**, eabi5735
- [97] H. Ansuini, C. Cicchini, A. Nicosia, M. Tripodi, R. Cortese, A. Luzzago, *Nucleic Acids Res.* **2002**, *30*, e78-e78
- [98] S. Bonetto, I. Carlván, D. Baty, *Peptides* **2005**, *26*, 2302-2313
- [99] C. E. Flynn, C. Mao, A. Hayhurst, J. L. Williams, G. Georgiou, B. Iverson, A. M. Belcher, *J. Mater. Chem.* **2003**, *13*, 2414-2421
- [100] R. R. Naik, L. L. Brott, S. J. Clarson, M. O. Stone, *J. Nanosci. Nanotechnol.* **2002**, *2*, 95-100
- [101] T. K. M. Nguyen, M. R. Ki, C. S. Lee, S. P. Pack, *Journal of Industrial and Engineering Chemistry* **2019**, *73*, 198-204
- [102] M. A. Bartel, J. R. Weinstein, D. V. Schaffer, *Gene Ther.* **2012**, *19*, 694-700
- [103] S. Valsesia-Wittmann, A. Drynda, G. Deleage, M. Aumailley, J.-M. Heard, O. Danos, G. Verdier, F.-L. Cosset, *J. Virol.* **1994**, *68*, 4609-4619
- [104] V. Biermann, C. Volpers, S. Hussmann, A. Stock, H. Kewes, G. Schiedner, A. Herrmann, S. Kochanek, *Hum. Gene Ther.* **2001**, *12*, 1757-1769
- [105] G. Destito, R. Yeh, C. S. Rae, M. G. Finn, M. Manchester, *Chem. Biol.* **2007**, *14*, 1152-1162
- [106] V. A. Belyi, M. Muthukumar, *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 17174-17178
- [107] J. A. Jones, T. W. Giessen, *Biotechnol Bioeng* **2021**, *118*, 491-505
- [108] Y. Azuma, R. Zschoche, D. Hilvert, *J. Biol. Chem.* **2017**, *292*, 10321-10327
- [109] P. L. Chariou, N. F. Steinmetz, *ACS Nano* **2017**, *11*, 4719-4730
- [110] M. Liang, K. Fan, M. Zhou, D. Duan, J. Zheng, D. Yang, J. Feng, X. Yan, *Proc. Natl. Acad. Sci.* **2014**, *111*, 14900-14905
- [111] B. Jiang, L. Yan, J. Zhang, M. Zhou, G. Shi, X. Tian, K. Fan, C. Hao, X. Yan, *ACS Appl. Mater. Interfaces.* **2019**, *11*, 9747-9755
- [112] S. Abe, K. Hirata, T. Ueno, K. Morino, N. Shimizu, M. Yamamoto, M. Takata, E. Yashima, Y. Watanabe, *J. Am. Chem. Soc.* **2009**, *131*, 6958-6960

- [113] Y. H. Chung, H. Cai, N. F. Steinmetz, *Adv. Drug. Deliv. Rev.* **2020**, *156*, 214-235
- [114] C. Aussignargues, B. C. Paasch, R. Gonzalez-Esquer, O. Erbilgin, C. A. Kerfeld, *Commun. Integr. Biol.* **2015**, *8*, e1039755
- [115] P. G. Stockley, et al., *J Biol Phys* **2013**, *39*, 277-287
- [116] T. Douglas, M. Young, *Nature* **1998**, *393*, 152-155
- [117] J. Sun, et al., *Proc. Natl. Acad. Sci.* **2007**, *104*, 1354-1359
- [118] M. Lach, C. Strelow, A. Meyer, A. Mews, T. Beck, *ACS Appl. Mater. Interfaces* **2022**, *14*, 10656-10668
- [119] M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, *Nat Struct Mol Biol* **2008**, *15*, 939-947
- [120] W. J. Altenburg, N. Rollins, P. A. Silver, T. W. Giessen, *Sci. Rep.* **2021**, *11*, 4951
- [121] C. E. Ashley, et al., *ACS Nano* **2011**, *5*, 5729-5745
- [122] J. E. Glasgow, S. L. Capehart, M. B. Francis, D. Tullman-Ercek, *ACS Nano* **2012**, *6*, 8658-8664
- [123] T. W. Giessen, P. A. Silver, *ChemBioChem* **2016**, *17*, 1931-1935
- [124] J. Cao, R. H. Guenther, T. L. Sit, S. A. Lommel, C. H. Opperman, J. A. Willoughby, *ACS Appl. Mater. Interfaces.* **2015**, *7*, 9546-9553
- [125] M. Loredo-Tovias, et al., *Nanoscale* **2017**, *9*, 11625-11631
- [126] L. Loo, R. H. Guenther, S. A. Lommel, S. Franzen, *Chem. Commun.* **2008**, 88-90
- [127] M. Brasch, A. de la Escosura, Y. Ma, C. Uetrecht, A. J. Heck, T. Torres, J. J. Cornelissen, *J. Am. Chem. Soc.* **2011**, *133*, 6878-6881
- [128] M. Truffi, L. Fiandra, L. Sorrentino, M. Monieri, F. Corsi, S. Mazzucchelli, *Pharmacol. Res.* **2016**, *107*, 57-65
- [129] D. C. Buehler, D. B. Toso, V. A. Kickhoefer, Z. H. Zhou, L. H. Rome, *Small* **2011**, *7*, 1432-1439
- [130] C. B. Chang, C. M. Knobler, W. M. Gelbart, T. G. Mason, *ACS Nano* **2008**, *2*, 281-286
- [131] T. G. Edwardson, S. Tetter, D. Hilvert, *Nat. Commun.* **2020**, *11*, 1-9
- [132] P. D. Garimella, A. Datta, D. W. Romanini, K. N. Raymond, M. B. Francis, *J. Am. Chem. Soc.* **2011**, *133*, 14704-14709
- [133] W. Wu, S. C. Hsiao, Z. M. Carrico, M. B. Francis, *Angew. Chem. Int. Ed.* **2009**, *48*, 9493-9497
- [134] S. Wang, A. T. Al-Soodani, G. C. Thomas, B. A. Buck-Koehntop, K. J. Woycechowsky, *Bioconj. Chem.* **2018**, *29*, 2332-2342
- [135] Q. Guo, G. Thomas, K. Woycechowsky, *RSC Adv.* **2017**, *7*, 34676-34686
- [136] L. Brillault, P. V. Jutras, N. Dashti, E. C. Thuenemann, G. Morgan, G. P. Lomonosoff, M. J. Landsberg, F. Sainsbury, *ACS Nano* **2017**, *11*, 3476-3484
- [137] J. M. Hooker, E. W. Kovacs, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 3718-3719
- [138] S. C. Andrews, *Adv. Microb. Physiol.* **1998**, *40*, 281-351
- [139] Y. Zhang, B. P. Orner, *Int. J. Mol. Sci.* **2011**, *12*, 5406-5421
- [140] G. C. Ford, P. M. Harrison, D. W. Rice, J. M. Smith, A. Treffry, J. L. White, J. Yariv, *Philos Trans R Soc Lond B Biol Sci* **1984**, *304*, 551-565
- [141] G. A. Clegg, J. E. Fitton, P. M. Harrison, A. Treffry, *Prog. Biophys. Mol. Biol.* **1981**, *36*, 53-86

- [142] S. C. Andrews, *Biochim. Biophys. Acta* **2010**, *1800*, 691-705
- [143] M. Gerl, R. Jaenicke, *Eur. Biophys. J.* **1987**, *15*, 103-109
- [144] D. Sato, H. Ohtomo, Y. Yamada, T. Hikima, A. Kurobe, K. Fujiwara, M. Ikeguchi, *Biochemistry* **2016**, *55*, 287-293
- [145] S. J. Y. Paul D. Hempstead, Alisdair R. Fernie David M. Lawson, Peter J. Artymiuk, David W. Rice, Geoffrey C. Ford and Pauline M. Harris, *J. Mol. Biol.* **1997**, *268*, 424-448
- [146] R. R. Crichton, J. P. Declercq, *Biochim. Biophys. Acta.* **2010**, *1800*, 706-718
- [147] T. Masuda, F. Goto, T. Yoshihara, B. Mikami, *J. Biol. Chem.* **2010**, *285*, 4049-4059
- [148] C. Lv, Y. Bai, S. Yang, G. Zhao, B. Chen, *Food Chem.* **2013**, *141*, 3851-3858
- [149] E. C. Theil, X. S. Liu, T. Tosha, *Inorg. Chim. Acta.* **2008**, *361*, 868-874
- [150] T. Takahashi, S. Kuyucak, *Biophys. J.* **2003**, *84*, 2256-2263
- [151] I. Bertini, D. Lalli, S. Mangani, C. Pozzi, C. Rosa, E. C. Theil, P. Turano, *J. Am. Chem. Soc.* **2012**, *134*, 6169-6176
- [152] T. Masuda, F. Goto, T. Yoshihara, B. Mikami, *Biochem. Biophys. Res. Commun.* **2010**, *400*, 94-99
- [153] P. M. Harrison, P. Arosio, *Biochim. Biophys. Acta, Bioenerg.* **1996**, *1275*, 161-203
- [154] T. Tosha, H. L. Ng, O. Bhattasali, T. Alber, E. C. Theil, *J. Am. Chem. Soc.* **2010**, *132*, 14562-14569
- [155] T. Douglas, D. R. Ripoll, *Protein Sci.* **1998**, *7*, 1083-1091
- [156] D. E. Mayer, J. S. Rohrer, D. A. Schoeller, D. C. Harris, *Biochemistry* **1983**, *22*, 876-880
- [157] M. Uchida, S. Kang, C. Reichhardt, K. Harlen, T. Douglas, *Biochim. Biophys. Acta.* **2010**, *1800*, 834-845
- [158] G. Zhao, F. Bou-Abdallah, P. Arosio, S. Levi, C. Janus-Chandler, N. D. Chasteen, *Biochemistry* **2003**, *42*, 3142-3150
- [159] X. Liu, E. C. Theil, *Acc. Chem. Res.* **2005**, *38*, 167-175
- [160] N. Song, J. Zhang, J. Zhai, J. Hong, C. Yuan, M. Liang, *Acc. Chem. Res.* **2021**, *54*, 3313-3325
- [161] X. Li, L. Qiu, P. Zhu, X. Tao, T. Imanaka, J. Zhao, Y. Huang, Y. Tu, X. Cao, *Small* **2012**, *8*, 2505-2514
- [162] T. R. Daniels, T. Delgado, J. A. Rodriguez, G. Helguera, M. L. Penichet, *Clin. Immunol.* **2006**, *121*, 144-158
- [163] L. Li, et al., *Proc. Natl. Acad. Sci.* **2010**, *107*, 3505-3510
- [164] M. Budiarta, W. Xu, L. Schubert, M. Meledina, A. Meledin, D. Wöll, A. Pich, T. Beck, *J. Colloid Interface Sci.* **2021**, *591*, 451-462
- [165] P. Sánchez, E. Valero, N. Gálvez, J. M. Domínguez-Vera, M. Marinone, G. Poletti, M. Corti, A. Lascialfari, *Dalton Trans.* **2009**, 800-804
- [166] Y. Zhao, et al., *ACS Nano* **2016**, *10*, 4184-4191
- [167] Z. Wang, et al., *ACS Nano* **2016**, *10*, 3453-3460
- [168] H. Wu, M. H. Engelhard, J. Wang, D. R. Fisher, Y. Lin, *J. Mater. Chem.* **2008**, *18*, 1779-1783
- [169] T. Tan, et al., *Adv. Sci.* **2018**, *5*, 1801012
- [170] C. Gu, T. Zhang, C. Lv, Y. Liu, Y. Wang, G. Zhao, *ACS Nano* **2020**, *14*, 17080-17090

- [171] L. Li, M. Muñoz-Culla, U. Carmona, M. P. Lopez, F. Yang, C. Trigueros, D. Otaegui, L. Zhang, M. Knez, *Biomaterials* **2016**, *98*, 143-151
- [172] A. Macone, S. Masciarelli, F. Palombarini, D. Quaglio, A. Boffi, M. C. Trabuco, P. Baiocco, F. Fazi, A. Bonamore, *Sci. Rep.* **2019**, *9*, 1-7
- [173] S. Tetter, D. Hilvert, *Angew. Chem. Int. Ed.* **2017**, *56*, 14933-14936
- [174] F. Palombarini, et al., *J. Nanobiotechnol.* **2021**, *19*, 1-12
- [175] P. Ochtrop, C. P. R. Hackenberger, *Curr. Opin. Chem. Biol.* **2020**, *58*, 28-36
- [176] O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174-2195
- [177] P. A. Szijj, K. A. Kostadinova, R. J. Spears, V. Chudasama, *Org. Biomol. Chem.* **2020**, *18*, 9018-9028
- [178] L. M. van der Sleen, K. M. Tych, *Nanomaterials* **2021**, *11*, 2424
- [179] J. M. Chalker, G. J. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, *4*, 630-640
- [180] N. Lundell, T. Schreitmüller, *Anal. Biochem.* **1999**, *266*, 31-47
- [181] G. T. Hermanson, in *Bioconjugate Techniques (Third Edition)* (Ed.: G. T. Hermanson), Academic Press, Boston, **2013**, pp. 229-258.
- [182] B. H. Northrop, S. H. Frayne, U. Choudhary, *Polym. Chem.* **2015**, *6*, 3415-3430
- [183] N. E. Sharpless, M. Flavin, *Biochemistry* **1966**, *5*, 2963-2971
- [184] O. Barbosa, C. Ortiz, Á. Berenguer-Murcia, R. Torres, R. C. Rodrigues, R. Fernandez-Lafuente, *RSC Adv.* **2014**, *4*, 1583-1600
- [185] F. M. Richards, J. R. Knowles, *J. Mol. Biol.* **1968**, *37*, 231-233
- [186] I. Migneault, C. Dartiguenave, M. J. Bertrand, K. C. Waldron, *BioTechniques* **2004**, *37*, 790-802
- [187] S. Mathapati, D. K. Bishi, S. Guhathakurta, K. M. Cherian, J. R. Venugopal, S. Ramakrishna, R. S. Verma, *Mater. Sci. Eng. C.* **2013**, *33*, 1561-1572
- [188] Y. Liu, L. Ma, C. Gao, *Mater. Sci. Eng. C.* **2012**, *32*, 2361-2366
- [189] A. F. S. A. Habeeb, *Arch. Biochem. Biophys.* **1967**, *119*, 264-268
- [190] E. F. Jansen, A. C. Olson, *Arch. Biochem. Biophys.* **1969**, *129*, 221-227
- [191] H. K. Manonmani, R. Joseph, *Process Biochem.* **1993**, *28*, 325-329
- [192] D. R. Walt, V. I. Agayn, *TrAC, Trends Anal. Chem.* **1994**, *13*, 425-430
- [193] J. Koehler Leman, et al., *PLoS Comp. Biol.* **2020**, *16*, e1007507
- [194] J. K. Leman, et al., *Nat. Methods* **2020**, *17*, 665-680
- [195] M. S. Lin, T. Head-Gordon, *J. Chem. Theory Comput.* **2008**, *4*, 515-521
- [196] T. Jacobs, B. Williams, T. Williams, X. Xu, A. Eletsy, J. Federizon, T. Szyperski, B. Kuhlman, *Science* **2016**, *352*, 687-690
- [197] S. L. Guffy, F. D. Teets, M. I. Langlois, B. Kuhlman, *J. Chem. Inf. Model.* **2018**, *58*, 895-901
- [198] A. M. Watkins, C. Geniesse, W. Kladwang, P. Zakrevsky, L. Jaeger, R. Das, *Sci. Adv.* **2018**, *4*, eaar5316
- [199] P. Sripakdeevong, W. Kladwang, R. Das, *Proc. Natl. Acad. Sci.* **2011**, *108*, 20573-20578
- [200] J. Adolf-Bryfogle, O. Kalyuzhniy, M. Kubitz, B. D. Weitzner, X. Hu, Y. Adachi, W. R. Schief, R. L. Dunbrack Jr, *PLoS Comput. Biol.* **2018**, *14*, e1006112

- [201] D. Baran, M. G. Pszolla, G. D. Lapidoth, C. Norn, O. Dym, T. Unger, S. Albeck, M. D. Tyka, S. J. Fleishman, *Proc. Natl. Acad. Sci.* **2017**, *114*, 10900-10905
- [202] G. Lapidoth, J. Parker, J. Prilusky, S. J. Fleishman, *Bioinformatics* **2019**, *35*, 1591-1593
- [203] C. H. Norn, G. Lapidoth, S. J. Fleishman, *Proteins: Struct., Funct., Bioinf.* **2017**, *85*, 30-38
- [204] N. A. Marze, S. S. Roy Burman, W. Sheffler, J. J. Gray, *Bioinformatics* **2018**, *34*, 3461-3469
- [205] S. DeLuca, K. Khar, J. Meiler, *PLoS One* **2015**, *10*, e0132508
- [206] I. W. Davis, D. Baker, *J. Mol. Biol.* **2009**, *385*, 381-392
- [207] C. B. Anfinsen, *Science* **1973**, *181*, 223-230
- [208] D. Baker, J. L. Sohl, D. A. Agard, *Nature* **1992**, *356*, 263-265
- [209] R. F. Alford, et al., *J. Chem. Theory Comput.* **2017**, *13*, 3031-3048
- [210] R. Das, D. Baker, *Annu. Rev. Biochem.* **2008**, *77*, 363-382
- [211] A. Leaver-Fay, et al., in *Methods Enzymol.*, Vol. 487 (Eds.: M. L. Johnson, L. Brand), Academic Press, **2011**, pp. 545-574.
- [212] RosettaCommons,
https://www.rosettacommons.org/docs/latest/rosetta_basics/structural_concepts/Internal-coordinates Accessed on: 05.06.2022
- [213] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, J. Meiler, *Biochemistry* **2010**, *49*, 2987-2998
- [214] RosettaCommons,
https://www.rosettacommons.org/docs/latest/rosetta_basics/structural_concepts/Mover Accessed on 06.06.2022
- [215] B. J. Bender, et al., *Biochemistry* **2016**, *55*, 4748-4763
- [216] A. Leaver-Fay, et al., in *Methods Enzymol.*, Vol. 523, Elsevier, **2013**, pp. 109-143.
- [217] K. T. Simons, C. Kooperberg, E. Huang, D. Baker, *J. Mol. Biol.* **1997**, *268*, 209-225
- [218] K. T. Simons, I. Ruczinski, C. Kooperberg, B. A. Fox, C. Bystroff, D. Baker, *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 82-95
- [219] A. Warshel, S. T. Russell, *Q. Rev. Biophys.* **1984**, *17*, 283-422
- [220] R. L. Baldwin, *J. Mol. Biol.* **2007**, *371*, 283-301
- [221] T. Lazaridis, M. Karplus, *Proteins: Struct., Funct., Bioinf.* **1999**, *35*, 133-152
- [222] R. E. Hubbard, M. Kamran Haider, in *Encyclopedia of Life Sciences*, John Wiley & Sons, Chichester, U.K., **2010**.
- [223] T. Kortemme, A. V. Morozov, D. Baker, *J. Mol. Biol.* **2003**, *326*, 1239-1259
- [224] M. J. O'Meara, et al., *J. Chem. Theory Comput.* **2015**, *11*, 609-622
- [225] J. S. Richardson, D. A. Keedy, D. C. Richardson, in *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*, World Scientific, **2013**, pp. 46-61.
- [226] G. Wang, R. L. Dunbrack Jr, *Bioinformatics* **2003**, *19*, 1589-1591
- [227] A. V. Finkelstein, A. Y. Badretdinov, A. M. Gutin, *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 142-150
- [228] D. Shortle, *Protein Sci.* **2003**, *12*, 1298-1302

- [229] M. V. Shapovalov, R. L. Dunbrack Jr, *Structure* **2011**, *19*, 844-858
- [230] H. Park, P. Bradley, P. Greisen Jr, Y. Liu, V. K. Mulligan, D. E. Kim, D. Baker, F. DiMaio, *J. Chem. Theory Comput.* **2016**, *12*, 6201-6212
- [231] A. E. Miklos, et al., *Chem. Biol.* **2012**, *19*, 449-455
- [232] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* **2003**, *302*, 1364-1368
- [233] D. Huson, O. Kohlbacher, A. Lupas, K. Nieselt, A. Zell, *German Conference on Bioinformatics, Gesellschaft für Informatik eV*, **2006**.
- [234] J. Meiler, D. Baker, *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 538-548
- [235] R. Moretti, B. J. Bender, B. Allison, J. Meiler, in *Computational Design of Ligand Binding Proteins*, Springer, **2016**, pp. 47-62.
- [236] S. T. Smith, J. Meiler, *PloS one* **2020**, *15*, e0240450
- [237] L. Hu, X. Wang, Y.-A. Huang, P. Hu, Z.-H. You, *Brief. Bioinformatics* **2021**, *22*
- [238] H. Park, H. Lee, C. Seok, *Curr. Opin. Struct. Biol.* **2015**, *35*, 24-31
- [239] S. Chaudhury, M. Berrondo, B. D. Weitzner, P. Muthu, H. Bergman, J. J. Gray, *PLOS ONE* **2011**, *6*, e22477
- [240] I. Rayment, *Structure* **2002**, *10*, 147-151
- [241] R. Fraczkiwicz, W. Braun, *J. Comput. Chem.* **1998**, *19*, 319-333
- [242] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, in *The proteomics protocols handbook*, John M. Walker, **2005**, pp. 571-607.
- [243] G. D. Martich, A. J. Boujoukos, A. F. Suffredini, *Immunobiology* **1993**, *187*, 403-416
- [244] D. S. Woulfe, *Expert Rev. Hematol.* **2010**, *3*, 81-91
- [245] C. Tripisciano, T. Eichhorn, S. Harm, V. Weber, *Biomed Res. Int.* **2014**, *2014*, 238160
- [246] K. Kubiak-Ossowska, B. Jachimska, P. A. Mulheran, *J. Phys. Chem. B* **2016**, *120*, 10463-10468
- [247] Y. L. Jeyachandran, E. Mielczarski, B. Rai, J. A. Mielczarski, *Langmuir* **2009**, *25*, 11614-11620
- [248] I. O. f. Standardization, in *Part 4: Selection of Tests for Interactions with Blood Vol. ISO 10993-4:2017*, DIN Deutsches Institut für Normung e. V. Beuth Verlag GmbH, Berlin, Germany, **2017**.
- [249] F. C. Barreto, et al., *Clin J Am Soc Nephrol* **2009**, *4*, 1551-1558
- [250] M. Hida, Y. Aiba, S. Sawamura, N. Suzuki, T. Satoh, Y. Koga, *Nephron* **1996**, *74*, 349-355
- [251] L. Kastl, et al., *ACS Nano* **2013**, *7*, 6605-6618
- [252] C. Balut, S. Despa, I. Lambrichts, M. Ameloot, P. Steels, I. Smets, *Kidney Int.* **2008**, *73*, 226-232
- [253] C. S. Pillay, E. Elliott, C. Dennison, *Biochem. J.* **2002**, *363*, 417-429
- [254] F. Zhang, Z. Ali, F. Amin, A. Feltz, M. Oheim, W. J. Parak, *ChemPhysChem* **2010**, *11*, 730-735
- [255] F. Zhang, E. Lees, F. Amin, P. Rivera_Gil, F. Yang, P. Mulvaney, W. J. Parak, *Small* **2011**, *7*, 3113-3127
- [256] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res.* **2004**, *32*, W665-W667
- [257] P. A. Sigala, E. A. Ruben, C. W. Liu, P. M. B. Piccoli, E. G. Hohenstein, T. J. Martínez, A. J. Schultz, D. Herschlag, *J. Am. Chem. Soc.* **2015**, *137*, 5730-5740

- [258] E. M. Williams, S. M. Jung, J. L. Coffman, S. Lutz, *ACS Synthetic Biology* **2018**, *7*, 2514-2517
- [259] L. S. R. Adamson, et al., *Science Advances* **2022**, *8*, eabl7346
- [260] B. Ahn, S.-G. Lee, H. R. Yoon, J. M. Lee, H. J. Oh, H. M. Kim, Y. Jung, *Angew. Chem. Int. Ed.* **2018**, *57*, 2909-2913
- [261] S. Levi, P. Santambrogio, B. Corsi, A. Cozzi, P. Arosio, *Biochem. J.* **1996**, *317*, 467-473
- [262] S. Haldar, L. E. Bevers, T. Tosha, E. C. Theil, *J. Biol. Chem.* **2011**, *286*, 25620-25627
- [263] C. Pozzi, F. Di Pisa, D. Lalli, C. Rosa, E. Theil, P. Turano, S. Mangani, *Acta Crystallographica Section D* **2015**, *71*, 941-953
- [264] K. J. Cho, H. J. Shin, J.-H. Lee, K.-J. Kim, S. S. Park, Y. Lee, C. Lee, S. S. Park, K. H. Kim, *J. Mol. Biol.* **2009**, *390*, 83-98
- [265] C. Lv, S. Zhang, J. Zang, G. Zhao, C. Xu, *Biochemistry* **2014**, *53*, 2232-2241
- [266] C. Bernacchioni, V. Ghini, E. C. Theil, P. Turano, *RSC Advances* **2016**, *6*, 21219-21227
- [267] C. Pozzi, F. Di Pisa, C. Bernacchioni, S. Ciambellotti, P. Turano, S. Mangani, *Acta Crystallogr., Sect. D* **2015**, *71*, 1909-1920
- [268] F. Bou-Abdallah, *Biochim. Biophys. Acta, Gen. Subj.* **2010**, *1800*, 719-731
- [269] Y. Shi, H. Tian, Y. Wang, Y. Shen, Q. Zhu, F. Ding, *Blood Purif.* **2019**, *47*, 351-360
- [270] R. Esmailbeiki, K. Krawczyk, B. Knapp, J.-C. Nebel, C. M. Deane, *Brief. Bioinformatics* **2015**, *17*, 117-131
- [271] H. C. Chen, et al., *ACS Appl. Mater. Interfaces.* **2016**, *8*, 19691-19700
- [272] P. A. Aronov, F. J.-G. Luo, N. S. Plummer, Z. Quan, S. Holmes, T. H. Hostetter, T. W. Meyer, *J. Am. Soc. Nephrol.* **2011**, *22*, 1769-1776
- [273] W. R. Wikoff, A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters, G. Siuzdak, *Proc. Natl. Acad. Sci.* **2009**, *106*, 3698-3703
- [274] W. Wang, B. A. Malcolm, *BioTechniques* **1999**, *26*, 680-682
- [275] M. D. Winn, et al., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235-242
- [276] W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125-132
- [277] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674
- [278] P. Emsley, K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126-2132
- [279] F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus, G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2017**, *73*, 112-122
- [280] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355-367
- [281] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12-21
- [282] M. A. Miteva, F. Guyon, P. Tuffiè $\frac{1}{2}$ ry, *Nucleic Acids Res.* **2010**, *38*, W622-W627
- [283] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774-797

10 Appendix

10.1 Amino acid sequences of proteins fabricated in the thesis

All variant are based on the negative-supercharged ferritin variant developed in prior work. [89]

Ftn^(neg) [89]

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSHEE
REHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**C**DFIETHYLNEQ**V**KAI**K**ELGDHVTNLRKMGAPESGLAEYLFDKHTLGDS**D**NES

Ftn^(neg)-1xCys: K53C, C130A

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNF**A**CYFLHQSHEE
REHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**A**DFIETHYLNEQ**V**KAI**K**ELGDHVTNLRKMGAPESGLAEYLFDKHTLGDS**D**NES

Ftn^(neg)-2xCys: K53C, E64C, C130A,

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNF**A**CYFLHQSHEE
RCHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**A**DFIETHYLNEQ**V**KAI**K**ELGDHVTNLRKMGAPESGLAEYLFDKHTLGDS**D**NES

Ftn^(neg)-3xCys: K53C, E64C, C130A, K143C

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNF**A**CYFLHQSHEE
RCHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**A**DFIETHYLNEQ**V****C**A**I**K**E**LGDHVTNLRKMGAPESGLAEYLFDKHTLGDS**D**NES

Ftn^(neg)-4xCys: K53C, E64C, C130A, K143C, S178C

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNF**A**CYFLHQSHEE
RCHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**A**DFIETHYLNEQ**V****C**A**I**K**E**LGDHVTNLRKMGAPESGLAEYLFDKHTLG**D****C**DNES

Ftn^(neg)-Ap4: E61V, E62A, D131F, D140W

TTASTSQVRQNYHQDSEEA**IN**RQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQS**HVA**
REHA**E**KLMKLQNQRGGRIFLQDIQ**P**DEDDWESGLNAME**E**EA**E**LEK**N**VNQSLLELHKLATD
KNDPHL**C**FFIETHYL**NW**Q**V**KAI**K**ELGDHVTNLRKMGAPESGLAEYLFDKHTLGDS**D**NES

Ftn^(neg)-Ap4-3A: E61V, E62A, T122A, D123A, N125A D131F, D140W

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSHVA
REHAEKLMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLAAA
KADPHLCDFIETHYLNWQVKAIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-Ap7: E61V, E62A, H128F, D131W, N139V, E140W, K143V

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSHVA
REHAEKLMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLATD
KNDPFLCWFIETHYLWVQVVAIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-Ap16: K49V, Y54F, H57W, Q58L, E61V, E62A, H65L, H128F, H136Y, D131W, H136Y
N139V, E140W, K143V, E147L, H151Y, N154A

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALVNF~~AK~~FFLWLSHVA
RELAEKLMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLATD
KNDPFLCWFIETY~~YL~~WVQVVAIKLLGDYVTALR~~R~~KMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-dock03: E64R, H65K, K68E, K71R, Q75D, R76K, H128E, D131E, F132H, T135K,
H136R, Y137H, N139R, E140R

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSH~~EE~~
RRKAEELMRLQNDKGGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLATD
KNDPELCEHIEKRHLRRQVKAIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-dock23: H57E, Q58R, H60I, E61G, E64G, K68D, T135D, H136K, Y137H, N139E,
E140K, A144N

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLERSIGE
RGHAEDLMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLATD
KNDPHLCDFIEDKHLEKQVKNIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-dock43: H57E, Q58R, E64R, H65K, K68E, T135D, H136R, Y137H, N139R, E140R,
E147Y

TTASTSQVRQNYHQDSEEAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLERSH~~EE~~
RRKAEELMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEEAELEKQNVNQSLELHKLATD
KNDPHLCDFIEDRHLRRQVKAIKYLG~~H~~VTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Ftn^(neg)-Pocket: A30S, Y54R, E55A, H57I, Q58I, E61K, E62D, A103S, E107D, K137R, A138Y, E146R, L147A, H151R

TTASTSQVRQNYHQDSEEAINRQINLELY**SSYVYLSMSYYFDRDDVALKNFAKRALIASHKD**
REHAEKLMKLNQRGGRIFLQDIQKPDEDDWESGLNAMEESLELDKKNVQSLLELHKLATD
KNDPHLCDFIETHYLNEQV**RYIKRAGDRVTNLRKMGAPESGLAEYLFDKHTLGDS**DNES

10.2 Additional information for chemical modification part

Table 10.1: Solvent accessible surface area of amino acids at Ftn^(neg) inner surface. The ratio of the sidechain to the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations for the selected positions. Calculations were done with the webservice GETAREA^[241] The calculations are based on the crystal structure (PDB ID: 5JKK). If the position is located in an interface between subunits was checked with the PDBePISA tool.^[283]

Position	Amino acid	SASA	At interface
	in Ftn ^(neg)	Ratio [%]	
52	A	17.3	yes
53	K	83.4	
54	Y	13.6	
58	Q	21.9	
61	E	36.2	
62	E	5.2	
65	H	14.8	
68	K	44.4	
69	L	0	
72	L	0.1	
75	Q	51.5	yes
128	H	52.6	
131	D	68.3	
136	H	38.7	
139	N	70.6	
140	E	43.7	
143	K	59.5	
146	K	77.5	
147	E	50	
150	D	46.8	
151	H	10.4	

Table 10.2: Solvent-accessible surface area. The ratio of the sidechain to the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations for the selected positions. Calculations were done with the webservice GETAREA^[241] The calculations are based on the crystal structure (PDB ID: 5JJK). The position S178 is near the end of the ferritin subunit and because of the lacking electron density not modeled, therefore the accessible surface area could not be calculated on this position.

Amino acid	Ratio [%]
K53	78.1
E64	65.7
K143	53.2
S178	N.A

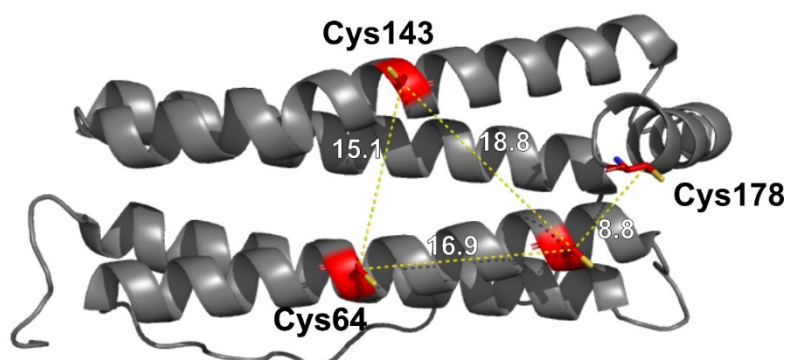


Figure 10.1: Distances between reactive cysteine sites. Distance between reactive cysteine sites given in Angstrom. Since position 178 is not modelled in respective crystal structure due a lack of electron density position 176 was selected for this illustration. As a result, the distance is probably larger in the actual structure.

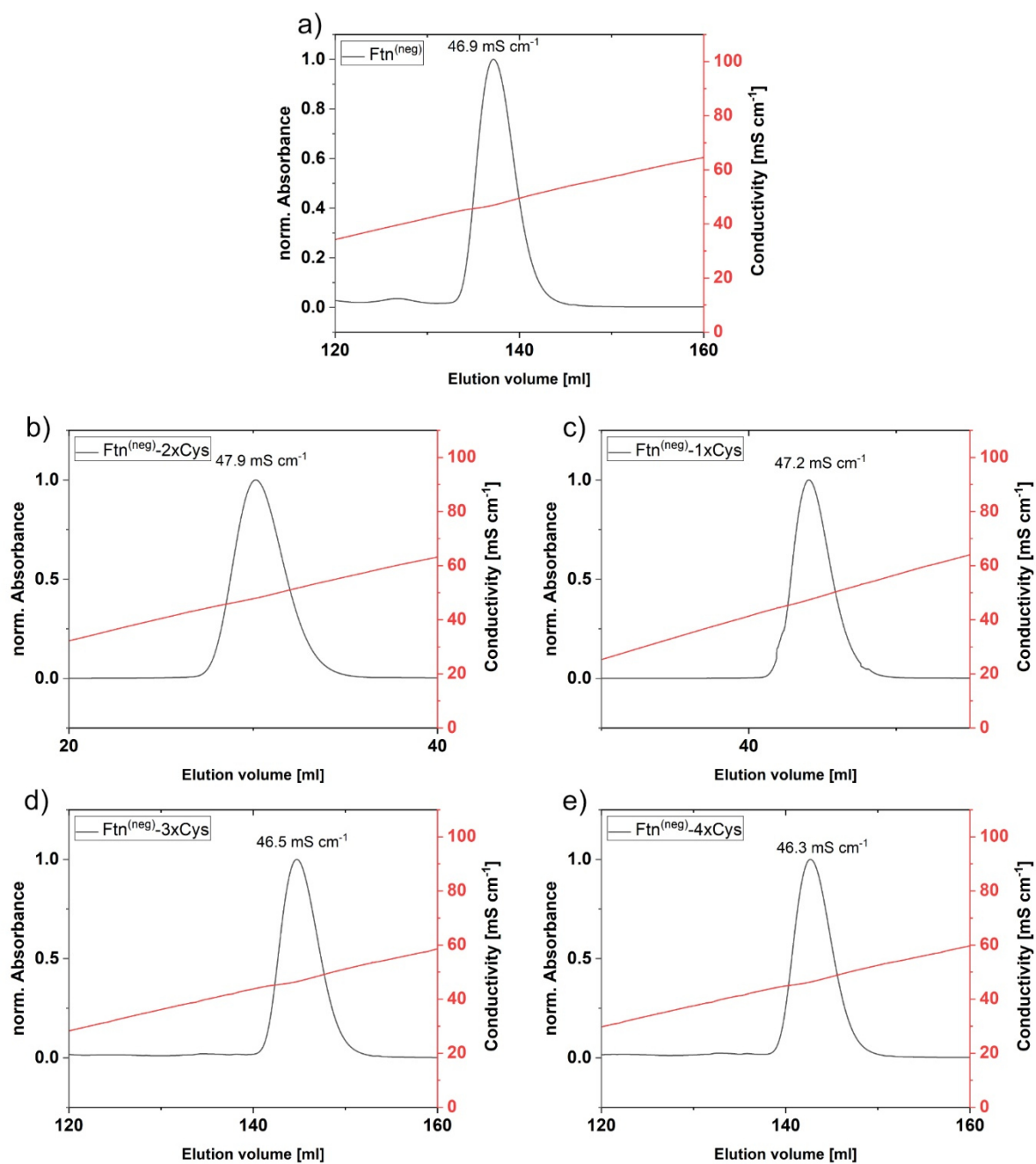


Figure 10.2: Ion-exchange chromatography (IEC) of Ftn^(neg)-Cysteine variants. The absorbance at 280 nm indicating the presence of protein is represented by the black line. The red line represents the conductivity of the buffer during the elution step and is a measure for the salt concentration. IEC chromatogram for a) Ftn^(neg), b) Ftn^(neg)-1xCys, c) Ftn^(neg)-2xCys, d) Ftn^(neg)-3xCys and e) Ftn^(neg)-4xCys.

Table 10.3: Conductivity of buffer at protein elution in IEC. Protein is eluted at a specific salt concentration measured by conductivity of the buffer. All Ftn^(neg)-Cys variants show roughly the same elution behavior indicating that mutations have no effect on the outer surface.

Protein	Conductivity [mS cm ⁻¹]
Ftn ^(neg) -1xCys	47.9
Ftn ^(neg) -2xCys	47.2
Ftn ^(neg) -3xCys	46.5
Ftn ^(neg) -4xCys	46.3

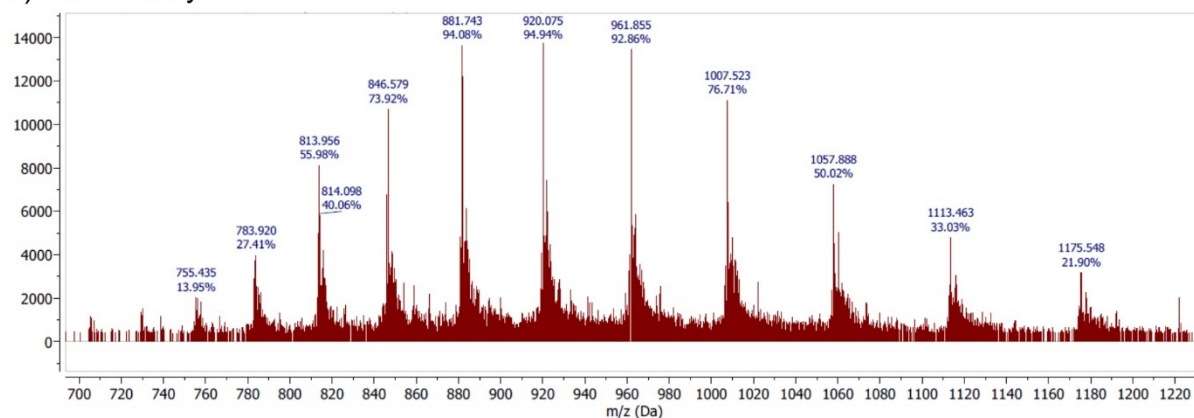
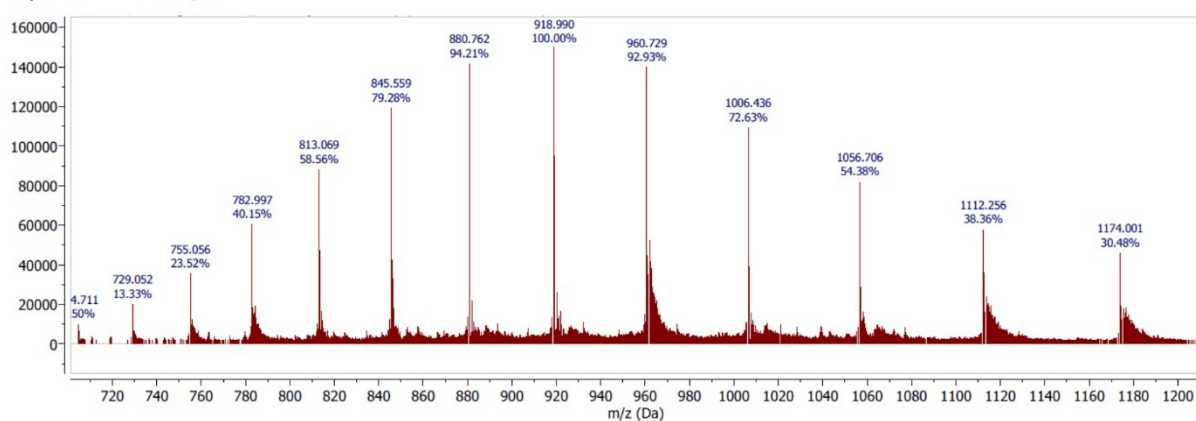
a) Ftn^(neg)-1xCysb) Ftn^(neg)-2xCys

Figure 10.3: ESI-MS spectra of Ftn^(neg)-1xCys and 2xCys.

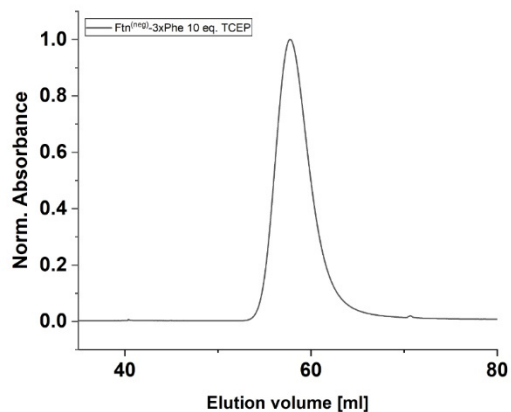
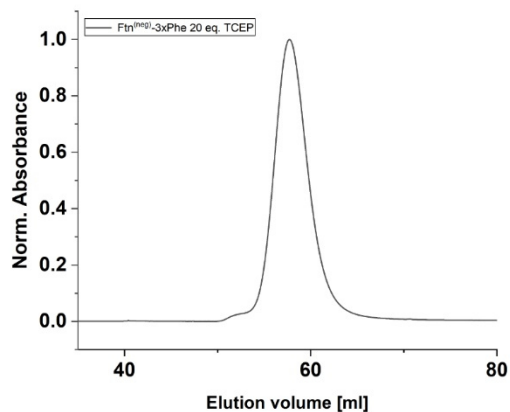
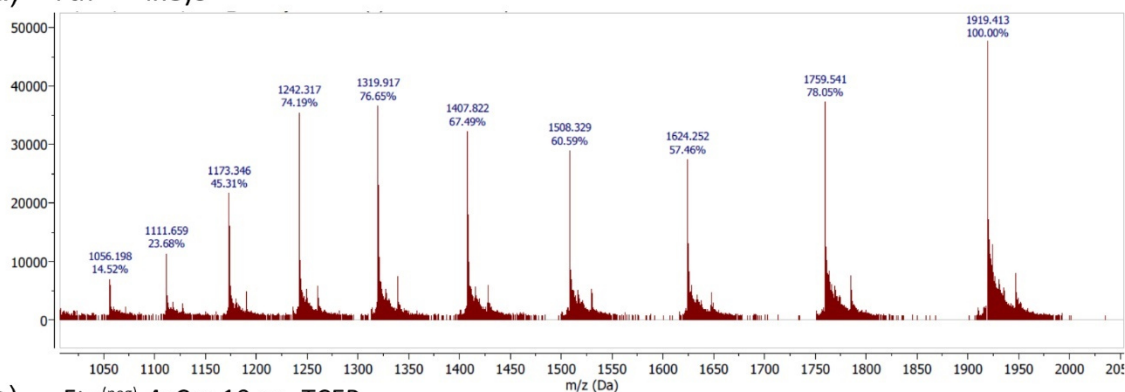
a) Ftn^(neg)-3xCys-Phe 10 eq. TCEPb) Ftn^(neg)-3xCys-Phe 20 eq. TCEP

Figure 10.4: Size exclusion chromatogram of Ftn^(neg)-3xCys after Phe coupling in presence of TCEP. SEC revealed elution at expected elution volumes and indicates complete cage assembly.

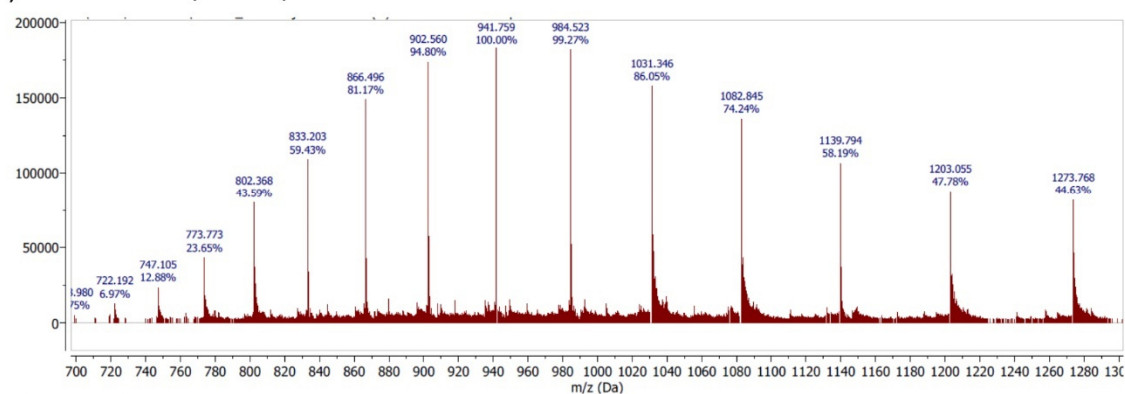
Table 10.4: Comparisons of UV absorption of Phe functionalized Ftn^(neg) cysteine variants. Ratio of UV absorption at 260 and 280 nm of Ftn^(neg)-3xCys and 4xCys after Phe functionalization with different amount of TCEP.

Protein	Ratio 260/280
Ftn ^(neg)	0.5
Ftn ^(neg) -3xCys 10 eq TCEP	1.35
Ftn ^(neg) -3xCys 20 eq TCEP	1.16
Ftn ^(neg) -4xCys 10 eq TCEP	1.43
Ftn ^(neg) -4xCys 20 eq TCEP	1.28

a) Ftn^(neg)-4xCys



b) Ftn^(neg)-4xCys 10 eq. TCEP



c) Ftn^(neg)-4xCys 10 eq. TCEP

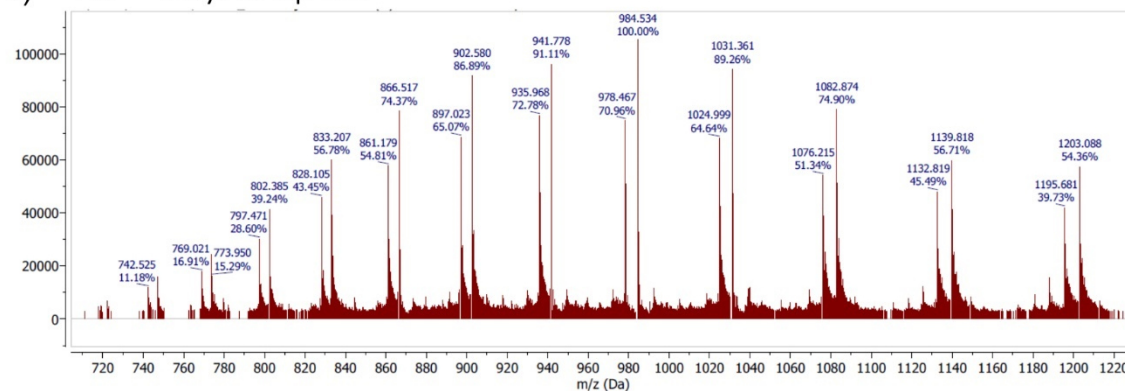


Figure 10.5: ESI-MS spectra of Ftn^(neg)-4xCys functionalized with Phe ligand. ESI-MS spectra for untreated Ftn^(neg)-4xCys (a) as a reference and coupled to Phe at the presence of 10 eq. (b) and 20 eq. (c) TCEP.

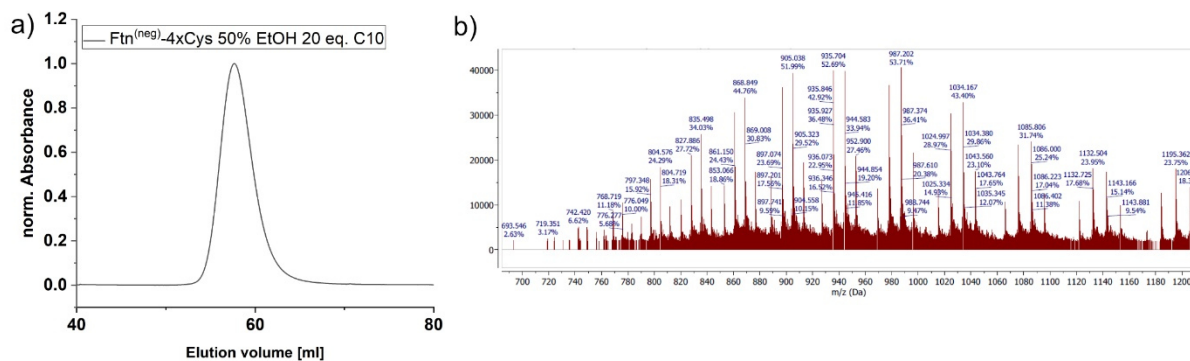


Figure 10.6: Ftn^(neg)-4xCys functionalization with 20 eq. C10 ligand at 50% EtOH in reaction mixture. a) SEC chromatogram after reaction reveals complete reassembly of protein cage. b) Respective ESI-MS spectrum showed mixture of subunits with different degree of functionalization

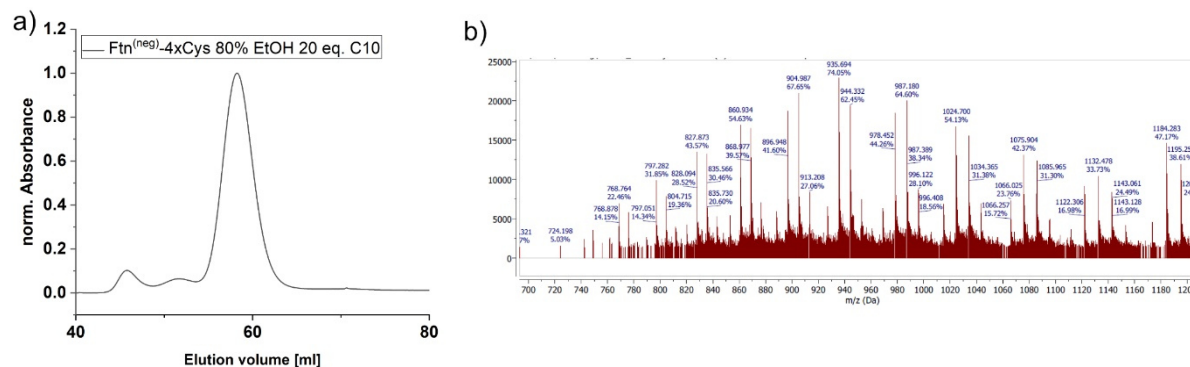


Figure 10.7: Ftn^(neg)-4xCys functionalization with 20 eq. C10 ligand at 80% EtOH in reaction mixture. a) SEC chromatogram after reaction reveals complete reassembly of protein cage. b) Respective ESI-MS spectrum showed mixture of subunits with different degree of functionalization

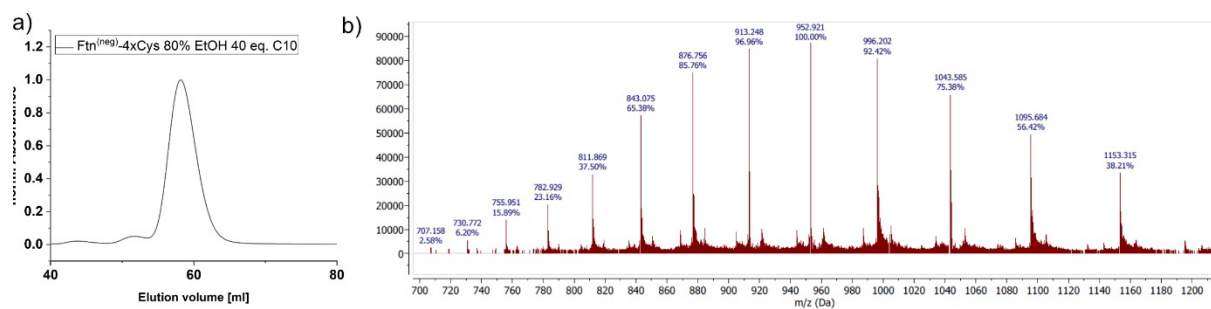
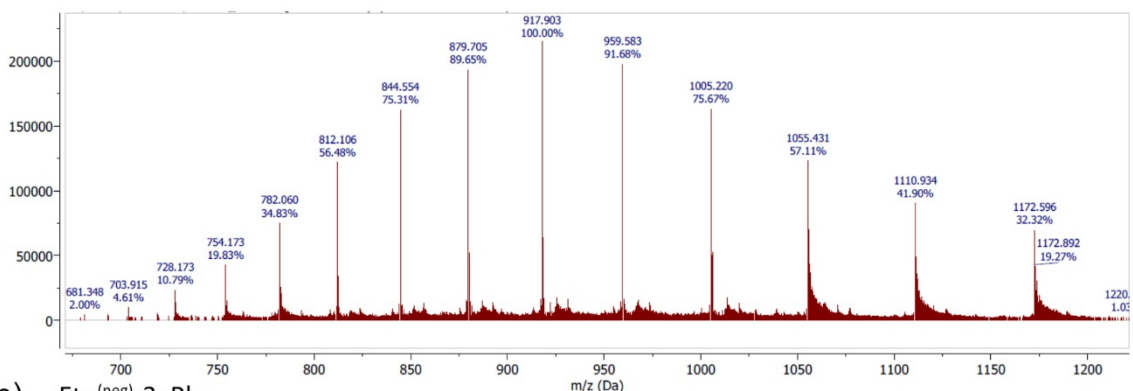
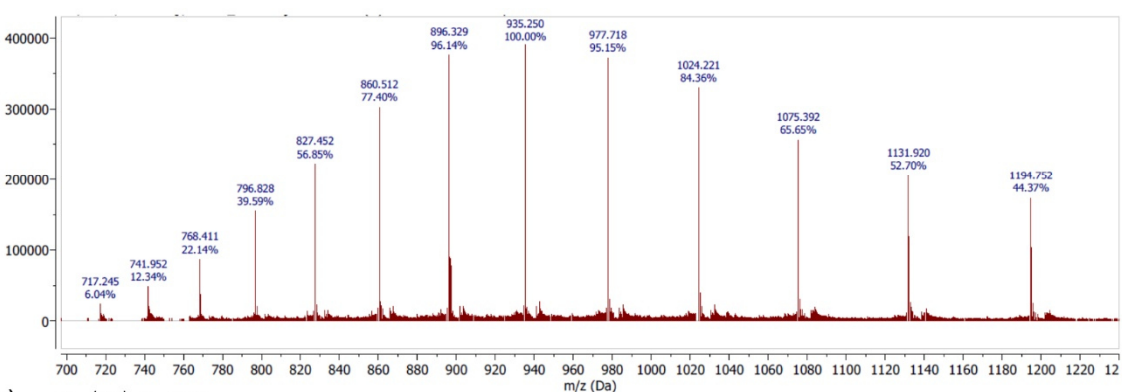


Figure 10.8: Ftn^(neg)-4xCys functionalization with 40 eq. C10 ligand at 80% EtOH in reaction mixture. a) SEC chromatogram after reaction reveals complete reassembly of protein cage. b) Respective ESI-MS spectrum showed mixture of subunits with different degree of functionalization

a) Ftn^(neg)-3xCys



b) Ftn^(neg)-3xPhe



c) Ftn^(neg)-3xC10

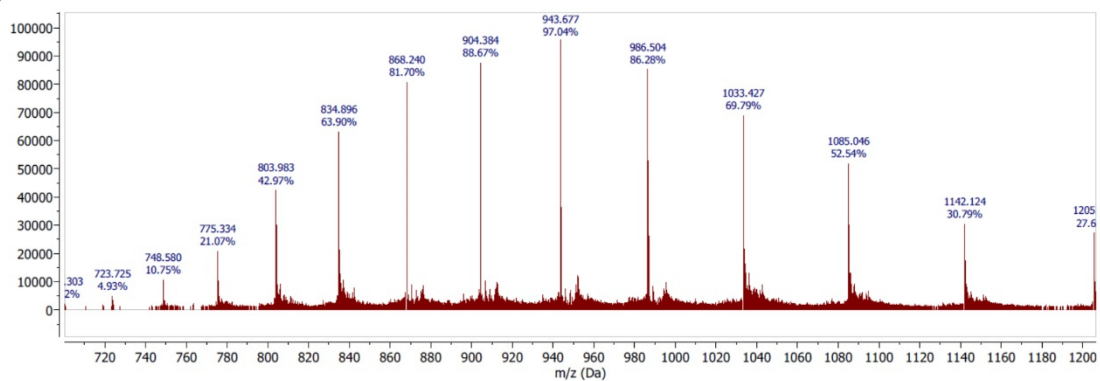
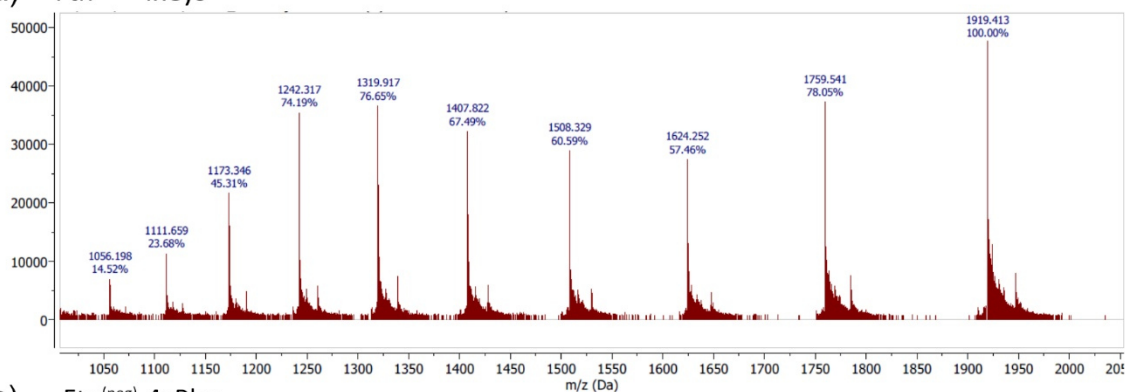
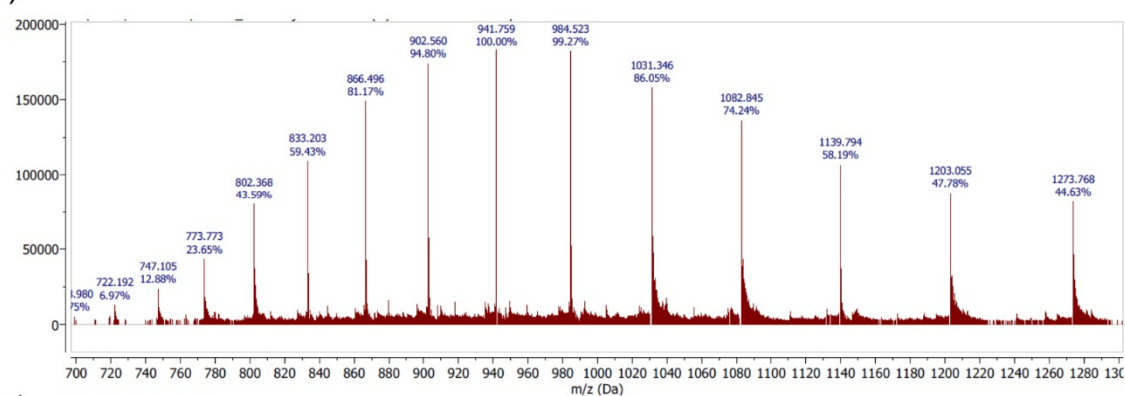


Figure 10.9: ESI-MS spectrum of Ftn^(neg)-3xCys variants. ESI-MS spectra for all spectra contain signal for only one species indicating pure samples.

a) Ftn^(neg)-4xCys



b) Ftn^(neg)-4xPhe



c) Ftn^(neg)-4xC10

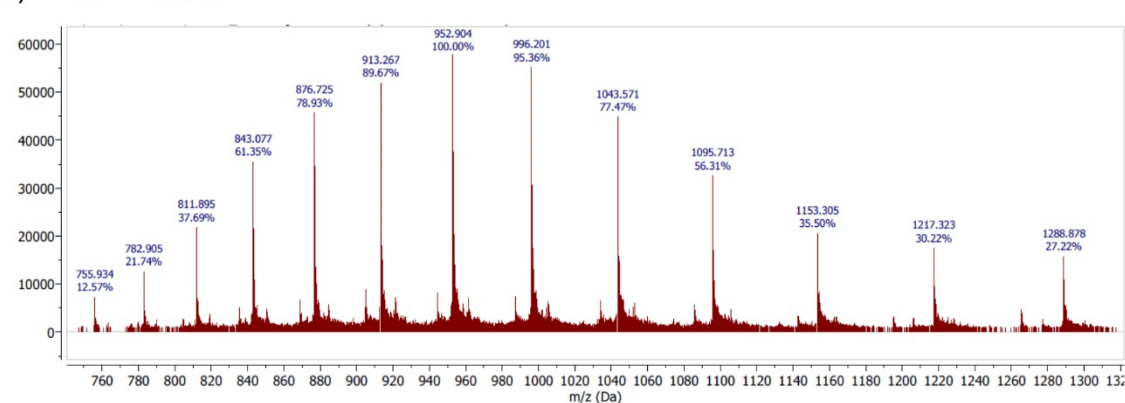


Figure 10.10 ESI-MS spectrum of Ftn^(neg)-4xCys variants. ESI-MS spectra for all spectra contain signal for only one species indicating pure samples.

Table 10.5: Summary of crystallographic data. Data statistics and refinement details for the crystal structures obtained from crystals of Ftn^(neg) functionalized with three C10 ligands

Ftn^(neg)-3xC10	
Data collection	
Wavelength (Å)	1.00
Space group	<i>P23</i>
Unit cell dimensions [a, (Å)]	181.44
Resolution range (Å)	44.01 – 2.0
Highest resolution shell (Å)	2.072 – 2.0
No. of observed reflections	2 767 050 (272 472)
No. of unique reflections ^[a]	133 407 (13 123)
Multiplicity	20.7 (20.8)
Completeness (%)	99.75 (98.91)
<I/σI>	46.05 (17.04)
R _{merge} (%)	5.56 (17.1)
R _{meas} (%)	5.70 (17.5)
Wilson B-factor	15.94
Refinement	
R _{work} (%)	14.4
R _{free} (%)	17.7
No. atoms	12512
macromolecules	11301
ions/glycerol	135
water	1076
B-factor (Å ²)	17.76
macromolecules	16.96
ions/glycerol	34.63
water	24.05

R.m.s deviations

bond lengths (Å) 0.016

bond angles (deg) 1.84

Ramachandran statistics (%)

favoured 98.38

outliers 0

Molprobability score 0.9

[a] Value in parentheses indicates number of reflections used for R_{free} calculation.

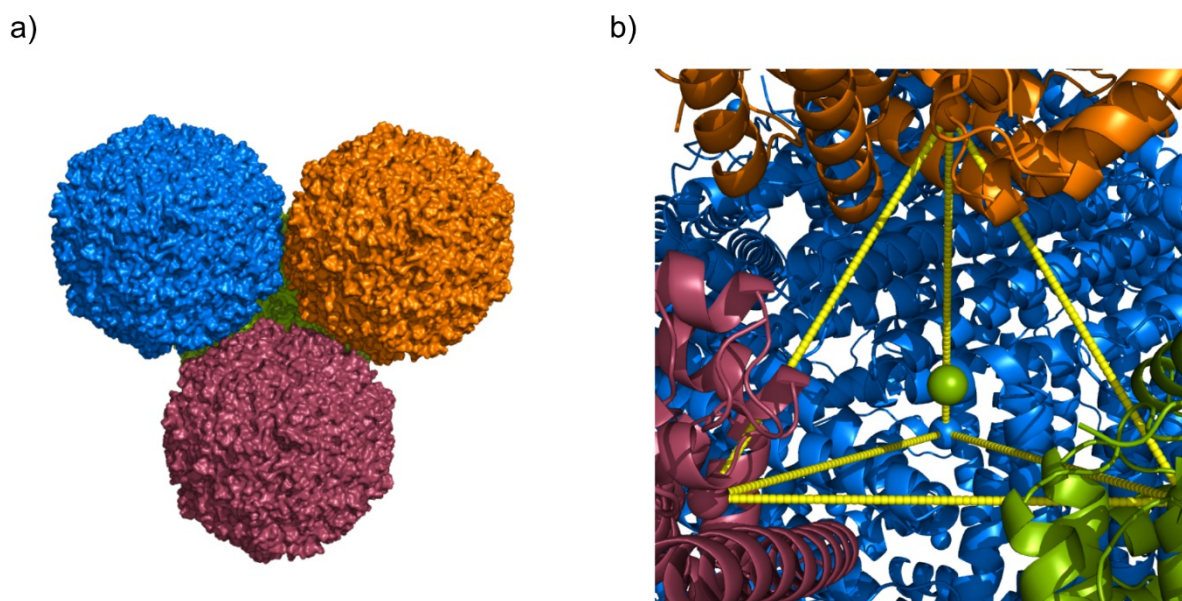


Figure 10.11: Size of solvent channel. a) Inspection of the unit cell of the $\text{Ftn}^{(\text{neg})}$ crystal structure (PDB ID: 5JJK) revealed the largest channel at the intersection of 4 protein container. b) To determine the size of this channel a tetrahedron was constructed between the Mg atoms located in the 3-fold channel. The center of the tetrahedron was calculated using Equation (7) with the position of the Mg atoms as vectors \vec{a} , \vec{b} , \vec{c} and \vec{d} . One atom is placed at the center point and the “Distancetoatom”¹ module from *PyMol* is used to find the distance to the nearest atom of the surrounding protein cages. The oxygen in the carboxy group of the amino acids aspartate at position 123 was the nearest with a distance between 15.483 – 15.494 Å. Therefore, it can be assumed that the channel between the container has a diameter of around 31 Å.

$$\vec{s} = \frac{1}{4}(\vec{a} + \vec{b} + \vec{c} + \vec{d}) \quad (7)$$

¹ <https://pymolwiki.org/index.php/Distancetoatom>; Authors: Andreas Warnecke and Jared Sampson

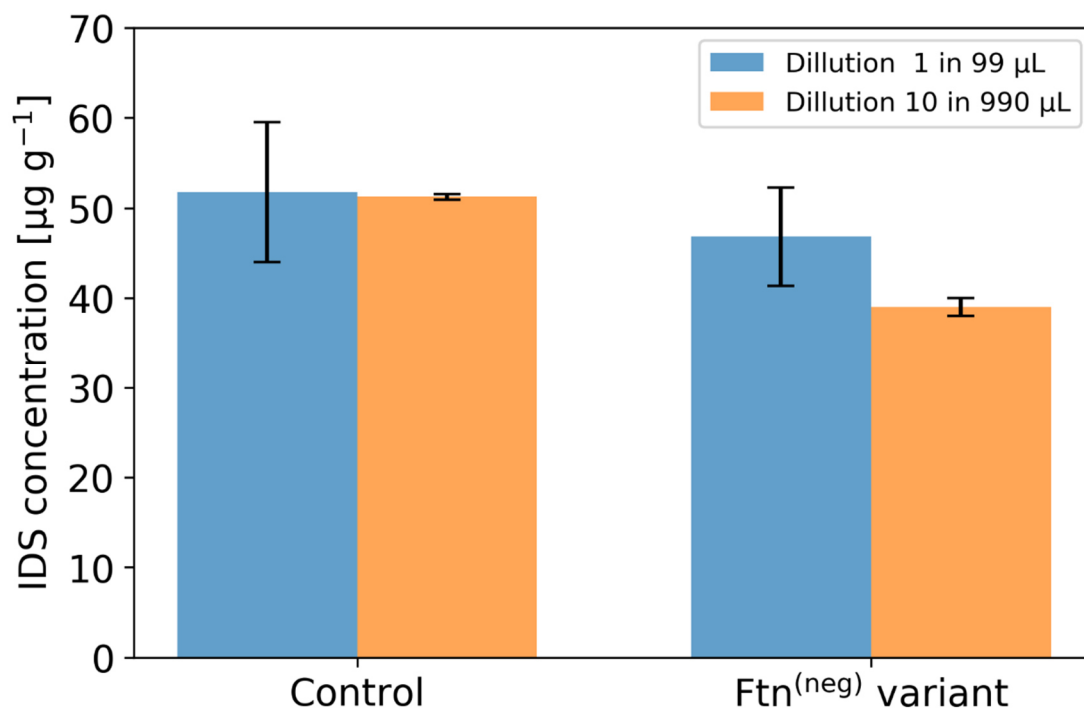


Figure 10.12: Standard deviation for PBUT measurement. Initial test with a 1:1000 dilution bears a high standard deviation between identical samples. This could be drastically decreased by separating the step

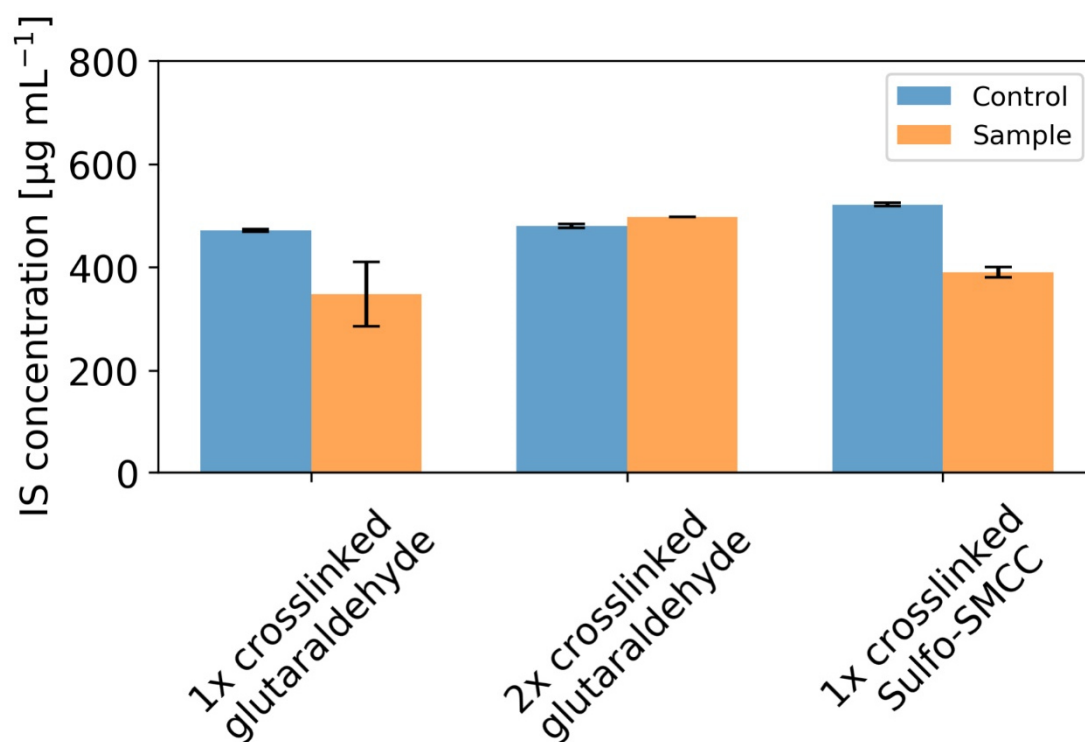


Figure 10.13: Reduction of toxin concentration for different crosslinking techniques. Difference in toxin concentration of a control sample to the sample treated with the adsorbent is shown. After a single treatment with glutaraldehyde toxin adsorption is still possible but crystals are unstable in higher concentration of BSA (Figure 5.14). After a second treatment respective crystals tolerate the BSA condition but no PBUT adsorption can be detected. Sulfo-SMCC crosslinking yield stable crystals still capable reducing the toxin concentration.

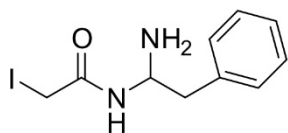
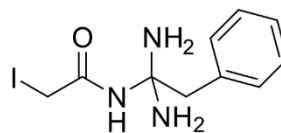
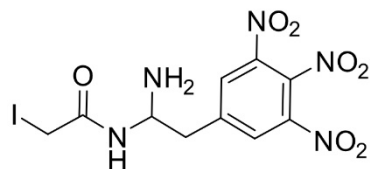
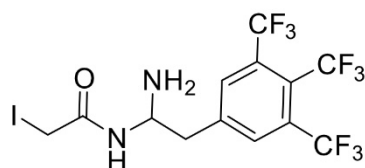
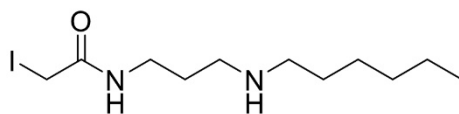
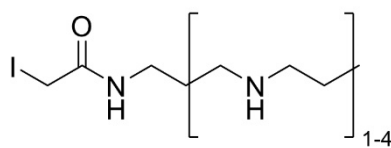
*N*-(1-amino-2-phenylethyl)-2-iodoacetamide*N*-(1,1-diamino-2-phenylethyl)-2-iodoacetamide*N*-(1-amino-2-(3,4,5-trinitrophenyl)ethyl)-2-iodoacetamide*N*-(1-amino-2-(3,4,5-tris(trifluoromethyl)phenyl)ethyl)-2-iodoacetamide*N*-(3-(hexylamino)propyl)-2-iodoacetamide2-iodo-*N*-(3-(propylamino)propyl)acetamide derivatives

Figure 10.14: Possible ligands for functionalization of Ftn^(neg)-Cys. The depicted ligands combine hydrophobic and polar positive characteristics for more efficient binding of the PBUTs. Aromatic ligands are further substituted with electron withdrawing groups to facilitate efficient π - π interactions with the electron rich PBUTs.

10.3 Additional information on protein redesign part

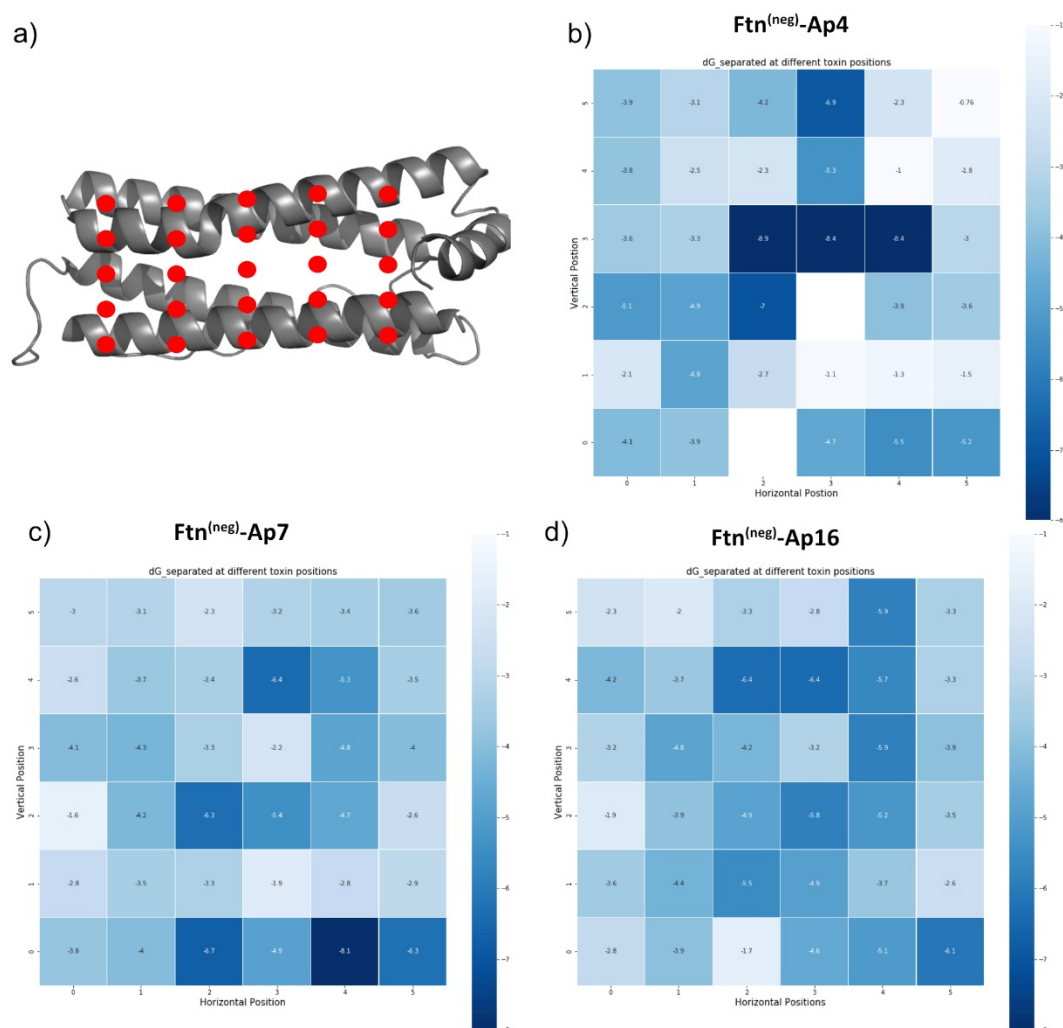


Figure 10.15: Binding energy as determined by Rosetta for Ftn^(neg)-Ap variants. Ligand docking protocols were performed with different start locations as indicated by red dots in a). The protocols performed a low- and high-resolution docking steps followed by a final minimization of side-chain energy. The binding energy is final determined by the by comparing the energy of the ligand-protein complex with the unbound protein. Results are shown as heatmaps for a) Ftn^(neg)-Ap4 b), c) Ap7 and d) Ap16.

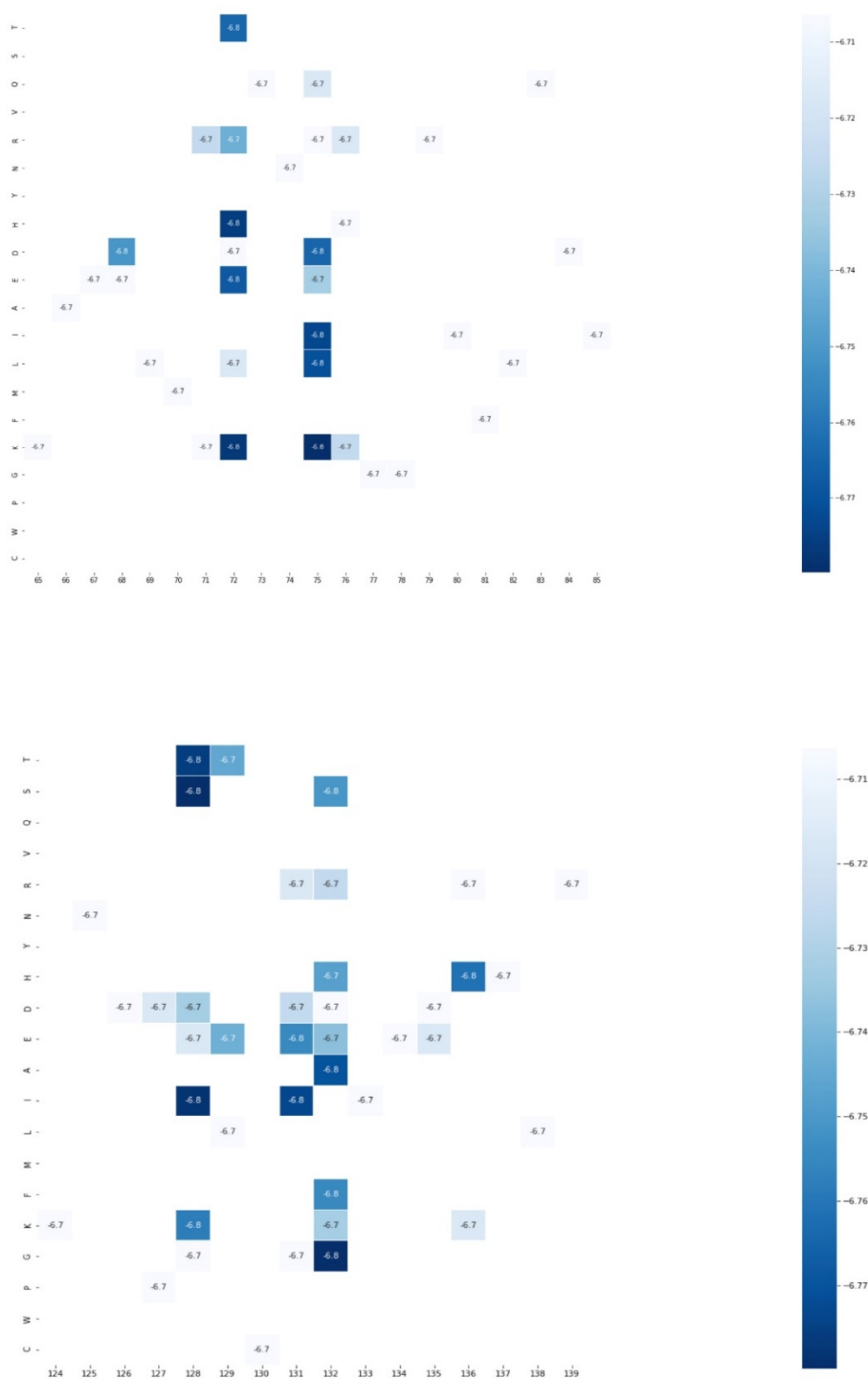


Figure 10.16: Influence of single amino acids on the overall $dG_{\text{separated}}$. First results from a script aiming to access the influence of a single amino acids in the modified sequences on the $dG_{\text{separated}}$ value of the model. The mean $dG_{\text{separated}}$ for all models were around -6.7. The script successfully identified positions and amino acids leading to an increase in $dG_{\text{separated}}$. The used test data contained the amino acid sequence of 30 models from a first-generation ligand-docking protocol and their respective $dG_{\text{separated}}$.

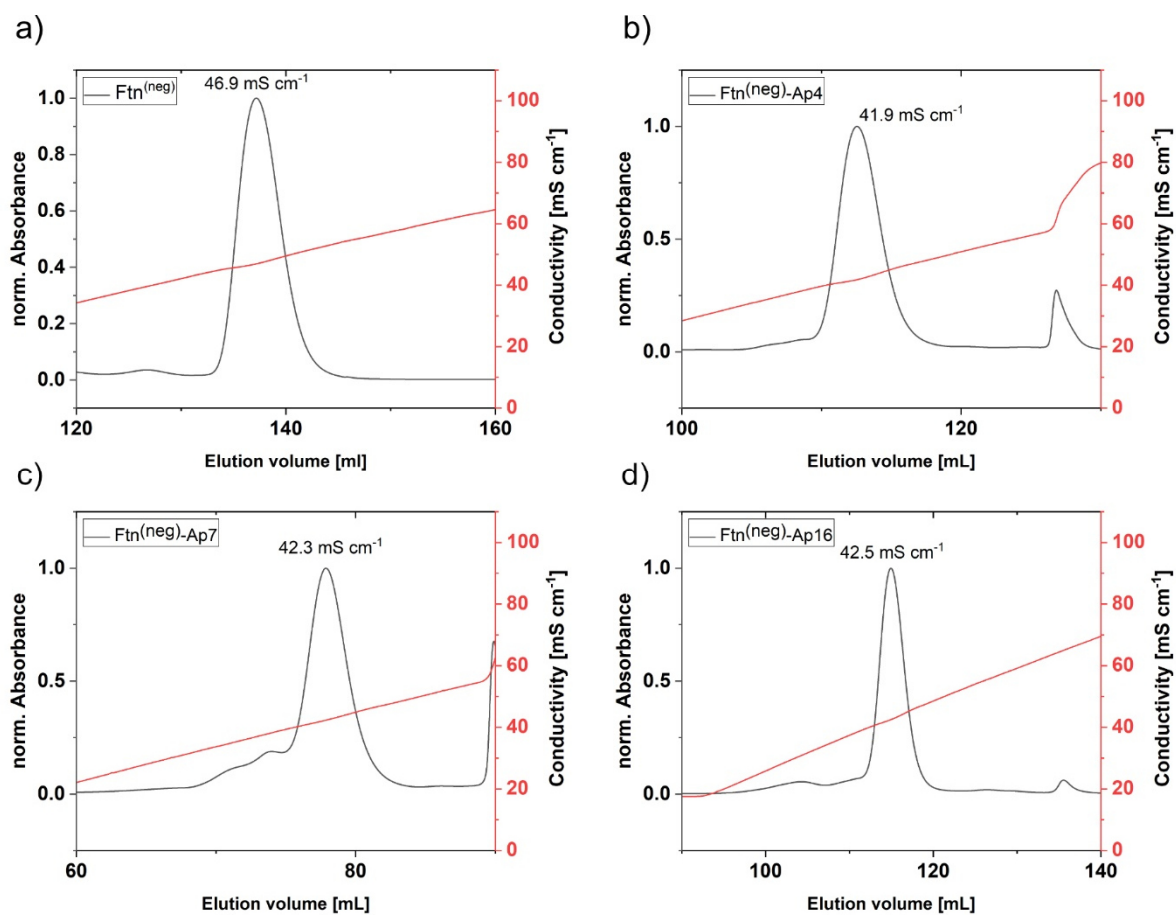


Figure 10.17: IEC chromatogram of ferritin variants with increased hydrophobic side chain density. IEC is performed for a) Ftn^(neg), b) Ftn^(neg)-Ap4, c) Ftn^(neg)-Ap7 and d) Ftn^(neg)-Ap16. Normalized absorbance at 280 nm (left y-axis in black) and conductivity (right y-axis in red) are plotted against the elution volume. The conductivity present at elution peak is indicated in each figure.

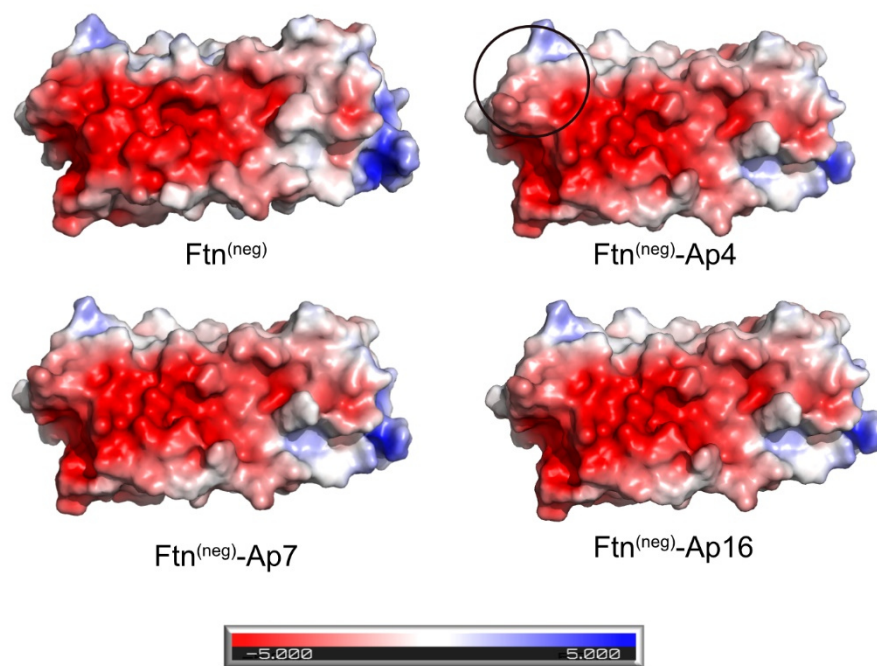


Figure 10.18: Surface potential of ferritin variants with increased density of hydrophobic residues. Surface potential was determined by the APBS tool.^[256] Only a small difference in potential between the unmodified and the modified variants can be detected in the upper left part of each variant. The region is indicated for the $Ftn^{(neg)}-Ap4$ with a black circle.

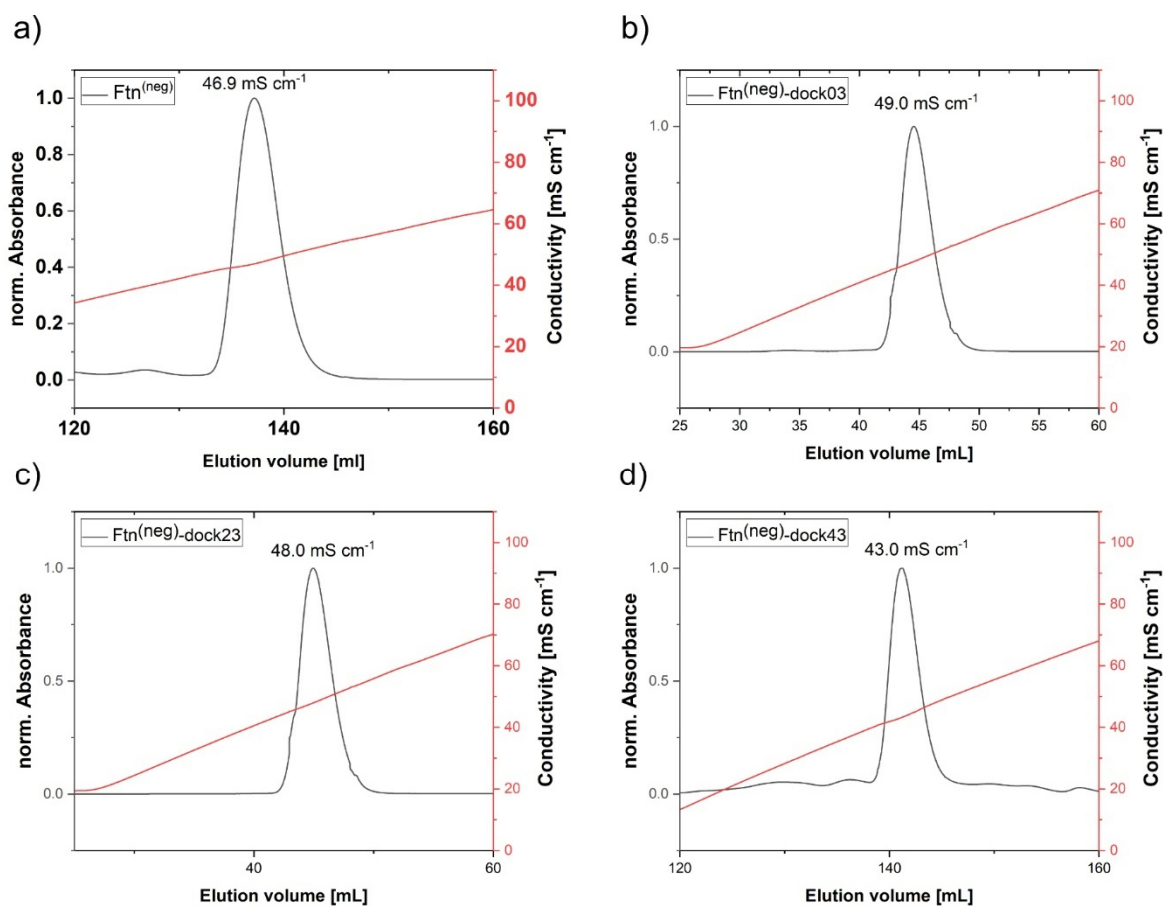


Figure 10.19: IEC chromatogram for Ftn^(neg)-docking variants. Ion-exchange chromatography were performed for each variant and the normalized absorbance at 280 nm (black) and the conductivity (red) are plotted against the elution volume for a) unmodified variants and the redesigned variant b) dock03 c) dock23 and d) dock43.

Table 10.6: Removed and added negative charged amino acids in redesigned Ftn^(neg) variants. The number of removed and added negative charged amino acids Glutamic acid (E) and Aspartic acid (D) are determined from the sequence of the redesigned ferritin variants.

Protein	Removed	Added	Net change
	D and E	D and E	
Ftn ^(neg) -Ap4	4	0	-4
Ftn ^(neg) -dock03	2	4	+2
Ftn ^(neg) -dock23	3	4	+1
Ftn ^(neg) -dock43	2	3	+1

Calculation of adsorption capacity if every subunit bind one IS molecule

$$m_{IS} = \frac{n_{Ferritin}}{M_{Ferritin}} * 24 * M_{Indoxyl\ sulfate} \quad (8)$$

$$M_{Indoxyl\ sulfate} = 213.21 \text{ g mol}^{-1}$$

$$M_{Ferritin} = 508560 \text{ g mol}^{-1}$$

For a mass of 1 g ferritin, an indoxyl sulfate mass of 10.06 mg is calculated. Resulting in an adsorption capacity of 10.06 mg g⁻¹.

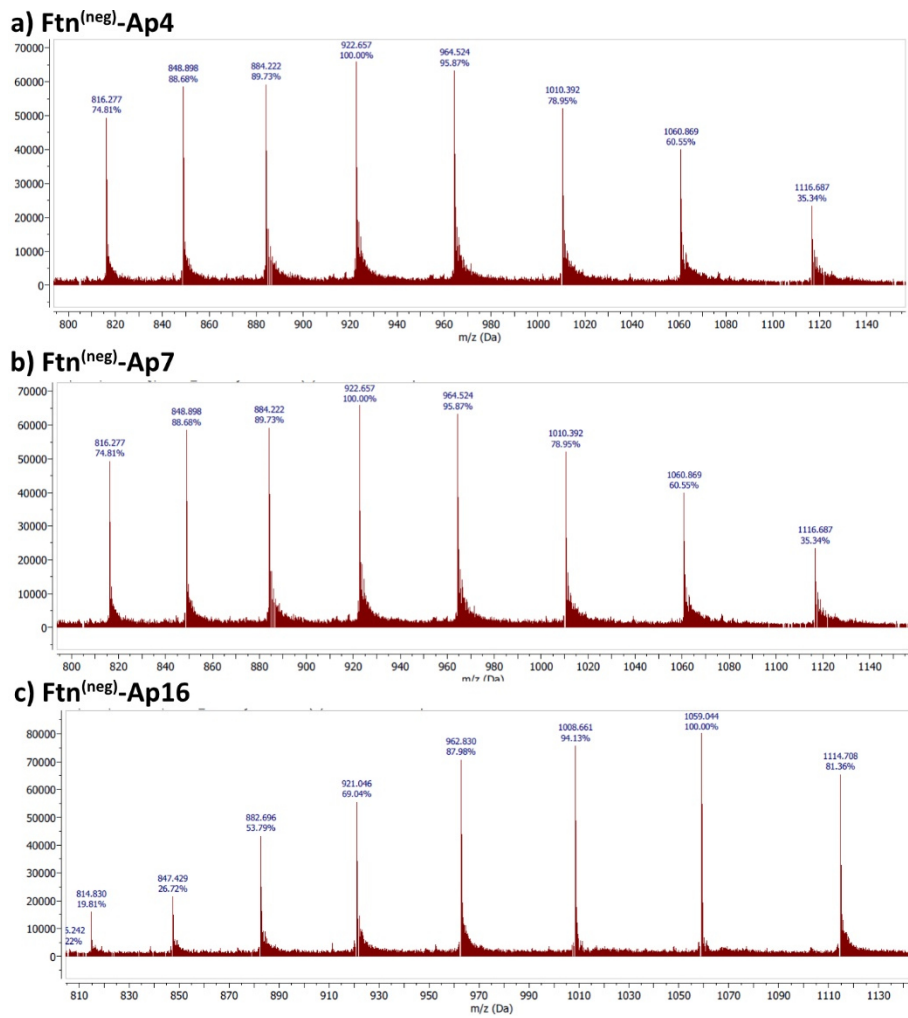


Figure 10.20: ESI-MS spectra for Ftn^(neg)-Ap variants. ESI-MS spectra showed separated signals for differently charged ions. The calculated mass in summarized in Table 5.5.

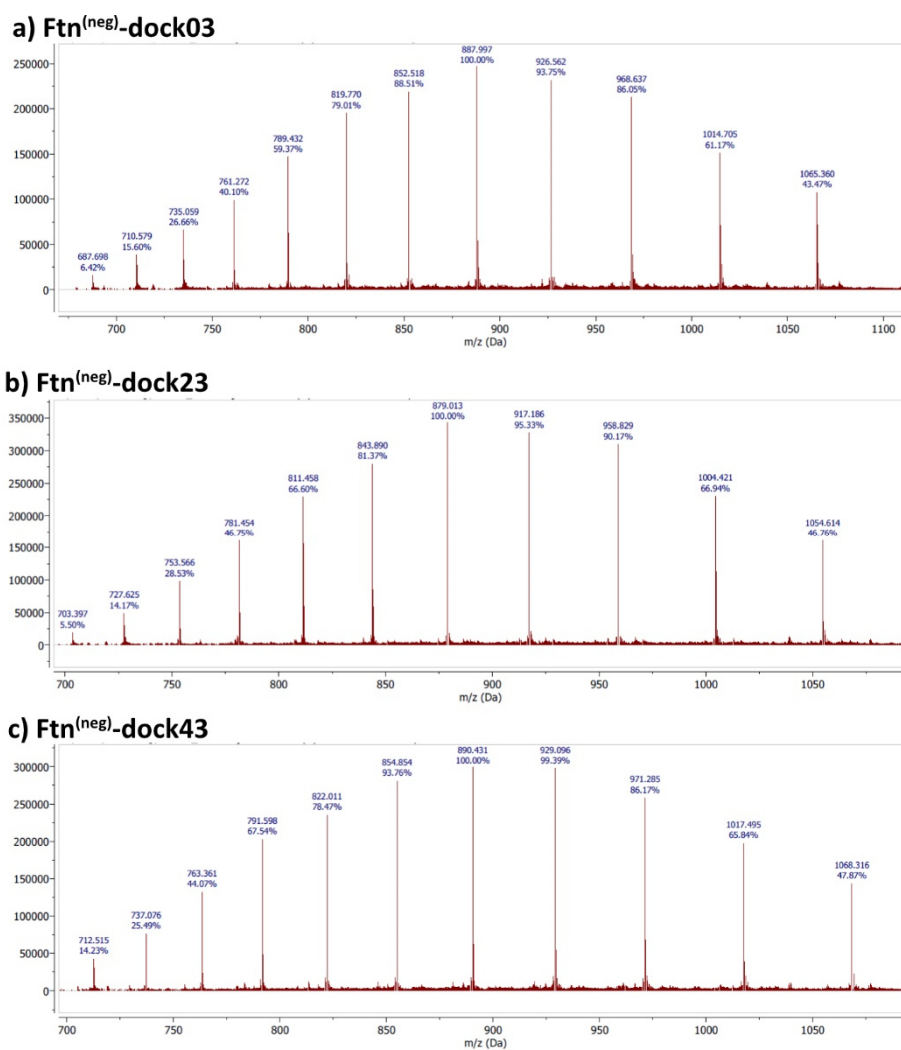


Figure 10.21: ESI-MS spectra for $\text{Ftn}^{(\text{neg})}$ -docking variants. ESI-MS spectra showed separated signals for differently charged ions.

Table 10.7: Overview variants for protein-protein docking. For each of the seven variants used for protein-protein docking modeling the amino acids at a given position are specified. Mutations are grouped in positions located at the interface and not located at the interface.

Variant	Interface mutation	Non interface mutation
Ftn ^(neg)	x	x
Ftn ^(neg) -m1	A18	x
Ftn ^(neg) -m3	x	C90, C102, H105
Ftn ^(pos)	x	x
Ftn ^(pos) -m1	A18	x
Ftn ^(pos) -m4	x	N98, C102, H105, N109
Ftn	A18	C90, N98 C102, H105, N109

Table 10.8: Summary of crystallographic data. Data statistics and refinement details for the crystal structures obtained from crystals of Ftn^(neg) docking variants

	Ftn^(neg)-Ap4	Ftn^(neg)-Ap7	Ftn^(neg)-Ap16
Data collection			
Wavelength (Å)	1.0	1.0	1.0
Space group	F432	P23	P23
Unit cell dimensions [a, (Å)]	177.52	181.17	180.97
Resolution range (Å)	44.38 – 1.5	43.94 – 2.001	45.24 – 1.5
Highest resolution shell (Å)	1.554 – 1.5	2.072 – 2.001	1.554 – 1.5
No. of observed reflections	2983933 (294425)	5197677 (541421)	12702453 (1246499)
No. of unique reflections ^[a]	38852 (3825)	68678 (6688)	312152 (31086)
Multiplicity	76.8 (76.9)	75.7 (80.3)	40.7 (40.1)
Completeness (%)	99.89 (99.16)	99.18 (99.23)	99.88 (99.93)
<I/σI>	28.20 (0.76)	31.52 (4.90)	21.90 (2.69)

Appendix

R _{merge} (%)	0.144 (6.727)	0.1778 (1.704)	0.1576 (1.571)
R _{meas} (%)	0.145 (6.771)	0.1791 (1.715)	0.1596 (1.591)
Wilson B-factor	25.59	27.73	14.21
Refinement			
R _{work} (%)	0.20	0.23	0.14
R _{free} (%)	0.23	0.27	0.16
No. atoms	1541	11858	12841
Macromolecules	1403	11408	11384
ions/glycerol	16	15	15
Water	122	435	1442
B-factor (Å ²)			
Macromolecules	30.66	31.08	15.08
ions/glycerol	55.59	55.10	27.45
water	40.11	33.21	26.50
R.m.s deviations			
bond lengths (Å)	0.015	0.014	0.015
bond angles (deg)	1.86	1.86	1.81
Ramachandran statistics (%)			
favoured	98.24	98.31	98.31
outliers	0	0	0
Molprobit score	0.72	1.97	1.02

[a] Value in parentheses indicates number of reflections used for R_{free} calculation.

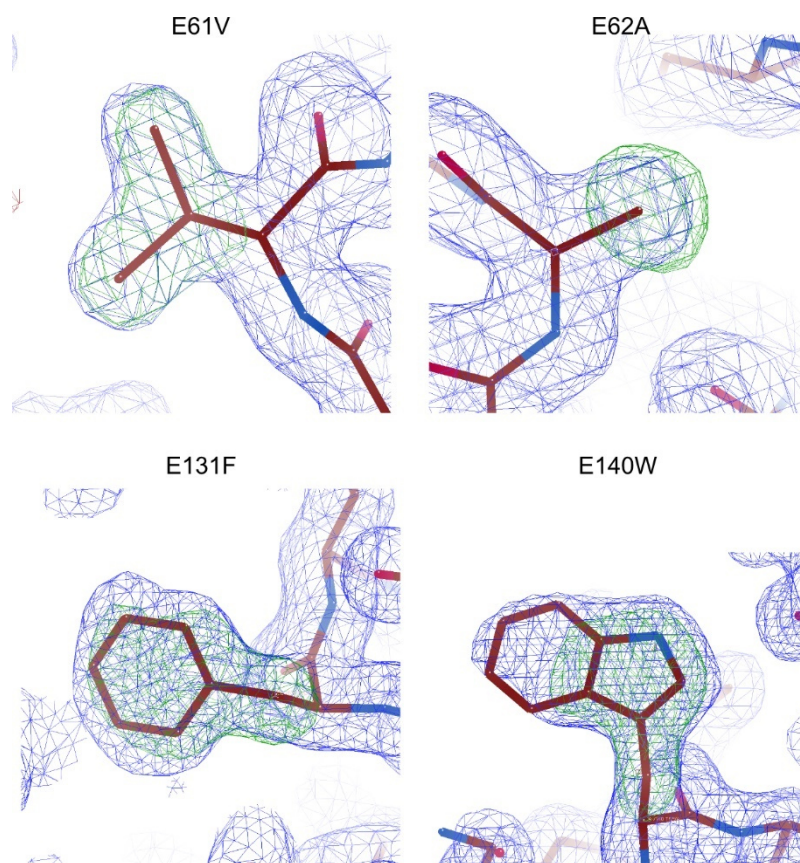


Figure 10.22: Electron density omit maps for $Ftn^{(neg)}$ -Ap4. Electron density ($2F_o - F_c$ omit map, blue) and difference electron density ($F_o - F_c$, green) map. $2F_o - F_c$ map (blue): 1 RMSD, $F_o - F_c$ (green): 5 RMSD. Maps were calculated without atoms of the introduced amino acids.

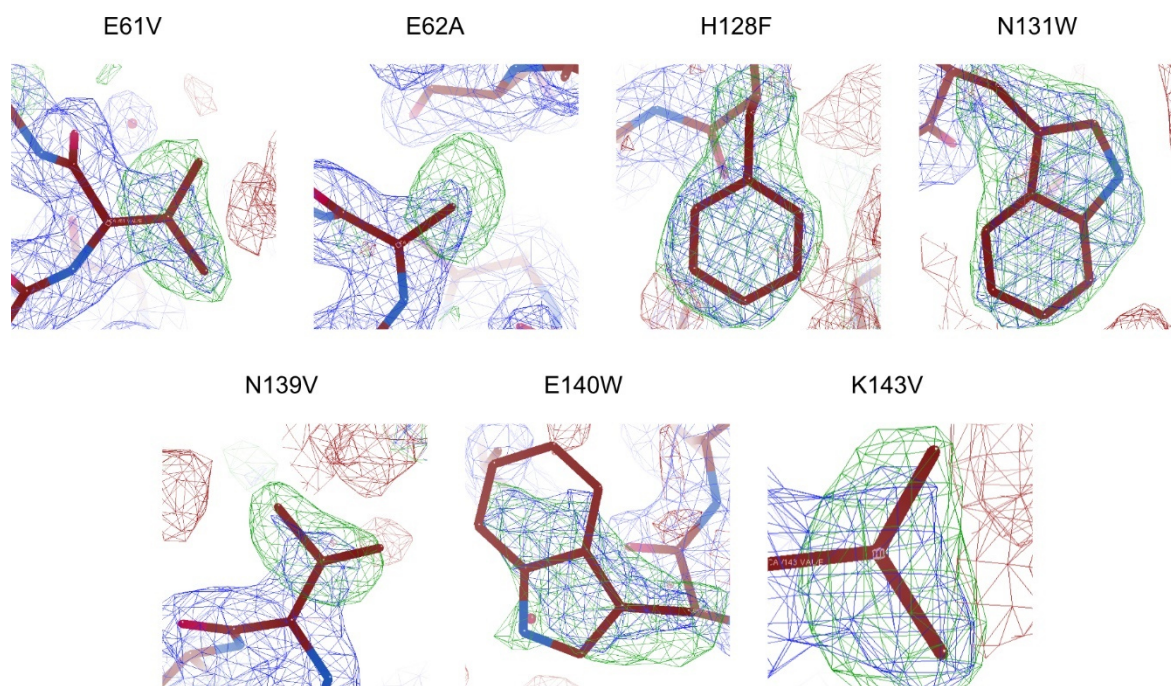


Figure 10.23: Electron density omit maps for Ftn^(neg)-Ap7. Electron density ($2F_o-F_c$ omit map, blue) and difference electron density (F_o-F_c , green) map. $2F_o-F_c$ map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated without atoms of the introduced amino acids.

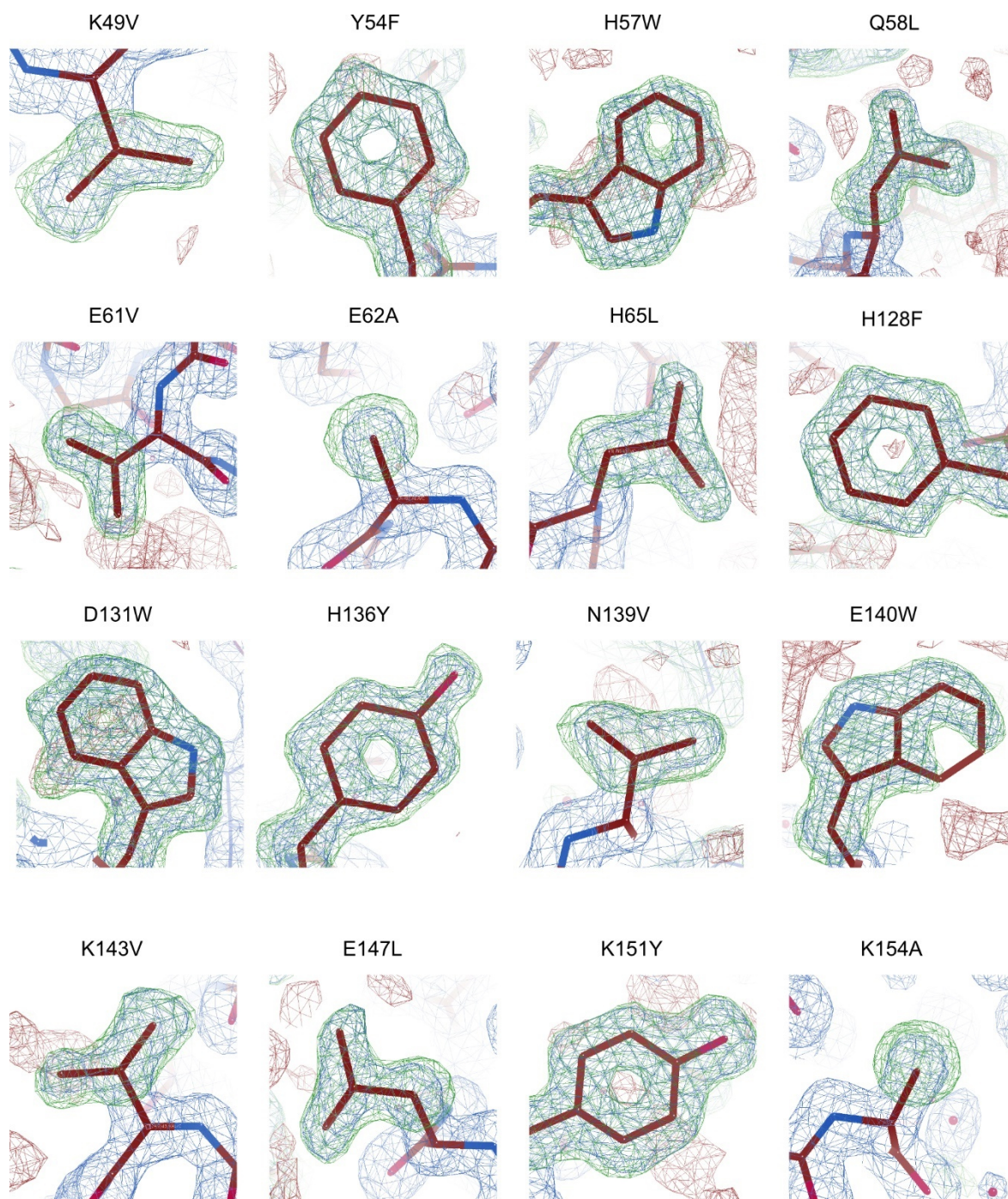


Figure 10.24: Electron density omit maps for Ftn^(neg)-Ap16. Electron density ($2F_o-F_c$ omit map, blue) and difference electron density (F_o-F_c , green) map. $2F_o-F_c$ map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated without atoms of the introduced amino acids.

Table 10.9: Summary of crystallographic data. Data statistics and refinement details for the crystal structures obtained from crystals of Ftn^(neg) docking variants

	Ftn ^(neg) -dock03	Ftn ^(neg) -dock23	Ftn ^(neg) -dock43
Data collection			
Wavelength (Å)	1.0	1.0	1.0
Space group	F432	P23	P23
Unit cell dimensions [a, (Å)]	179.25	181.21	182.2
Resolution range (Å)	44.81 – 1.802	43.95 – 2.0	44.19 – 2.0
Highest resolution shell (Å)	1.866 – 1.802	2.072 – 2.0	2.072 – 2.0
No. of observed reflections	1833717 (180933)	2767527 (278294)	5633193 (572517)
No. of unique reflections ^[a]	23409 (2294)	133050 (13145)	135273 (13428)
Multiplicity	78.3 (78.8)	20.8 (21.2)	41.6 (42.6)
Completeness (%)	99.81 (98.65)	99.82 (99.98)	99.82 (99.97)
<I/σI>	21.11 (1.09)	22.06 (6.09)	24.05 (6.11)
R _{merge} (%)	0.1826 (2.341)	0.1145 (0.61)	0.1774 (0.8651)
R _{meas} (%)	0.1838 (2.356)	0.1173 (0.625)	0.1796 (0.8754)
Wilson B-factor	33.11	21.21	18.59
Refinement			
R _{work} (%)	0.21	0.17	0.17
R _{free} (%)	0.25	0.20	0.20
No. atoms			
macromolecules	1457	11304	11464
ions/glycerol	4	10	15
water	122	615	627
B-factor (Å ²)			
macromolecules	33.25	23.52	20.20
ions/glycerol	37.82	56.71	89.11
water	38.58	24.97	21.70

Appendix

R.m.s deviations

bond lengths (Å)	0.016	0.016	0.016
------------------	-------	-------	-------

bond angles (deg)	1.83	1.83	1.85
-------------------	------	------	------

Ramachandran statistics (%)

favoured	98.24	98.31	98.53
----------	-------	-------	-------

outliers	0	0	0
----------	---	---	---

Molprobit score	1.39	1.31	1.38
-----------------	------	------	------

[a] Value in parentheses indicates number of reflections used for R_{free} calculation.

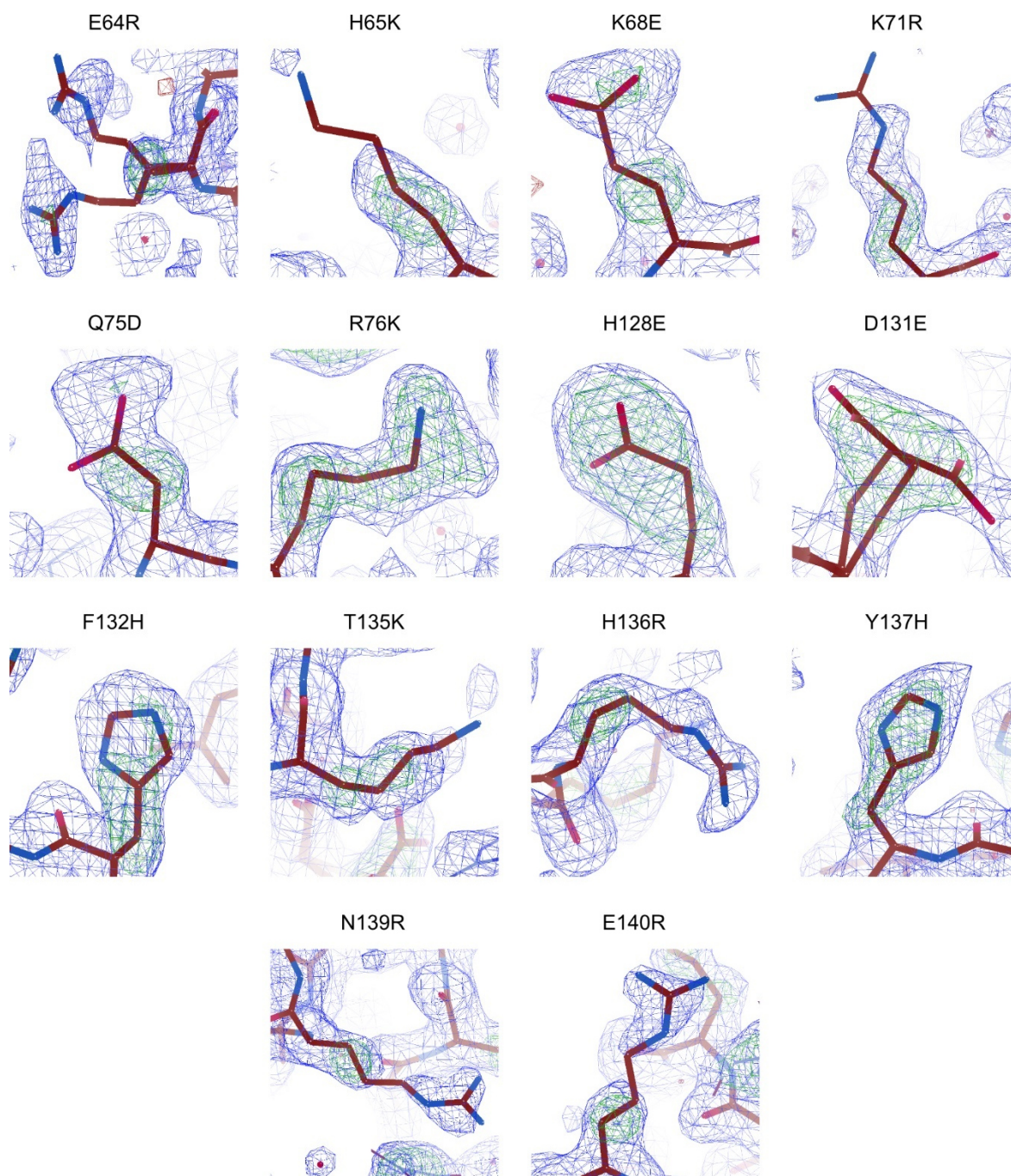


Figure 10.25: Electron density omit maps for Ftn^(neg)-dock03. Electron density ($2F_o-F_c$ omit map, blue) and difference electron density (F_o-F_c , green) map. $2F_o-F_c$ map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated without atoms of the introduced amino acids.

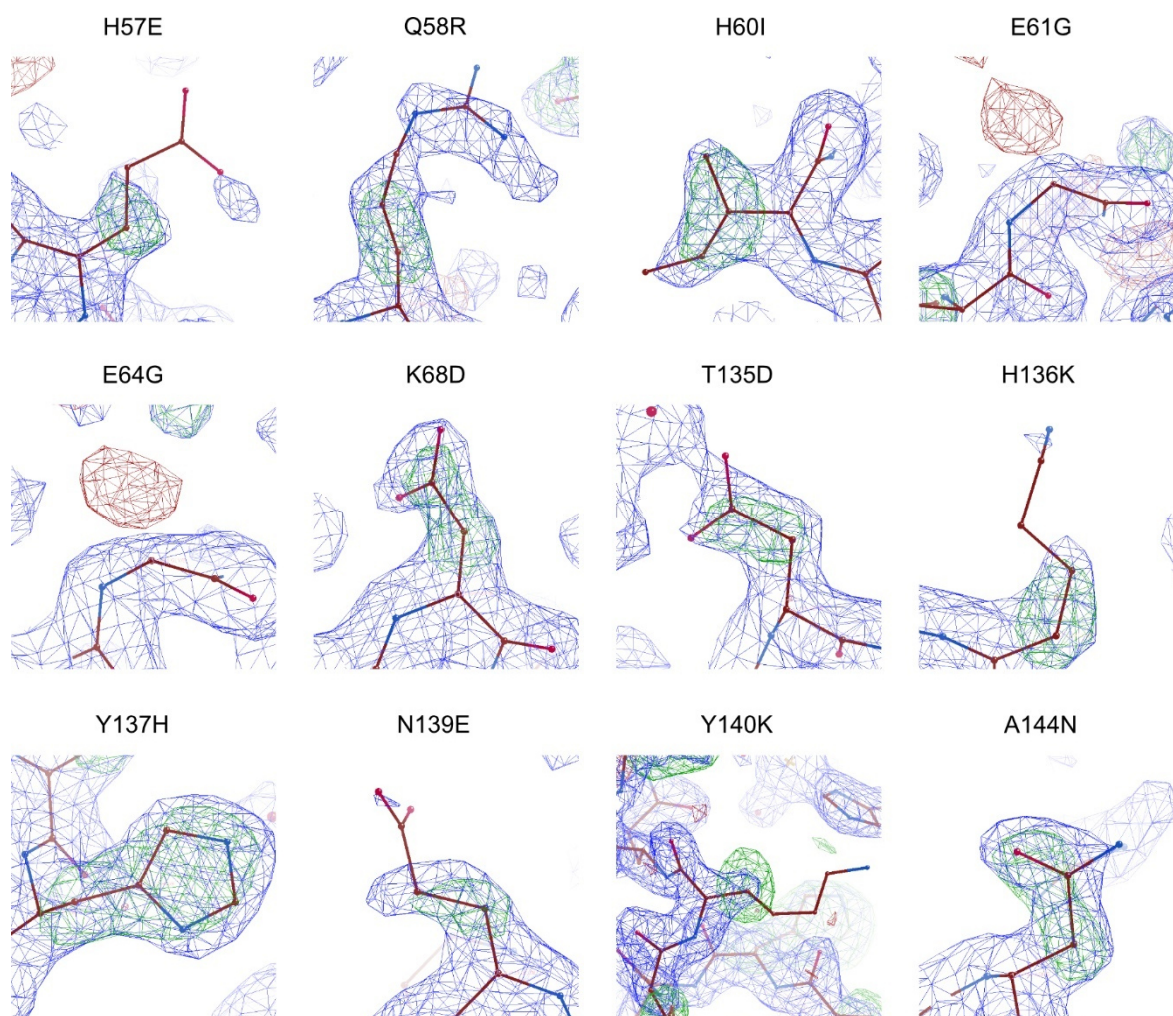


Figure 10.26: Electron density omit maps for Ftn^(neg)-dock23. Electron density (2F_o-F_c omit map, blue) and difference electron density (F_o-F_c, green) map. 2F_o-F_c map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated with a model without atoms of the introduced amino acids. For mutation E61G and E64G an alanine was introduced in the model for calculation of the omit maps. The red density at this point indicates that no side chain is present at this site.

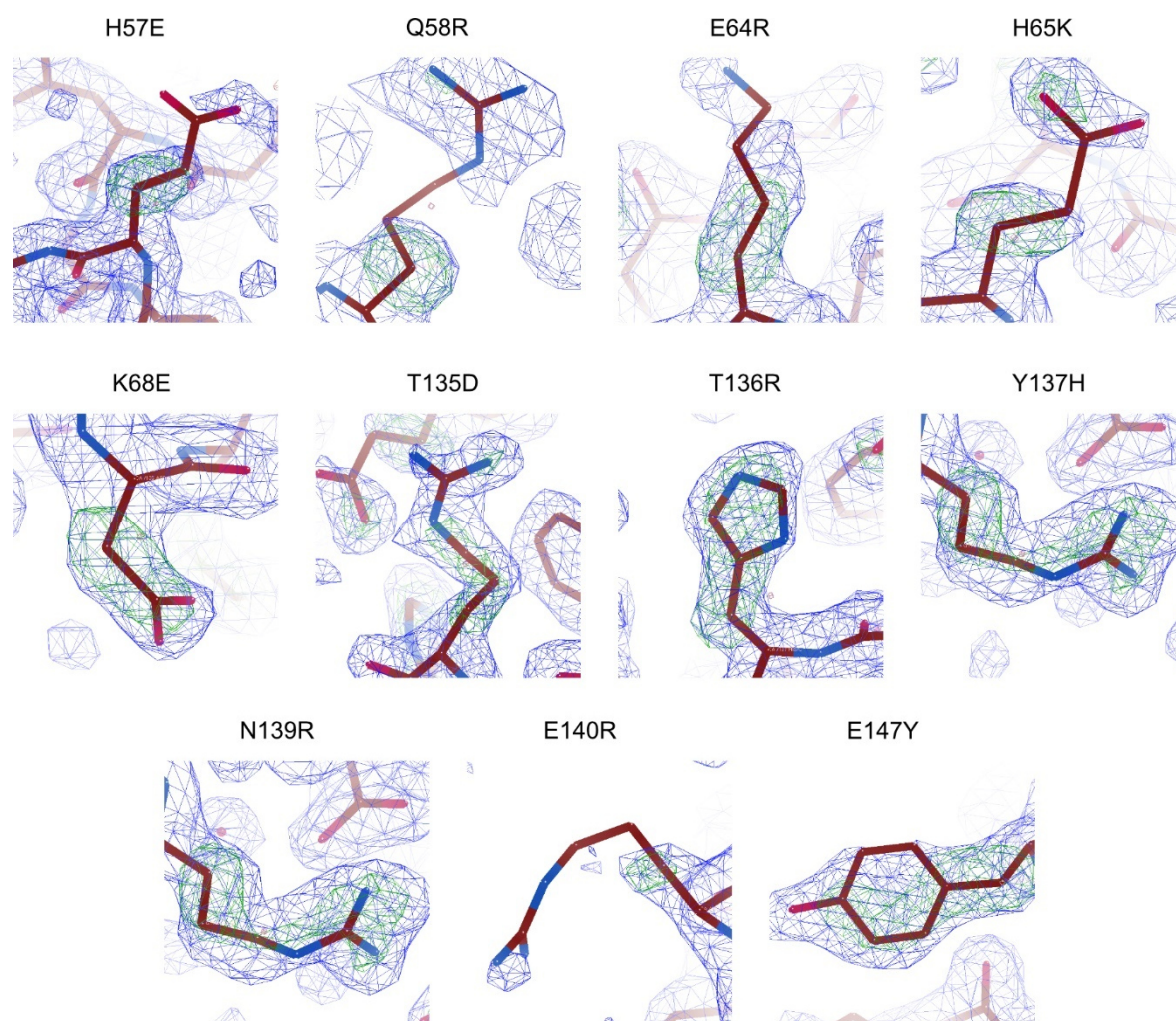


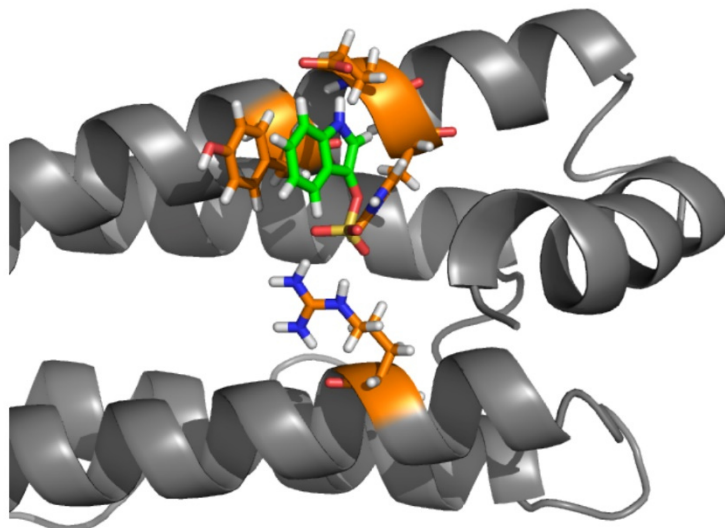
Figure 10.27: Electron density omit maps for $Ftn^{(neg)}$ -dock43. Electron density ($2F_o-F_c$ omit map, blue) and difference electron density (F_o-F_c , green) map. $2F_o-F_c$ map (blue): 1 RMSD, F_o-F_c (green): 5 RMSD. Maps were calculated without atoms of the introduced amino acids.

Table 10.10. Highest capacities for respective toxin

Toxin	Variant	Capacity	Form
IDS			
	Ftn^(neg)-Ap4-3A	985 ± 11	Non-crystalline
	Ftn^(neg)-Ap16	686 ± 2	Non-crystalline
	Ftn^(neg)-Dock03	685 ± 140	Non-Crystalline
	Ftn^(neg)-Ap4	672 ± 27	Non-crystalline
	Ftn^(neg)-Dock03	521 ± 46	crystalline
pCS			
	Ftn^(neg)-Ap4	697 ± 61	Non-crystalline
	Ftn^(neg)-Ap16	664 ± 58	Crystalline
	Ftn^(neg)-Ap16	598 ± 58	Non-crystalline
	Ftn^(neg)-Dock23	570 ± 5	Non-crystalline
	Ftn^(neg)-Dock23	470 ± 130	Crystalline
PAA			
	Ftn^(neg)-Dock43	13225 ± 840	Non-crystalline
	Ftn^(neg)	9659 ± 1100	Non-crystalline

Results for AP7 are excluded

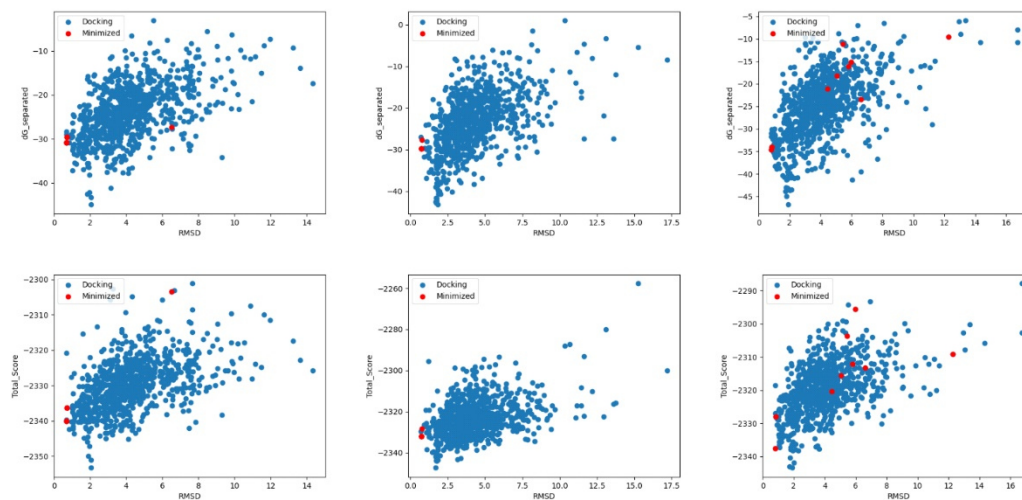
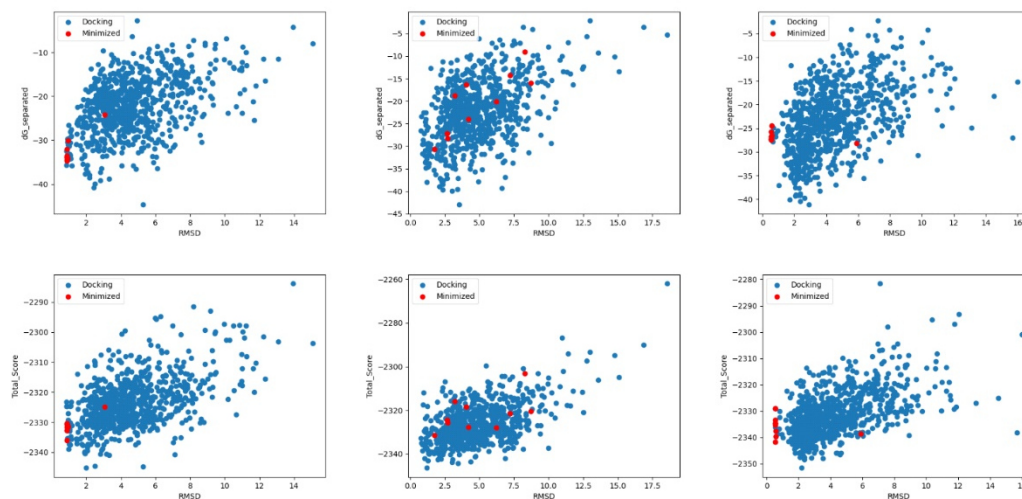
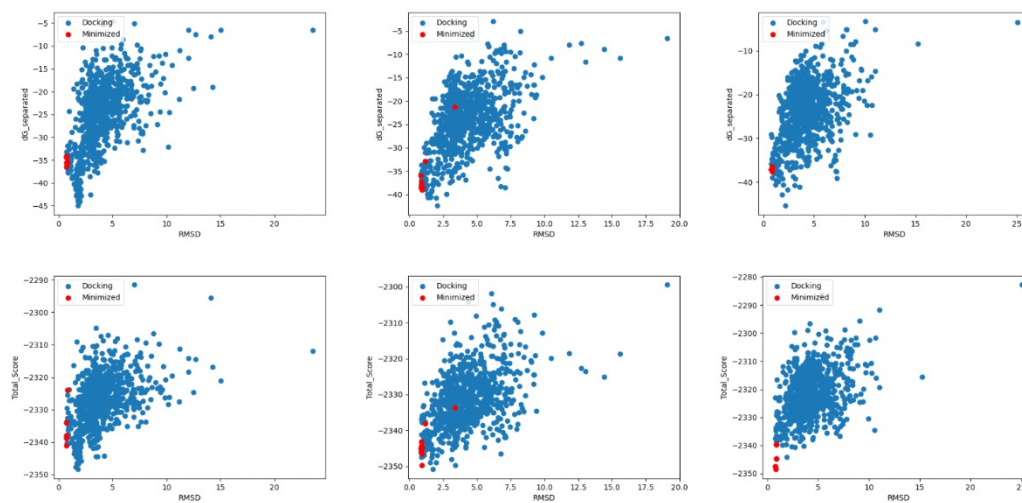
		60		140	
Ftn(neg)	NFAKYFLHQS	HEEREHAEKL	FIETHYLN	EQVKAIKELG	D
Dock43ER.RK..E.	...DRH.R	R.....Y...	.
Dock43_new RRK..E.	...DRH.R	R.....Y...	E



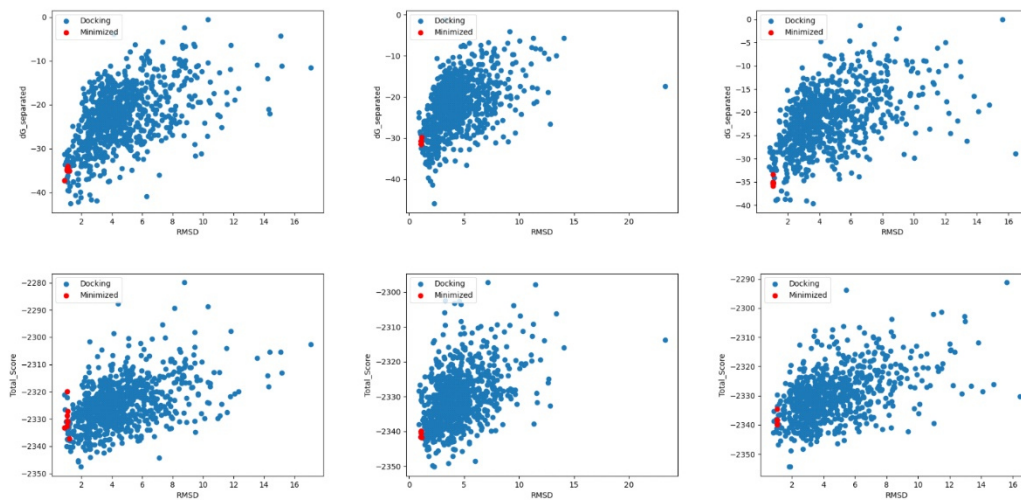
dG_separated: -14.26

Figure 10.28: Improved binding site of Ftn^(neg)-dock43. Binding site of dock43 was improved by the addition of two more residues highlighted in red in the sequence alignment. Especially the new hydrogen bridges between the sulfate group of the toxin and the functional group of the introduced arginine (Y54R) improve the affinity of this binding site.

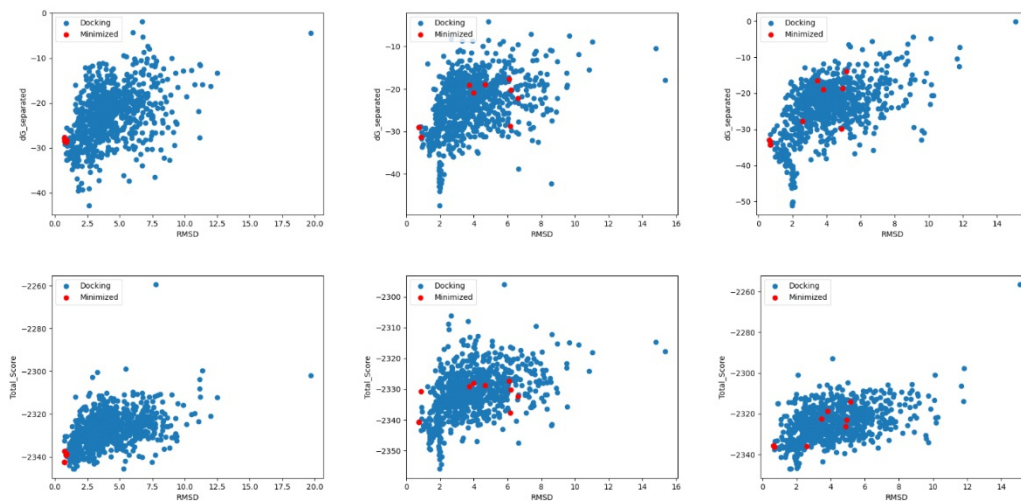
10.4 Results from protein-protein docking with binary protein container

Ftn^(pos) - Ftn^(neg) Contact A1Ftn^(pos) - Ftn^(neg) Contact A2Ftn^(pos) - Ftn^(neg)m1 Contact A1

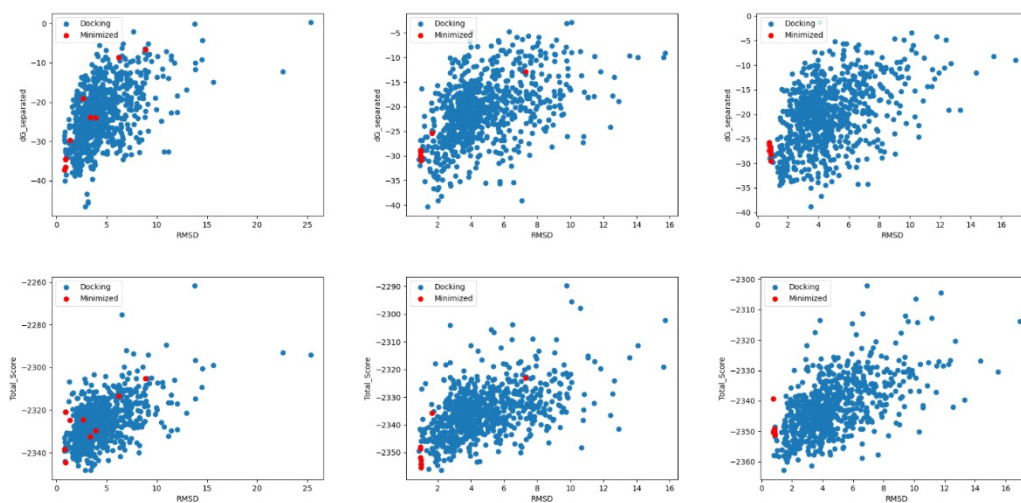
Ftn^(pos) - Ftn^(neg)m1 Contact A2



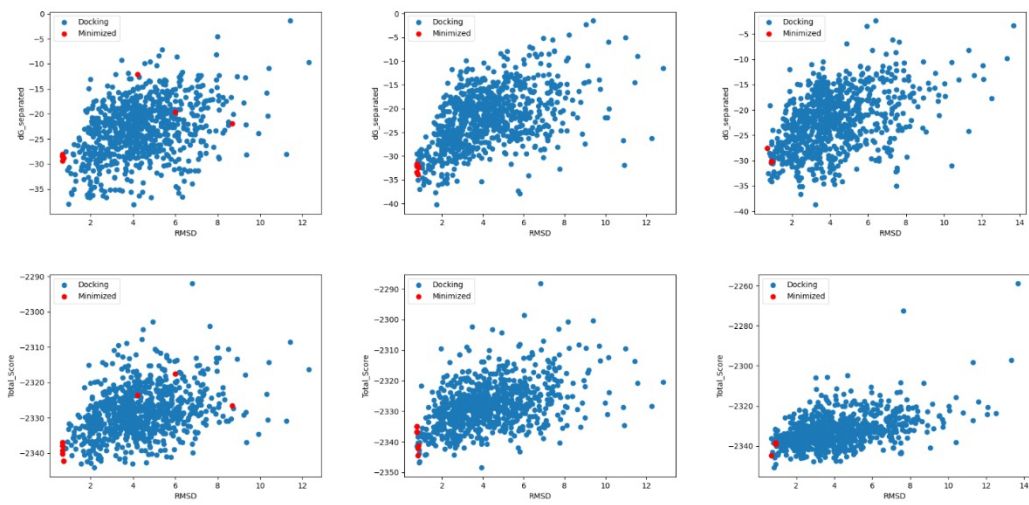
Ftn^(pos) - Ftn^(neg)m3 Contact A1



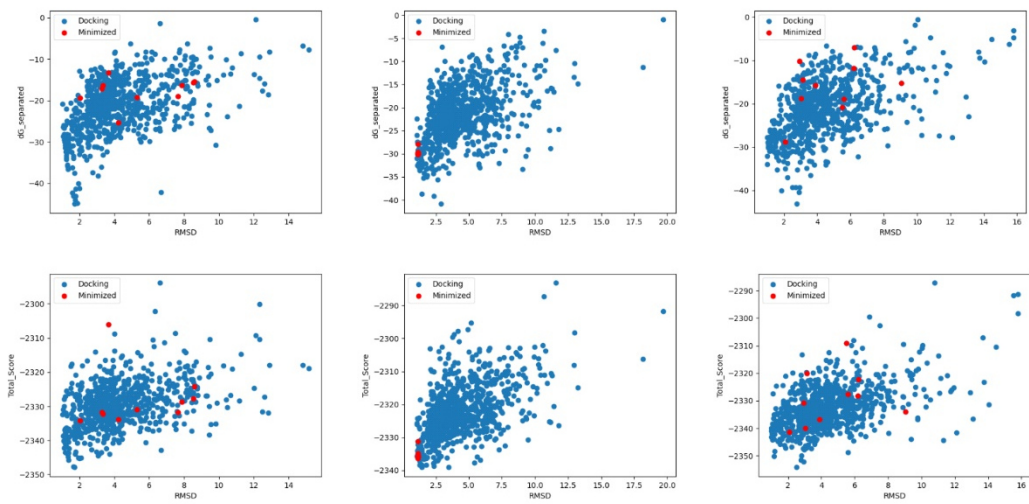
Ftn^(pos) - Ftn^(neg)m3 Contact A2



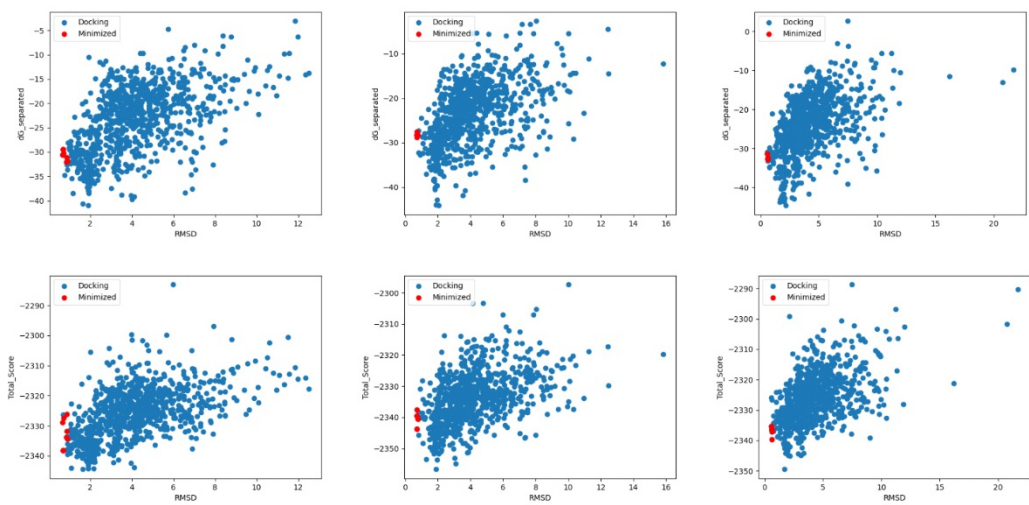
Ftn^(pos) - Ftn Contact A1



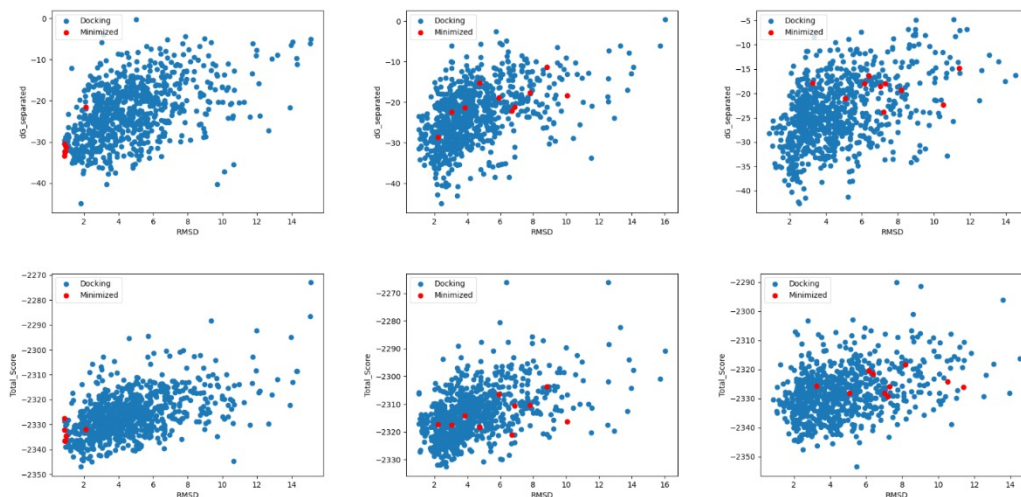
Ftn^(pos) - Ftn Contact A2



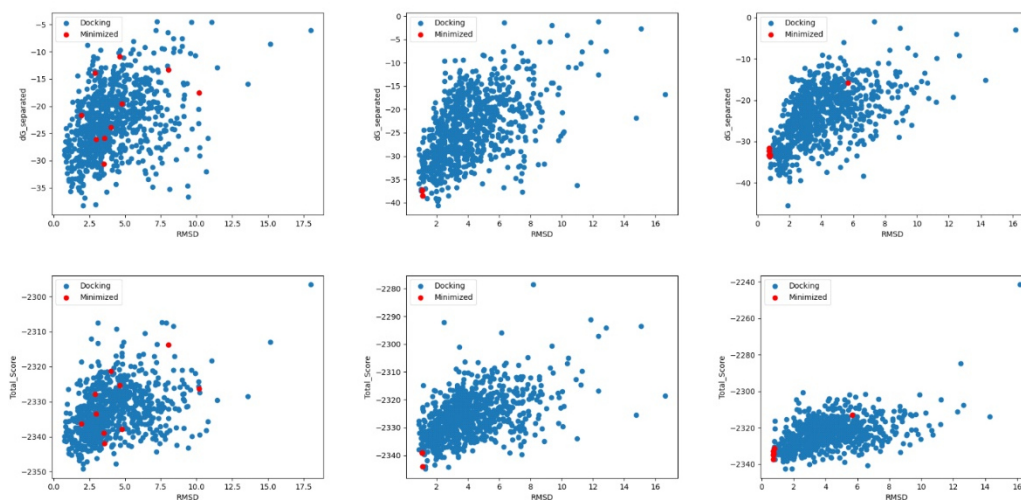
Ftn^(pos)m1- Ftn^(neg) Contact A1



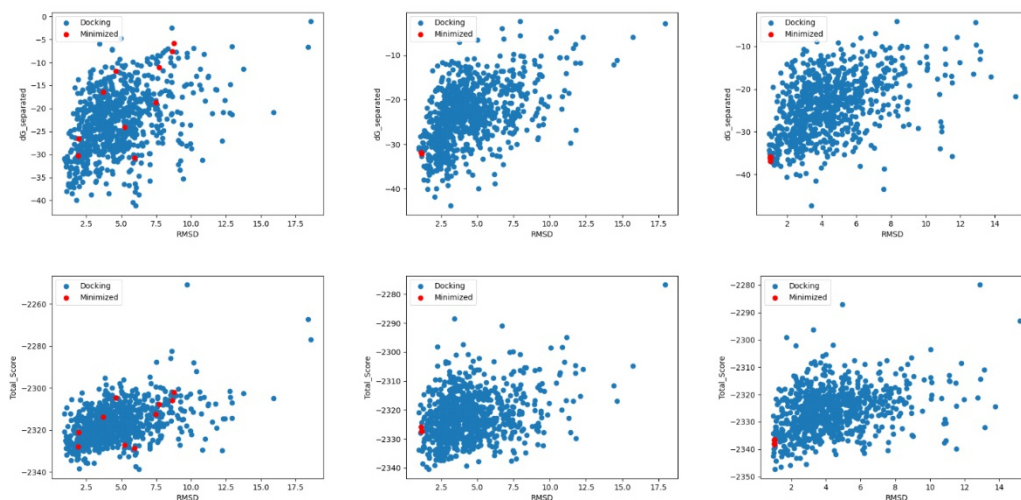
Ftn^(pos)m1- Ftn^(neg) Contact A2



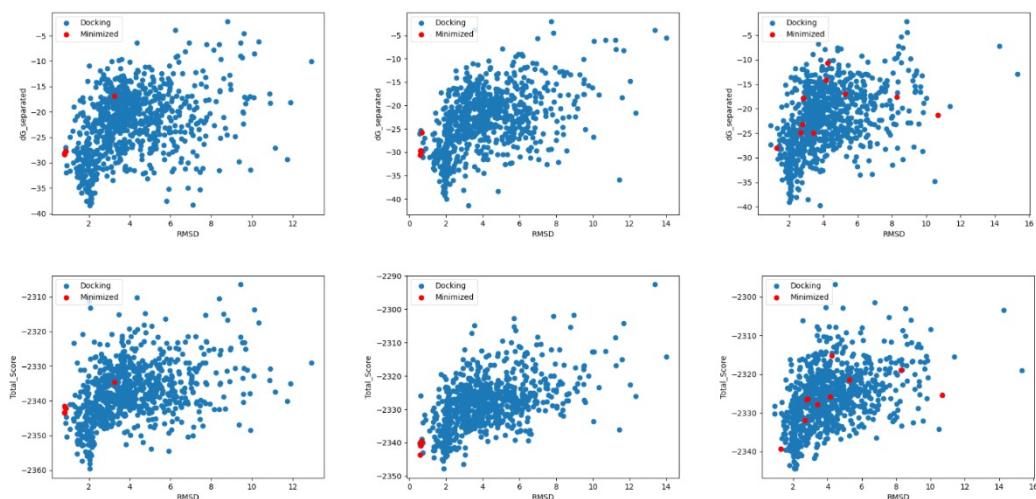
Ftn^(pos)m1- Ftn^(neg)m1 Contact A1



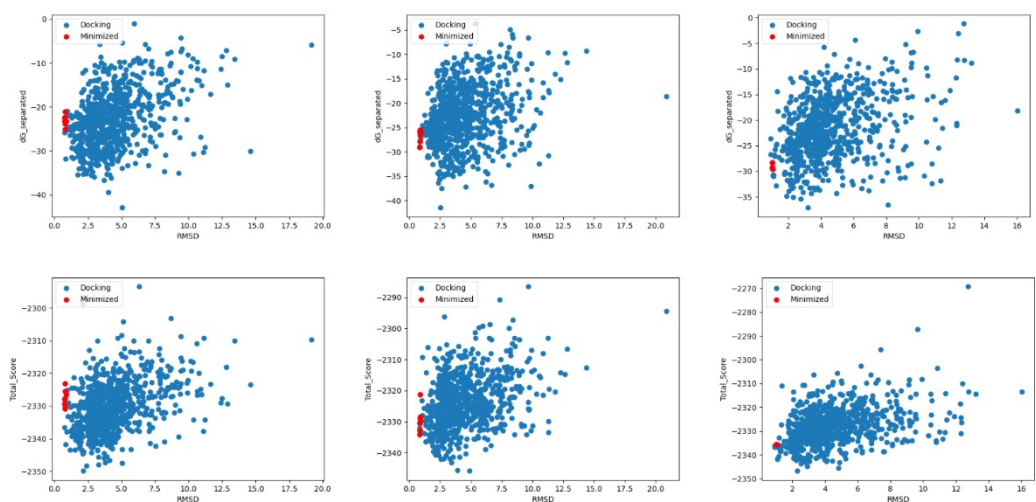
Ftn^(pos)m1- Ftn^(neg)m1 Contact A2



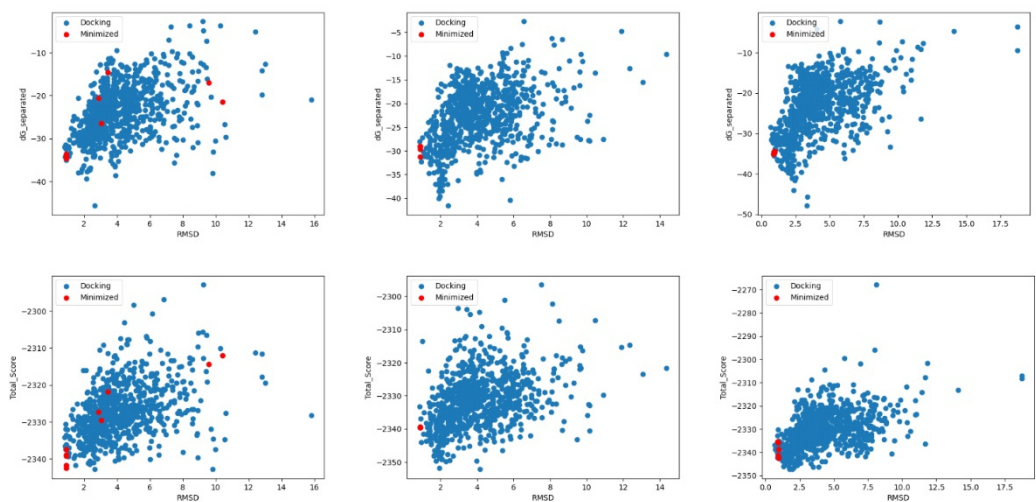
Ftn^(pos)m1- Ftn^(neg)m3 Contact A1



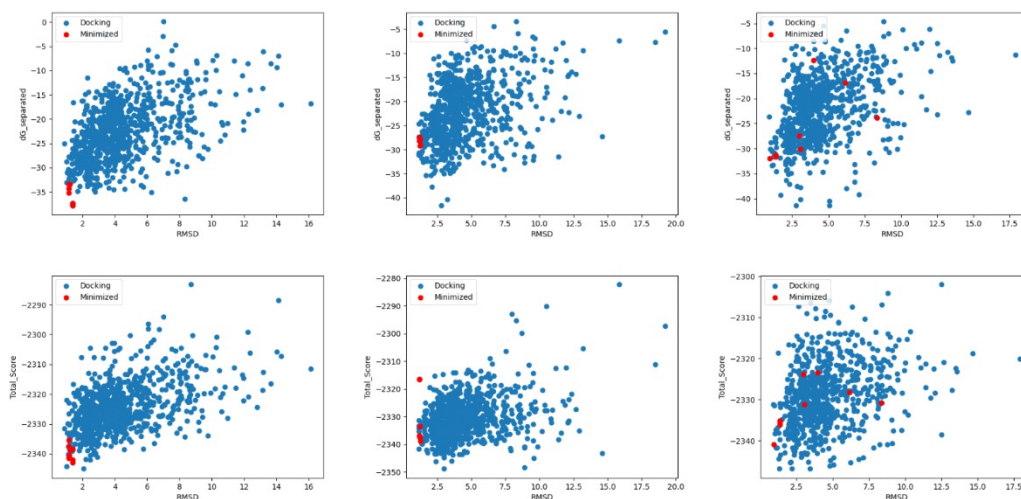
Ftn^(pos)m1- Ftn^(neg)m3 Contact A2



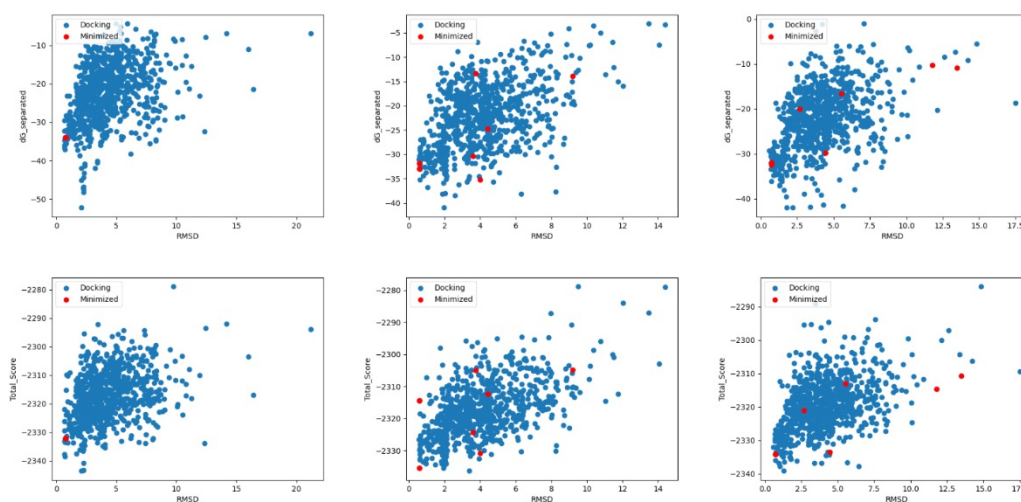
Ftn^(pos)m1- Ftn Contact A1



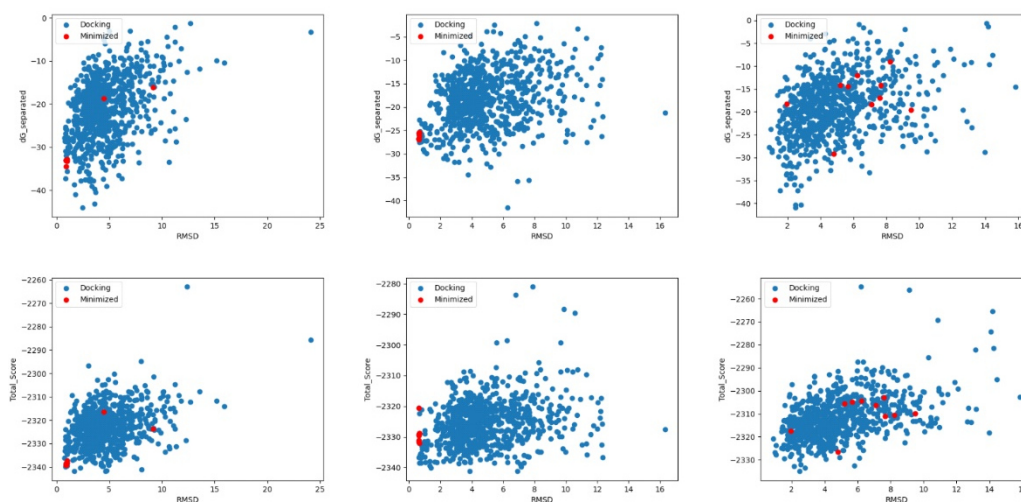
Ftn^(pos)m1- Ftn Contact A2



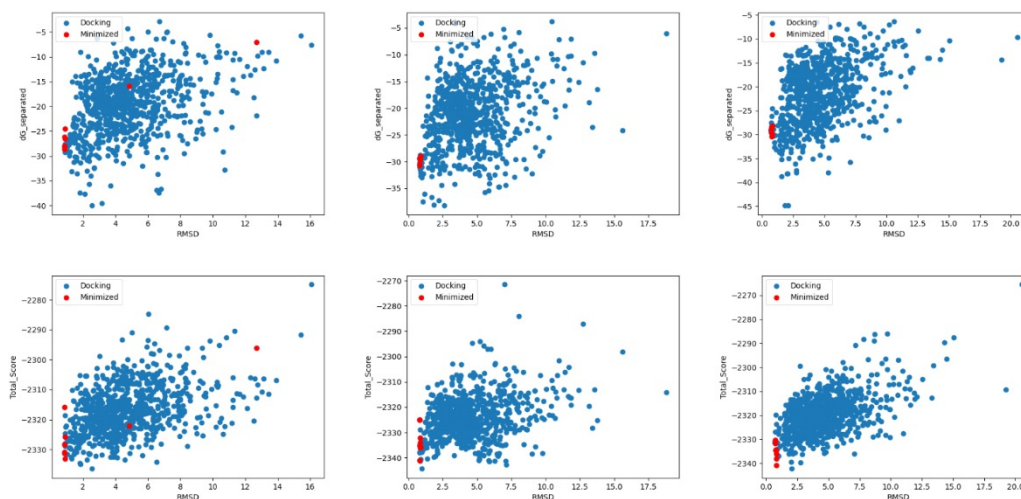
Ftn^(pos)m4- Ftn^(neg) Contact A1



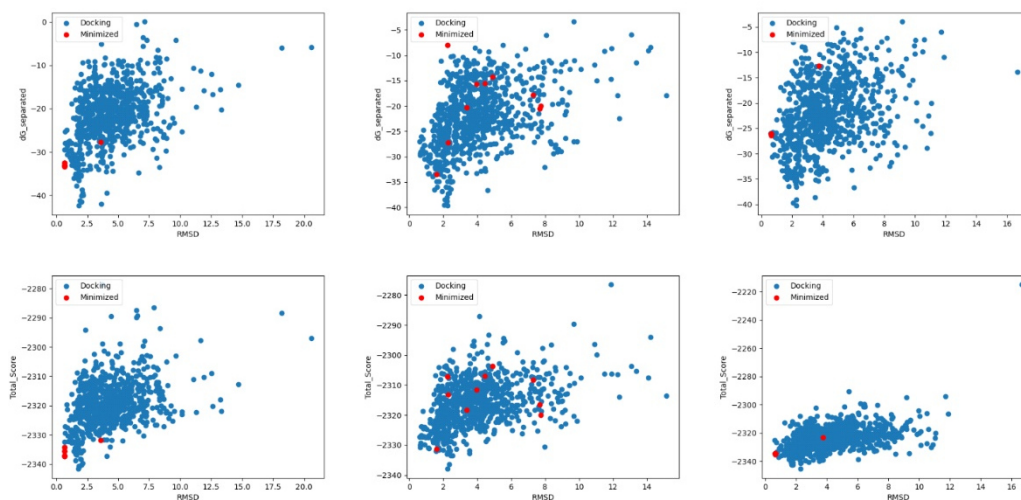
Ftn^(pos)m4- Ftn^(neg)m1 Contact A1



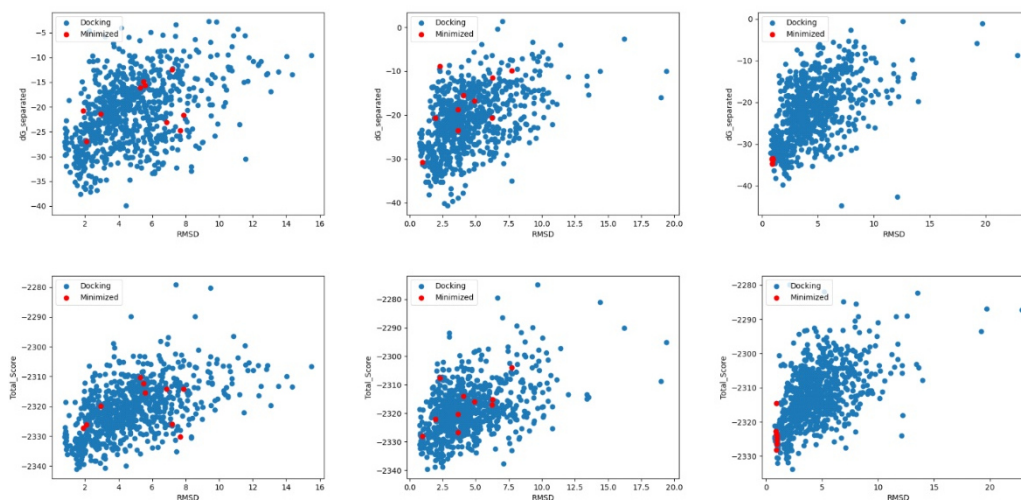
Ftn^(pos)m4- Ftn^(neg)m1 Contact A2



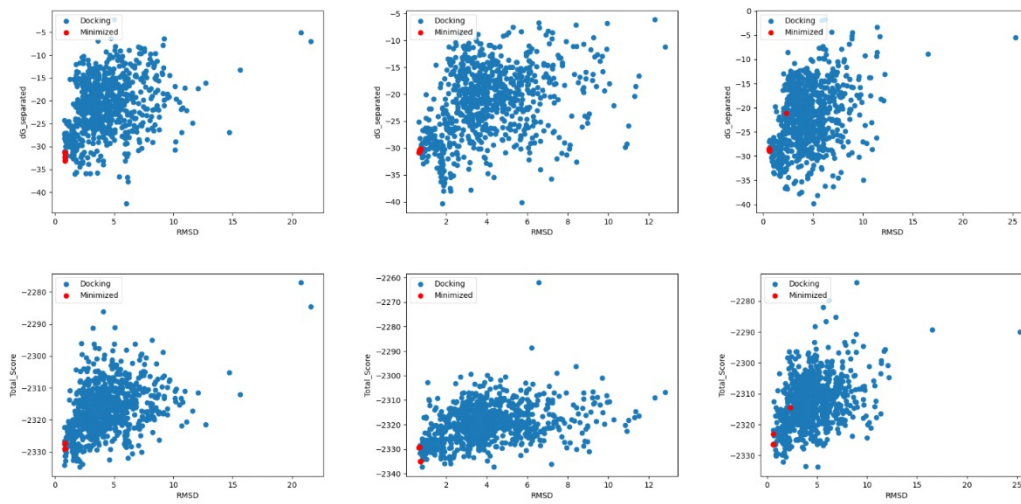
Ftn^(pos)m4- Ftn^(neg)m3 Contact A1



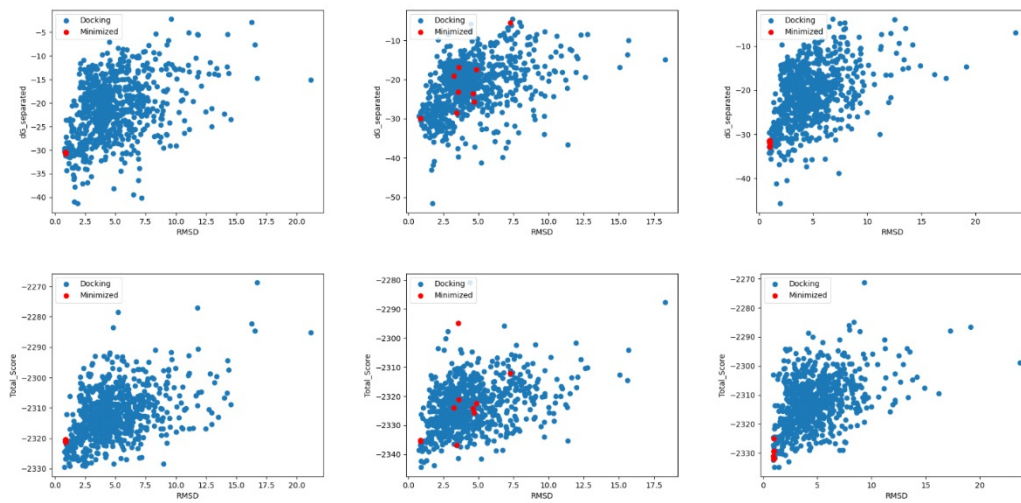
Ftn^(pos)m4- Ftn^(neg)m3 Contact A2



Ftn^(pos)m4- Ftn^(neg) Contact A1



Ftn^(pos)m4- Ftn^(neg) Contact A2



10.5 Scripts and files used for computational protein design

10.5.1 Compilation of Rosetta

```
#!/bin/bash
#SBATCH --job-name=Unpack
#SBATCH --partition=std
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e

source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
./scons.py bin mode=release extras=mpi
```

10.5.2 Generate symmetry definition file

10.5.2.1 Bash Script

```
#!/bin/bash
#SBATCH --job-name=Symm
#SBATCH --nodes=2
#SBATCH --tasks-per-node=16
#SBATCH --time=00:10:00
#SBATCH --export=NONE
```

```
set -e
```

```
source /sw/batch/init.sh
```

```
/work/bay1927/rosetta_software/rosetta_src_2020.08.61146_bundle/main
/source/src/apps/public/symmetry/make_symmdef_file.pl -m CRYST -r 24
-s P 4 3 2 -p 2CEI_reduced.pdb > 2CEI_p432_r24.symm |tee logfile
```

10.5.2.2 Input files

- PDB file of ferritin (PDB-ID: 2CEI) without heavy metal ions and water and manually added mutations for Ftn^(neg).

10.5.3 Relax protocol

10.5.3.1 Script

```
#!/bin/bash
#SBATCH --job-name=Relax
#SBATCH --partition=std
#SBATCH --nodes=12
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
module list
export LD_LIBRARY_PATH="$LD_RUN_PATH"

export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/release/linux/4.14/64/x86/gcc/8.2/mpi/

mpirun
//work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/bin/relax.mpi.linuxgccrelease @options_relax
```

10.5.3.2 Input files

- Symmetry definition file
 - will be generated by the script in 10.5.2
- PDB input file
 - PDB file of ferritin (PDB-ID: 2CEI) without heavy metal ions and water and manually added mutations for Ftn^(neg).
- Weights.wts file

```
METHOD WEIGHTS ref 0.773742 0.443793 -1.63002 -1.96094 0.61937
0.173326 0.388298 1.0806 -0.358574 0.761128 0.249477 -1.19118 -
0.250485 -1.51717 -0.32436 0.165383 0.20134 0.979644 1.23413
0.162496
fa_atr 1
fa_rep 0.55
fa_sol 0.9375
fa_intra_rep 0.005
fa_elec 0.875
pro_close 1.25
hbond_sr_bb 1.17
hbond_lr_bb 1.17
hbond_bb_sc 1.17
hbond_sc 1.1
dslf_fa13 1.25
rama 0.25
omega 0.625
```



```
fa_dun 0.7
p_aa_pp 0.4
yhh_planarity 0.625
ref 1
```

- **Options_relax**

```
-database
/work/bay4301/protein_design/rosetta_3.11/rosetta_src_2019.35.60890_
bundle/main/database
-s pseudo_D1.pdb
-ex1
-ex2
#-ex3
#-ex4
-use_input_sc
-constrain_relax_to_start_coords
-nstruct 500
#-minimize_sidechains
-run:seed_offset 100000
#-linmem_ig 10
-symmetry::symmetry_definition pD1_p432_r24.symm
-score:weights weights.wts
```

10.5.4 Fixbb protocol

10.5.4.1 Script

```
#!/bin/bash
#SBATCH --job-name=Relax
#SBATCH --partition=std
#SBATCH --nodes=12
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
module list
export LD_LIBRARY_PATH="$LD_RUN_PATH"

export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/
release/linux/4.14/64/x86/gcc/8.2/mpi/
```

```
mpirun
//work/bay4301/protein_design/rosetta_3.11/rosetta_src_2019.35.6089
0_bundle/main/source/bin/fixbb.mpi.linuxgccrelease @options_fixbb
```

10.5.4.2 Input files

- Symmetry definition file
 - will be generated by the script in 10.5.2
- PDB input file
 - energy minimized structure from relax run 10.5.3
- Weights.wts file
 - See section 10.5.3.2
- Options_fixbb

```
-database
/work/bay4301/protein_design/rosetta_3.11/rosetta_src_2019.35.6089_
bundle/main/database
-s pD1_relaxed.pdb
-resfile res_apolar+FWY
-ex1
-ex2
-use_input_sc
# -constrain_relax_to_start_coords
-nstruct 2000
-run:seed_offset 100000
-symmetry::symmetry_definition 2CEI_p432_r24.symm
-score:weights weights.wts
```

10.5.4.3 Residue files

For Ftn^(neg)-Ap4

```
#Header
NATAA
#Body
start
61 A PIKAA AVLIPWFE
62 A PIKAA AVLIPWFE
131 A PIKAA FWYD
140 A PIKAA AVLIPWFE
147 A PIKAA AVLIPWFE
150 A PIKAA AVLIPWFD
```

For Ftn^(neg)-Ap7

Appendix

```
#Header
NATAA
#Body
start
53 A PIKAA FWYK
61 A PIKAA AVLIPWFE
62 A PIKAA AVLIPWFE
131 A PIKAA FWYD
139 A PIKAA AVLIPWFN
140 A PIKAA AVLIPWFE
146 A PIKAA AVLIPWFK
147 A PIKAA AVLIPWFE
150 A PIKAA AVLIPWFD
```

For Ftn^(neg)-Ap16

```
#Header
NATAA
#Body
start
49 A PIKAA AVLIPWFK
50 A PIKAA AVLIPWFN
52 A PIKAA AVLIPWF
53 A PIKAA AVLIPWFK
54 A PIKAA AVLIPWFY
57 A PIKAA AVLIPWFH
58 A PIKAA AVLIPWFQ
61 A PIKAA AVLIPWFE
62 A PIKAA AVLIPWFE
64 A PIKAA AVLIPWFE
65 A PIKAA AVLIPWFH
68 A PIKAA AVLIPWFK
69 A PIKAA AVLIPWF
72 A PIKAA AVLIPWF
75 A PIKAA AVLIPWFQ
76 A PIKAA AVLIPWFR
128 A PIKAA AVLIPWFH
131 A PIKAA AVLIPWFD
135 A PIKAA AVLIPWFT
136 A PIKAA AVLIPWFH
139 A PIKAA AVLIPWFN
140 A PIKAA AVLIPWFE
143 A PIKAA AVLIPWFK
146 A PIKAA AVLIPWFK
147 A PIKAA AVLIPWFE
150 A PIKAA AVLIPWFD
151 A PIKAA AVLIPWFH
154 A PIKAA AVLIPWFN
```

Ferritin with removed ferroxidase site and iron transport pathway

```
#Header
```

Appendix

NATAA

#Body

start

27 A PIKAA GAVILMFYWPRHK
34 A PIKAA GAVILMFYWPRHK
61 A PIKAA GAVILMFYWPRHK
62 A PIKAA GAVILMFYWPRHK
65 A PIKAA GAVILMFYWPRHK
107 A PIKAA GAVILMFYWPRHK
118 A PIKAA GAVILMFYWPRHK
122 A PIKAA GAVILMFYWPRHK
123 A PIKAA GAVILMFYWPRHK
124 A PIKAA GAVILMFYWPRHK
125 A PIKAA GAVILMFYWPRHK
126 A PIKAA GAVILMFYWPRHK
127 A PIKAA GAVILMFYWPRHK
128 A PIKAA GAVILMFYWPRHK
129 A PIKAA GAVILMFYWPRHK
130 A PIKAA GAVILMFYWPRHK
131 A PIKAA GAVILMFYWPRHK
134 A PIKAA GAVILMFYWPRHK
135 A PIKAA GAVILMFYWPRHK
136 A PIKAA GAVILMFYWPRHK
137 A PIKAA GAVILMFYWPRHK
140 A PIKAA GAVILMFYWPRHK
141 A PIKAA GAVILMFYWPRHK

For Ftn^(neg)-3A

Header

NATAA

#Body

start

123 A PIKAA A
131 A PIKAA A
134 A PIKAA A

For Ftn^(neg)-5A

Header

NATAA

#Body

start

122 A PIKAA A
131 A PIKAA A
132 A PIKAA A
134 A PIKAA A
135 A PIKAA A

For Ftn^(neg)-8A

```
Header
NATAA
#Body
Start
122 A PIKAA A
123 A PIKAA A
125 A PIKAA A
131 A PIKAA A
132 A PIKAA A
134 A PIKAA A
135 A PIKAA A
139 A PIKAA A
```

For Ftn^(neg)-12A

```
Header
NATAA
#Body
start
122 A PIKAA A
123 A PIKAA A
124 A PIKAA A
125 A PIKAA A
126 A PIKAA A
127 A PIKAA A
128 A PIKAA A
129 A PIKAA A
131 A PIKAA A
134 A PIKAA A
135 A PIKAA A
140 A PIKAA A
```

10.5.5 Script to sort the Relax and Fixbb output

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=all
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --time=00:10:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh

grep -H 'SCORE' score.sc |sort -k2 -n|head -n20|nl -w 3 -s ' ' >Top20
```

10.5.6 Protein-ligand docking

10.5.6.1 Script to generate ligand params file

10.5.6.1.1 Script

```
#!/bin/bash
#SBATCH --job-name=Relax
#SBATCH --partition=std
#SBATCH --nodes=12
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh

python
//work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.6089
0_bundle/main/source/scripts/python/public/molfile_to_params.py -n 3IDS -p
3IDS --conformers-in-one-file 3IDS_conformer.sdf
```

The content of the param file should look like following

```
NAME 3ID
IO_STRING 3ID Z
TYPE LIGAND
AA UNK
ATOM C1 aroC X 0.07
ATOM O4 OH X -0.48
ATOM S1 S X 0.02
ATOM O1 OOC X -0.58
ATOM O2 OOC X -0.58
ATOM O3 OOC X -0.58
ATOM C2 aroC X 0.07
ATOM N1 Ntrp X -0.43
ATOM C3 aroC X 0.07
ATOM C4 aroC X 0.07
ATOM C5 aroC X 0.07
ATOM C6 aroC X 0.07
ATOM C7 aroC X 0.07
ATOM C8 aroC X 0.07
ATOM H6 Haro X 0.30
ATOM H5 Haro X 0.30
ATOM H4 Haro X 0.30
ATOM H3 Haro X 0.30
ATOM H1 Hpol X 0.61
ATOM H2 Haro X 0.30
BOND_TYPE O1 S1 2
BOND_TYPE S1 O2 2
BOND_TYPE S1 O3 1
BOND_TYPE S1 O4 1
BOND_TYPE O4 C1 1
BOND_TYPE C1 C2 4
BOND_TYPE C1 C8 4
BOND_TYPE C2 N1 4
BOND_TYPE C2 H2 1
BOND_TYPE N1 H1 1
```

```

BOND_TYPE  N1  C3  4
BOND_TYPE  C3  C4  4
BOND_TYPE  C3  C8  4
BOND_TYPE  C4  C5  4
BOND_TYPE  C4  H3  1
BOND_TYPE  C5  C6  4
BOND_TYPE  C5  H4  1
BOND_TYPE  C6  C7  4
BOND_TYPE  C6  H5  1
BOND_TYPE  C7  C8  4
BOND_TYPE  C7  H6  1
CHI 1  C1  O4  S1  O1
CHI 2  C2  C1  O4  S1
NBR_ATOM  C1
NBR_RADIUS 5.281226
ICOOR_INTERNAL  C1      0.000000    0.000000    0.000000    C1  O4  S1
ICOOR_INTERNAL  O4      0.000000   180.000000    1.351279    C1  O4  S1
ICOOR_INTERNAL  S1      0.000000    55.385017    1.640004    O4  C1  S1
ICOOR_INTERNAL  O1     59.355374    74.912598    1.463389    S1  O4  C1
ICOOR_INTERNAL  O2    -121.001449    74.923360    1.463128    S1  O4  O1
ICOOR_INTERNAL  O3    -119.504337    76.646799    1.463464    S1  O4  O2
ICOOR_INTERNAL  C2      3.468428    52.627197    1.374613    C1  O4  S1
ICOOR_INTERNAL  N1    -179.662451    72.176240    1.366771    C2  C1  O4
ICOOR_INTERNAL  C3      0.042555    69.605580    1.377423    N1  C2  C1
ICOOR_INTERNAL  C4     179.518199    49.640673    1.400205    C3  N1  C2
ICOOR_INTERNAL  C5    -179.925639    62.439472    1.398710    C4  C3  N1
ICOOR_INTERNAL  C6     -0.171186    58.935549    1.393767    C5  C4  C3
ICOOR_INTERNAL  C7      0.189164    59.369305    1.399949    C6  C5  C4
ICOOR_INTERNAL  C8      0.034575    60.555520    1.409540    C7  C6  C5
ICOOR_INTERNAL  H6     179.775227    59.169694    1.086215    C7  C6  C8
ICOOR_INTERNAL  H5     179.820772    59.994903    1.086531    C6  C5  C7
ICOOR_INTERNAL  H4     179.936479    60.604960    1.086964    C5  C4  C6
ICOOR_INTERNAL  H3     179.788635    58.567883    1.084278    C4  C3  C5
ICOOR_INTERNAL  H1     179.712635    55.346529    1.008869    N1  C2  C3
ICOOR_INTERNAL  H2     179.714476    49.048477    1.082605    C2  C1  N1
PDB_ROTAMERS 3ID_conformers.pdb

```

10.5.6.1.2 Input files

- Ligand chemical structure file
- Conformer ensemble for the ligand also in sdf format

10.5.6.2 Script to start ligand docking protocol

```

#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"

```

```

export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/
release/linux/4.14/64/x86/gcc/8.2/mpi/
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/src/relea
se/linux/4.14/64/x86/gcc/8.2/mpi/

mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@options_design.txt -nstruct 5000 > log.txt

```

10.5.6.2.1 Ligand docking RosettaScript

```

<ROSETTASCRIPTS>

  <SCOREFXNS>
    <ScoreFunction name="ligand_soft_rep" weights="ligand_soft_rep">
    </ScoreFunction>
    <ScoreFunction name="hard_rep" weights="ligand">
    </ScoreFunction>
  </SCOREFXNS>

  <TASKOPERATIONS>
    <DetectProteinLigandInterface name="design_interface" cut1="6.0" cut2="8.0"
cut3="10.0" cut4="12.0" design="1" resfile="resfile_dock22.txt"/> # see Note 19
  </TASKOPERATIONS>
  <LIGAND_AREAS>
    <LigandArea name="inhibitor_dock_sc" chain="X" cutoff="6.0"
add_nbr_radius="true" all_atom_mode="false"/>
    <LigandArea name="inhibitor_final_sc" chain="X" cutoff="6.0"
add_nbr_radius="true" all_atom_mode="false"/>
    <LigandArea name="inhibitor_final_bb" chain="X" cutoff="7.0"
add_nbr_radius="false" all_atom_mode="true" Calpha_restraints="0.3"/>
  </LIGAND_AREAS>

  <INTERFACE_BUILDERS>
    <InterfaceBuilder name="side_chain_for_docking"
ligand_areas="inhibitor_dock_sc"/>
    <InterfaceBuilder name="side_chain_for_final"
ligand_areas="inhibitor_final_sc"/>
    <InterfaceBuilder name="backbone" ligand_areas="inhibitor_final_bb"
extension_window="3"/>
  </INTERFACE_BUILDERS>

  <MOVEMAP_BUILDERS>
    <MoveMapBuilder name="docking" sc_interface="side_chain_for_docking"
minimize_water="false"/>
    <MoveMapBuilder name="final" sc_interface="side_chain_for_final"
bb_interface="backbone" minimize_water="false"/>
  </MOVEMAP_BUILDERS>

  <SCORINGGRIDS ligand_chain="X" width="15">
    <ClassicGrid grid_name="classic" weight="1.0"/>
  </SCORINGGRIDS>

  <MOVERS>

    <FavorNativeResidue name="favor_native" bonus="1.00" /> # see Notes 21 and 22
      <Transform name="transform" chain="X" box_size="5.0"
move_distance="0.1" angle="5" cycles="500" repeats="1"
temperature="5" rmsd="4.0" /> # see Note 23
    <HighResDocker name="high_res_docker" cycles="6" repack_every_nth="3"
scorefxn="ligand_soft_rep" movemap_builder="docking"/>
    <PackRotamersMover name="design_interface" scorefxn="hard_rep"
task_operations="design_interface"/>
    <FinalMinimizer name="final" scorefxn="hard_rep" movemap_builder="final"/>

```



```

    <InterfaceScoreCalculator name="add_scores" chains="X" scorefxn="hard_rep" />
    <ParsedProtocol name="low_res_dock">
      <Add mover_name="transform"/>
    </ParsedProtocol>
    <ParsedProtocol name="high_res_dock">
      <Add mover_name="high_res_docker"/>
      <Add mover_name="final"/>
    </ParsedProtocol>
  </MOVERS>

  <PROTOCOLS>

    <Add mover_name="favor_native"/>
    <Add mover_name="low_res_dock"/>
    <Add mover_name="design_interface"/> # see Note 24
    <Add mover_name="high_res_dock"/>
    <Add mover_name="add_scores"/>
  </PROTOCOLS>
</ROSETTASCRIPTS>

```

10.5.6.2.2 Input files

- PDB input file
 - energy minimized structure from relax run 10.5.3
 - Ligand added manually to the structure
- Ligand pdb structure
- Ligand_confomer.pdb
 - Generated by scripts in 10.5.6.1
- Ligand-params file
- Residue file

```

ALLAA
AUTO
start
1 X NATAA

```

- Options_design file

```

-in
  -file
    -s 'input/D1_3IDS.pdb'
    -extra_res_fa input/3ID.params
  -packing
    -ex1
    -ex2
    -no_optH false
    -flip_HNQ true
    -ignore_ligand_chi true
  -parser
    -protocol design.xml
    -mistakes
    -restore_pre_talaris_2013_behavior true

```

10.5.6.3 Script to filter ligand-docking output

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/
release/linux/4.14/64/x86/gcc/8.2/mpi/
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/src/relea
se/linux/4.14/64/x86/gcc/8.2/mpi/

perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle//main/source/src/apps/public/enzdes/DesignSelect.pl -d <(grep
SCORE score.sc) -c metric_thresholds.txt -tag_column last >
filtered_designs.sc awk '{print $NF ".pdb"}' filtered_designs.sc>
filtered_pdb.txt
```

10.5.6.3.1 Input files

- **Metric_thresholds.txt**

```
req total_score value < -800 # measure of protein stability
req if_X_fa_rep value < 1.0 # measure of ligand clashes
req ligand_is_touching_X value > 0.5 # 1.0 if ligand is in pocket
req interface_delta_X value < -8.0
output sortmin interface_delta_X # binding energy
```

10.5.6.4 Script to start InterfaceAnalyzer for the ligand-docking models

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
```

```
#source /sw/base/env/gcc-8.2.0_openmpi-3.1.3
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/release/linux/4.14/64/x86/gcc/8.2/mpi/
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/src/release/linux/4.14/64/x86/gcc/8.2/mpi/

awk -F '[:,space:]]+' '{print $57 }' filtered_pdbs.txt > pdblist #
the -F '[:,space:]]+' sets the distance between two columns to one space or more
for k in `cat pdblist`

do
    echo $k
    echo "${PWD}/${k}.pdb" >> PDB_for_filter_2.txt
done

mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/bin/InterfaceAnalyzer.mpi.linuxgccrelease -
interface_A_X -compute_packstat true -pack_separated -score:weights
ligandprime -no_nstruct_label -out:file:score_only
design_interfaces.sc -l PDB_for_filter_2.txt -extra_res_fa
input/3ID.params
```

10.5.6.5 Script to sort output of the InterfaceAnalyzer

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
#source /sw/base/env/gcc-8.2.0_openmpi-3.1.3
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/release/linux/4.14/64/x86/gcc/8.2/mpi/
```

```
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/src/relea
se/linux/4.14/64/x86/gcc/8.2/mpi/
```

```
perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/src/apps/public/enzdes/DesignSelect.pl -d <(grep
SCORE design_interfaces.sc) -c metric_thresholds_2.txt -tag_column
last > design_interfaces_filtered.sc
awk '{print $NF ".pdb"}' design_interfaces_filtered.sc >
filtered_pdb2.txt
```

10.5.6.5.1 Input files

- **Metric_thresholds_2.txt**

```
req packstat value > 0.55 # packing metric; 0-1 higher better
req sc_value value > 0.45# shape complementarity;0-1 higher
better
req delta_unsatHbonds value < 0.5# unsatisfi ed hydrogen bonds
on binding
req dG_separated/dSASAx100 value < -0.5 # binding energy per
contact area
output sortmin dG_separated # binding energy
```

10.5.6.6 Script for automated ligand docking

All input files from the previous ligand docking scripts are needed.

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
#source /sw/base/env/gcc-8.2.0_openmpi-3.1.3
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/external/
release/linux/4.14/64/x86/gcc/8.2/mpi/
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/roset
ta_3.11/rosetta_src_2019.35.60890_bundle/main/source/build/src/relea
se/linux/4.14/64/x86/gcc/8.2/mpi/
```

```
#Create a origin_directory with input structures, design.xml,
metric_thresholds_1&2.txt ( change H-bond in
metric_thresholds_2.txt to 0.5)

# Create directory for the Results
mkdir Results

# First Generation

cp -r Blanko D1_Dock_22_Gen1
cd D1_Dock_22_Gen1

#Start Rosetta design
mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@options_design.txt -nstruct 5000 > log.txt

#Filter1
perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle//main/source/src/apps/public/enzdes/DesignSelect.pl -d
<(grep SCORE score.sc) -c metric_thresholds.txt -tag_column last >
filtered_designs.sc awk '{print $NF ".pdb"}' filtered_designs.sc >
filtered_pdbs.txt

#Calculate metrics
awk -F '[:,space:]+' '{print $57 }' filtered_pdbs.txt > pdblast2 #
the -F '[:,space:]+' sets the distance between two columns to one
space or more
for k in `cat pdblast2`
do
    #echo $k
    echo "${PWD}/${k}.pdb" >> PDB_for_filter_2.txt
done

mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/bin/InterfaceAnalyzer.mpi.linuxgccrelease -
interface A_X -compute_packstat true -pack_separated -score:weights
ligandprime -no_nstruct_label -out:file:score_only
design_interfaces.sc -l PDB_for_filter_2.txt -extra_res_fa
input/3ID.params

#Filter 2

perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/src/apps/public/enzdes/DesignSelect.pl -d <(grep
SCORE design_interfaces.sc) -c metric_thresholds_2.txt -tag_column
last > design_interfaces_filtered.sc

# Extract Fasta
```

Appendix

```
awk '{print $NF ".pdb"}' design_interfaces_filtered.sc >
filtered_pdbs2.txt
perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/scripts/python/public/pdb2fasta.py $(cat
filtered_pdbs2.txt) > selected_sequences.fasta

#Copy results to results folder an rename them
cp design_interfaces_filtered.sc
../Results/Gen1_design_interfaces.sc
cp selected_sequences.fasta ../Results/Gen1_selected_sequences.fasta
awk -F '[[[:space:]]]+' '{print $23 }'
design_interfaces_filtered.sc > pdblist3 # the -F '[[[:space:]]]+'
sets the distance between two columns to one space or more
sed -i '2,$d' pdblist3
for pdb in `cat pdblist3`
do
cp ${pdb}.pdb ../Results/Gen1_best_${pdb}.pdb
done
sed '3,$d' selected_sequences.fasta >
selected_sequences_cutted.fasta
cp selected_sequences_cutted.fasta ../Results/Gen1_Best.fasta
cd ..

#Main loop for every Gen after 1
for i in {2..10}
do
cp -r Blanko D1_Dock_22_Gen${i}

prevGen=$(bc <<< "${i} -1")
echo $prevGen
cd D1_Dock_22_Gen${prevGen}
pwd
# Select best PDB and copy it
awk -F '[[[:space:]]]+' '{print $23 }'
design_interfaces_filtered.sc > pdblist # the -F '[[[:space:]]]+'
sets the distance between two columns to one space or more
sed -i '2,$d' pdblist
for pdb in `cat pdblist`
do
cp ${pdb}.pdb
../D1_Dock_22_Gen${i}/input/${pdb}.pdb
cp ${pdb}.pdb
../D1_Dock_22_Gen${i}/input/D1_3IDS.pdb
done

cd ../D1_Dock_22_Gen${i}
#Start Rosetta design
mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@options_design.txt -nstruct 5000 > log.txt
#Filter1
perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
```

```
_bundle//main/source/src/apps/public/enzdes/DesignSelect.pl -d
<(grep SCORE score.sc) -c metric_thresholds.txt -tag_column last >
filtered_designs.sc awk '{print $NF ".pdb"}' filtered_designs.sc>
filtered_pdbs.txt
    #Calculate metrics
    awk -F '[[[:space:]]]+' '{print $57 }' filtered_pdbs.txt >
pdblist2 # the -F '[[[:space:]]]+' sets the distance between two
columns to one space or more
    for k in `cat pdblist2`
    do
        echo $k
        echo "${PWD}/${k}.pdb" >> PDB_for_filter_2.txt
    done

    mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/bin/InterfaceAnalyzer.mpi.linuxgccrelease -
interface A_X -compute_packstat true -pack_separated -score:weights
ligandprime -no_nstruct_label -out:file:score_only
design_interfaces.sc -l PDB_for_filter_2.txt -extra_res_fa
input/3ID.params
    #Filter 2

    perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/src/apps/public/enzdes/DesignSelect.pl -d <(grep
SCORE design_interfaces.sc) -c metric_thresholds_2.txt -tag_column
last > design_interfaces_filtered.sc

    # Extract Fasta
    awk '{print $NF ".pdb"}' design_interfaces_filtered.sc >
filtered_pdbs2.txt
    perl
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890
_bundle/main/source/scripts/python/public/pdb2fasta.py $(cat
filtered_pdbs2.txt) > selected_sequences.fasta

    #Copy results to results folder an rename them
    cp design_interfaces_filtered.sc
../Results/Gen${i}_design_interfaces.sc
    cp selected_sequences.fasta
../Results/Gen${i}_selected_sequencrs.fasta
    awk -F '[[[:space:]]]+' '{print $23 }'
design_interfaces_filtered.sc > pdblist3 # the -F '[[[:space:]]+'
sets the distance between two columns to one space or more
    sed -i '2,$d' pdblist3
        for pdb in `cat pdblist3`
        do
            cp ${pdb}.pdb ../Results/Gen${i}_best_${pdb}.pdb
        done
    sed '3,$d' selected_sequences.fasta >
selected_sequences_cutted.fasta
    cp selected_sequences_cutted.fasta
../Results/Gen${i}_Best.fasta
    cd ..
```

done

10.5.6.1 Script to start ligand docking protocol at various starting positions

```
#!/bin/bash
#SBATCH --job-name=Symmiqq
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
#source /sw/base/env/gcc-8.2.0_openmpi-3.1.3
module switch env env/gcc-8.2.0_openmpi-3.1.3

export LD_LIBRARY_PATH="$LD_RUN_PATH"
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11
/rosetta_src_2019.35.60890_bundle/main/source/build/external/release/linux/
4.14/64/x86/gcc/8.2/mpi/
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/bay4301/protein_designe/rosetta_3.11
/rosetta_src_2019.35.60890_bundle/main/source/build/src/release/linux/4.14/
64/x86/gcc/8.2/mpi/
step=-0.2
step1=-0.2

for i in {0..2}
do
    #move along the chain
    x=$(bc <<< "-18.988178 + ${i} * ${step1} * -17.134375")
    y=$(bc <<< "-3.909102 + ${i} * ${step1} * -34.925604")
    z=$(bc <<< " 17.752204 + ${i} * ${step1} * -1.303496")

    for j in {0..5}
    do
        # move across chains
        x1=$(bc <<< "${x} + ${j} * ${step} * -3.444504")
        y1=$(bc <<< "${y} + ${j} * ${step} * -2.097589")
        z1=$(bc <<< "${z} + ${j} * ${step} * 16.24169")
        sed -i "33s/.*/<Coordinates
x=\"${x1}\" y=\"${y1}\" z=\"${z1}\" \\/> /" dock.xml
        mkdir dock_${i}${j}
        cd dock_${i}${j}
        cp ../options_above.txt options_above.txt
        cp ../dock.xml dock.xml
        pwd

        mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle
/main/source/bin/rosetta_scripts.mpi.linuxgccrelease @options_above.txt -
nstruct 5 > log.txt

        #process_id=$!
        #wait $process_id
        cd ..
    done
done
```



```

done

done

for i in {3..5}
do
    #entlang der chain
    x=$(bc <<< "-18.988178 + ${i} * ${step1} * -17.134375")
    y=$(bc <<< "-3.909102 + ${i} * ${step1} * -34.925604")
    z=$(bc <<< " 17.752204 + ${i} * ${step1} * -1.303496")

    for j in {0..5}
    do
        # zwischen der chain

        x1=$(bc <<< "${x} + ${j} * ${step} * -7.675404")
        y1=$(bc <<< "${y} + ${j} * ${step} * 1.7000292")
        z1=$(bc <<< "${z} + ${j} * ${step} * 16.0404052")
        sed -i "33s/.*/<Coordinates
x="\${x1}\\" y="\${y1}\\" z="\${z1}\\" \/> /" dock.xml
        mkdir dock_${i}${j}
        cd dock_${i}${j}
        cp ../options_above.txt options_above.txt
        cp ../dock.xml dock.xml
        pwd

        mpirun
/work/bay4301/protein_designe/rosetta_3.11/rosetta_src_2019.35.60890_bundle
/main/source/bin/rosetta_scripts.mpi.linuxgccrelease @options_above.txt -
nstruct 5 > log.txt

        cd ..
    done
done

```

10.5.7 Protein-protein docking

10.5.7.1 Script to start full protein-protein docking

```

#!/bin/bash
#SBATCH --job-name=Docking
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
module list
export LD_LIBRARY_PATH="$LD_RUN_PATH"

```

```
export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/fcvx142/Rosetta_3.13/rosetta_src_2021.16.61629_bundle/main/source/build/src/release/linux/4.14/64/x86/gcc/8.2/mpi/
```

```
mpirun
/work/fcvx142/Rosetta_3.13/rosetta_src_2021.16.61629_bundle/main/source/bin/rosetta_scripts.mpi.linuxgccrelease @docking.options -
parser:protocol docking_full.xml -out:suffix _full -nstruct 1000 >&
docking_full.log
```

10.5.7.2 protein-protein docking RosettaScript for complete docking

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
  </SCOREFXNS>
  <TASKOPERATIONS>
    <InitializeFromCommandline name="ifcl"/>
    <RestrictToRepacking name="rtr" />
    <RestrictToInterfaceVector name="rtiv" chain1_num="1,2" chain2_num="3,4"
      CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75"
      vector_dist_cutoff="9.0" />
  </TASKOPERATIONS>
  <FILTERS>
  </FILTERS>
  <MOVERS>
    MINIMIZATION MOVERS
    Single cycle of FastRelax to minimize backbone of docking partners
    <FastRelax name="minimize_interface" scorefxn="REF2015" repeats="1"
      task_operations="ifcl,rtr,rtiv" />

    DOCKING MOVERS
    <Docking name="dock_low" score_low="score_docking_low" score_high="REF2015"
      fullatom="0" local_refine="0" optimize_fold_tree="1" conserve_foldtree="0"
      ignore_default_docking_task="0" design="0" task_operations="ifcl" jumps="1"/>
    <Docking name="dock_high" score_low="score_docking_low" score_high="REF2015"
      fullatom="1" local_refine="1" optimize_fold_tree="1" conserve_foldtree="0"
      design="0" task_operations="ifcl" jumps="1"/>

    <SaveAndRetrieveSidechains name="srsc" allsc="0" />
  </MOVERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
    Run docking protocol
    <Add mover="dock_low"/>
    <Add mover="srsc" />
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

```
<Add mover="dock_high" />

Minimize interface
<Add mover="minimize_interface" />

</PROTOCOLS>
<OUTPUT scorefxn="REF2015" />
</ROSETTASCRIPTS>
```

10.5.7.3 Script to start protein-protein docking

```
#!/bin/bash
#SBATCH --job-name=DockingMinimize
#SBATCH --partition=std
#SBATCH --nodes=1
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
module list
export LD_LIBRARY_PATH="$LD_RUN_PATH"

export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/fcvx142/Rosetta_3.13/rosetta_
src_2021.16.61629_bundle/main/source/build/src/release/linux/4.14/64
/x86/gcc/8.2/mpi/

mpirun
/work/fcvx142/Rosetta_3.13/rosetta_src_2021.16.61629_bundle/main/sou
rce/bin/rosetta_scripts.mpi.linuxgccrelease @docking.options -
parser:protocol docking_minimize.xml -out:suffix _minimize -nstruct
10 >& docking_minimize.log
```

10.5.7.4 Protein-Protein docking RosettaScript for minimized docking

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
</SCOREFXNS>
  <TASKOPERATIONS>
    <InitializeFromCommandline name="ifcl"/>
    <RestrictToRepacking name="rtr" />
```

```

    <RestrictToInterfaceVector name="rtiv" chain1_num="1,2" chain2_num="3,4"
    CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75"
    vector_dist_cutoff="9.0" />
</TASKOPERATIONS>
<FILTERS>
</FILTERS>
<MOVERS>
    MINIMIZATION MOVERS
    Single cycle of FastRelax to minimize backbone of docking partners
    <FastRelax name="minimize_interface" scorefxn="REF2015" repeats="1"
    task_operations="ifcl,rtr,rtiv" />

    DOCKING MOVERS
    <Docking name="dock_low" score_low="score_docking_low" score_high="REF2015"
    fullatom="0" local_refine="0" optimize_fold_tree="1" conserve_foldtree="0"
    ignore_default_docking_task="0" design="0" task_operations="ifcl" jumps="1"/>
    <Docking name="dock_high" score_low="score_docking_low" score_high="REF2015"
    fullatom="1" local_refine="1" optimize_fold_tree="1" conserve_foldtree="0"
    design="0" task_operations="ifcl" jumps="1"/>

    <SaveAndRetrieveSidechains name="srsc" allsc="0" />
</MOVERS>
<APPLY_TO_POSE>
</APPLY_TO_POSE>
<PROTOCOLS>
    Add mover=dock_low/>
    Add mover=srsc />
    Add mover=dock_high />
    <Add mover="minimize_interface" />
</PROTOCOLS>
<OUTPUT scorefxn="REF2015" />
</ROSETTASCRIPTS>

```

10.5.7.4.1 Input files for full and minimized protein-protein docking

- PDB input file with four subunits belonging to the interface of two container
- Docking_options

```

-docking
  -partners AB_GH
  -dock_pert 3 8
  -dock_mcm_trans_magnitude 0.1
  -dock_mcm_rot_magnitude 5.0
-s QC12-D1_0000_renumb.pdb
-use_input_sc
-ex1
-ex2
-out
  -file
    -scorefile docking.fasc
-score:weights ref2015.wts

```

10.5.7.5 Script to start protein-protein docking analysis script

```
#!/bin/bash
#SBATCH --job-name=DockingAnalysis
#SBATCH --partition=std
#SBATCH --nodes=8
#SBATCH --tasks-per-node=16
#SBATCH --time=12:00:00
#SBATCH --export=NONE

set -e
source /sw/batch/init.sh
module switch env env/gcc-8.2.0_openmpi-3.1.3
module list
export LD_LIBRARY_PATH="$LD_RUN_PATH"

export
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/work/fcvx142/Rosetta_3.13/rosetta_src_2021.16.61629_bundle/main/source/build/src/release/linux/4.14/64/x86/gcc/8.2/mpi/

mpirun
/work/fcvx142/Rosetta_3.13/rosetta_src_2021.16.61629_bundle/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@docking_analysis.options -in:file:s *full*pdb *minimize*pdb >&
docking_analysis.log
sort -nk 7 docking_analysis.csv
pymol
QC12-D1_native_renumb.pdb QC12-D1_0000_renumb_full_0000.pdb
```

10.5.7.6 Script for protein-protein docking analysis

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
  </SCOREFXNS>
  <TASKOPERATIONS>
    <RestrictToInterfaceVector name="rtiv" chain1_num="1,2" chain2_num="3,4"
CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75"
vector_dist_cutoff="9.0" />
  </TASKOPERATIONS>
  <MOVERS>
    <InterfaceAnalyzerMover name="iface_analyzer" scorefxn="REF2015" packstat="0"
pack_input="0" pack_separated="1" fixedchains="A,B" tracer="0" />
  </MOVERS>
  <FILTERS>
    <AverageDegree name="avg_degree" threshold="0" distance_threshold="10"
task_operations="rtiv" />
    <Rmsd name="rmsd" superimpose="1" threshold="2500" >
      <span begin_res_num="9" end_res_num="41" />
      <span begin_res_num="75" end_res_num="120" />
      <span begin_res_num="181" end_res_num="213" />
      <span begin_res_num="247" end_res_num="292" />
      <span begin_res_num="353" end_res_num="385" />
      <span begin_res_num="419" end_res_num="464" />
      <span begin_res_num="525" end_res_num="557" />
      <span begin_res_num="591" end_res_num="636" />
```

```

        </Rmsd>
    </FILTERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        <Add mover="iface_analyzer" />
        <Add filter="avg_degree" />
        <Add filter="rmsd" />
    </PROTOCOLS>
    <OUTPUT scorefxn="REF2015" />
</ROSETTASCRIPTS>

```

10.5.7.6.1 Input files

- **Docking_analysis_options**

```

-parser:protocol docking_analysis.xml
-use_input_sc
-ex1
-ex2
-out:file:scorefile docking_analysis.csv
-native QC12-D1_native_renumb.pdb
-out:no_nstruct_label
-out:file:score_only
-score:weights ref2015.wts

```

10.5.8 Script for identifying the influence on single amino acids and the overall dG_separated

```

#!/usr/bin/env python
# coding: utf-8

import os
import pandas as pd
import numpy as np
from statistics import mean

notebook_path = os.path.abspath("Bioinfo.ipynb")
notebook_path
docking_results = os.path.join(os.path.dirname(notebook_path),
"Dock_03_results.txt")
df = pd.read_csv('Dock_03_results.txt', sep= ',', header=None,
skiprows=1)
#df.head()

length = len(df.index)      # sets the length of the dataframe this
is attributed to the amount of structures which passed the
filters in the design
for i in range(length):    # iterate through each line of the
dataframe
    sequence=df.iloc[i,1]  # define sequence as thing in data
frame at row 1 and line i
    sequence=list(sequence) # separate sequencee in its singel
character with teh list command

```

```
dG = df.iloc[i,2]          # define the dG values similar to
sequence
name = df.iloc[i,0]       # define name of the fastas
sequence.append(dG)       # append the dG value to the sequence
sequence.append(name)     # append the name value to the
sequence
sequence                  # now we have a list with each AA
as a singel character. The last two entries are the dG value and
the name of the coressponding pdb
lst.append(sequence)      # append the list o another list
and create a list of lists over all construct in the input file

#print(lst)

positions = list(map(list, zip(*lst))) # seperate the list after
the indices of the sub list. As a result you get a list with 174
entries where the last two are the name and the dG value

len(positions)
positions[1] # This and all entries till 171 contain the list
of the AA off all structures
#positions[172] # This line contains the dG seperate value
#positions[173] # This line contains the name of the pdb

# Now we have seperated our input data in a List of List. Each
sublist of this list represents all different AA in all different
Fasta files from the input, that are at the same position in the
fasta. Addional we have a list with the dG seperated. Since the
indxes are equal, the AA at postion x can be attributed to the
dG at position x. This data structure allows to write a code
whcih loops though the singel list and calculate the mean value
for each amino acid. This is than stored in a new dataframe and
finally polttet in a heat map

All_dG = positions[172]
mean_dG = mean(All_dG)
mean_dG

All_AA = "ARNDBCEQZGHILKMFPSTWYV" # List of All AA
All_AA = list(All_AA) # Separates
results = pd.DataFrame() # Create Dataframe for results
for i in range(173): # This Loops lopps around the different
Lists in our List of List. Since each list in the List of list
attributed to all AA in all fastas. This is equal to the Positions
in the later protein
    pos = positions[i] # here a list from the list of list
"positions" is opend and stored in a variable pos. Pos than
contains the list of AA at the given position for all pdbs
    position = i+5 # Python numbering start at 0 our PDB
at 5 so i need to add 5 to the loop variable i
    List_dG = [] # in this list all the dG values are
stored. It is refreshed for every new position
    for idx, val in enumerate(pos): # not so sure what happens
here in detail. copied an paste it from stack overflow. I guess
```

```
it iterates over the whole list and stores the AA in the variable
val
    for AA in All_AA:                # iterates throug the list
of all AA made prior
    AA_of_interest = AA              # sets AA to AA of interest
    if AA_of_interest == val:        # if AA of interest ist
equal to AA do following ..
        dG = positions[172][idx]     # search dG in the
positions list at line 172 where all dGs are stored
        List_dG.append(dG)          # added all dG for the
same AA to the dG list
        mean_dG_AA = mean(List_dG)  # calcualte the mean
value of this list
        results.loc[AA_of_interest, position] =
mean_dG_AA # add the values to an array
```

```
# Import packages needed to plot
import seaborn as sns
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(100,15));
sns.heatmap(results,cmap = 'Blues_r', square = True,
annot=True,linewidths=.1);
```

```
results_cut1 = results.iloc[:, 60:81]
plt.figure(figsize=(100,15));
sns.heatmap(results_cut1,cmap = 'Blues_r', square = True,
annot=True,linewidths=.1);
```

```
results_cut2 = results.iloc[:, 119:135]
plt.figure(figsize=(100,15))
sns.heatmap(results_cut2,cmap = 'Blues_r', square = True,
annot=True,linewidths=.1)
```


11 List of abbreviations

AaLS	<i>Aquifex aeolicus</i> lumazine synthase
Bfr	Bacterioferritin
C10	2-Bromo-N-decylacetamide
CCMV	Cowpea chlorotic mottle virus
CKD	Chronic kidney diseases
Da	Dalton
DOF	Degrees of freedom
Dps	DNA-binding proteins from starved cells
DTT	Dithiothreitol
EtOH	Ethanol
ESI-MS	Electron spray ionization mass spectrometry
Fixbb	Fixed backbone
FPAD	Fractionated plasma separation adsorption and dialysis
Ftn	Ferritin
Ftn ^(neg)	Negative-supercharged ferritin
Ftn ^(pos)	Positive-supercharged ferritin
GFR	Globular filtration rate
Gua	Guanidine hydrochloride
HAS	Human serum albumin
IEC	Ion-exchange chromatography
IPTG	Isopropyl- β -d-thiogalactopyranosid
IS	Indoxyl sulfate
LPS	Lipopolysaccharides
MgOAc	Magnesium acetate
MOF	Metal-organic framework
MPI	Message passing interface
NHS	N-Hydroxysuccinimide
PAA	Phenylacetic acid
<i>p</i> -CS	<i>p</i> -Cresyl sulfate
Phe	2-Iodo-N-phenylacetamide
RRT	Renal replacement therapy
SASA	Surface accessible surface area
SCXRD	Single-crystal X-ray diffraction
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

List of abbreviations

SEC	Size exclusion chromatography
SNARF	Seminaphtharhodafluor
Sulfo-SMCC	Sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate
TEM	Transmission electron microscopy
TMV	Tobacco mosaic virus
Tris	Tris(hydroxymethyl)aminomethane
TCEP	Tris(2-carboxyethyl)phosphine
UV/Vis	Ultraviolet visible

12 List of chemicals

Chemical	Supplier	H and P statement
Acetic acid	VWR	GHS02; GHS05 H: 226-314 P: 210-280-301+330+331-303+361+353-305+351+338
Ammonium sulfate	AppliChem	GHS07; GHS09 H: 302-315-319-335-411 P: 261-264-270-271-273-280-301+312-302+352-304+340-305+351+338-312-321-330-332-313-337-313-362-391-403+233-404-501
Ampicillin sodium salt	AppliChem	GHS07; GHS08 H: 317-334 P: 261-272-280-284-321-302+352-304+341-333+313-342+311-362+364-501
Brilliant Blue	Roth	H: 412 P: 273
2-Bromo-N-decyl acetamide	Merck	GHS07 H: 315-319-335-413 P: 273-302+352-305+351+338
Dithiothreitol	Sigma Aldrich	GHS07 H:302-315-319-335 P: 280-302+352-305+351+338-308+311

List of figures

Ethanol	Fischer Scientific	GHS02; GHS05 H: 225-319 P: 210-240-305+351+338-403+233
Glutaraldehyde (25%)	Merck	GHS05; GHS06; GHS08; GHS09 H:301-330-314-317-334-335-410 P: 260-280-304+340-310-305+351+338-403+233
Glycerol	VWR	-
Hydrochloric acid (37%)	VWR	GHS05; GHS07 H: 290-314-335 P: 260-280-303+361+353-304+340+310-305+351+338
Isopropyl- β -d-thiogalactopyranosid (IPTG)	Roth	GHS07 H: 319-335-315 P: 280-302+352-304+340-305+351+338-312
2-Iodo-N-phenylacetamide	Abcr	GHS05; GHS07 H: 302-315-319-335 P: 280-261-304-340-312-301-330-305-351-338
Sodium chlorid	Roth	-
Sodium hydroxid		GHS05 H: 290-314 P: 280-301+330+331-305+351+338-308+310
Tris(2-carboxyethyl) phosphine (TCEP)	Roth	GHS05 H: 314 P: 280-305+351+338-310

List of figures

Tris(hydroxymethyl)aminomethan (Tris)	Roth	-
Tryptone	AppliChem	-
Uranyl acetate		GHS06, GHS08, GHS10 H: 300+330-373-411 P:260-264-270-271-273-284-301+310-304+340-310- 314-320-321-330-391-403+233-405-501
Yeast	AppliChem	-

13 List of figures

Figure 2.1: Structure of the protein-bound uremic toxins.	3
Figure 2.2 : Dimensions and shapes of different protein cages.....	7
Figure 2.3: Schematic visualization of the three distinct surfaces of a protein cage.	8
Figure 2.4: Structure of the ferritin cage.....	12
Figure 2.5: Bioconjugation with an α -halocarbonyl agent.....	14
Figure 2.6: Mechanism of thiol-maleimide click reaction.	15
Figure 2.7: Reaction scheme for active ester with primary or secondary amines.	16
Figure 2.8: Glutaraldehyde structures and possible reaction pathways.	17
Figure 2.9: Main elements in a Rosetta protocol.....	20
Figure 4.1: Schematic overview of the strategy for protein functionalization.	29
Figure 5.1: Schematic representation of the cysteine anchor sites.	32
Figure 5.2: SEC chromatogram and SDS-PAGE for Ftn ^(neg) -Cys variants.	33
Figure 5.3: Chemical structure of hydrophobic ligands.	35
Figure 5.4: Functionalization of Ftn ^(neg) -1xCys with Phe-ligand.	36
Figure 5.5: ESI-MS spectra of functionalized Ftn ^(neg) -3xCys variants.	37
Figure 5.6: SEC chromatogram comparing the elution volume of functionalized and non-functionalized Ftn ^(neg) cysteine variants.....	39
Figure 5.7: Negatively stained TEM images of functionalized and non-functionalized Ftn ^(neg) cysteine variants.....	39
Figure 5.8: Crystals of functionalized Ftn ^(neg) -Cys variants.	41
Figure 5.9: Electron density omit maps for Ftn ^(neg) -3xC10.	42
Figure 5.10: Phase diagram for crystallization of macromolecules.....	43
Figure 5.11: Crystal size at different precipitant and protein concentrations.	44
Figure 5.12: Protein crystals of Ftn ^(neg) -4xPhe.....	45
Figure 5.13: Chemical structure of applied crosslinker.....	46
Figure 5.14: Stability of Ftn ^(neg) crystals after different cross-linking procedures.....	47
Figure 5.15: Non-crystalline protein material.	48

List of figures

Figure 5.16: Endotoxin assay for the crystalline adsorbent.	50
Figure 5.17: Platelet activation assay for crystalline protein-based adsorbent.	52
Figure 5.18: Biocompatibility of non-crystalline ferritin material.	52
Figure 5.19: BSA adsorption assay.	53
Figure 5.20: Decreasing IS control concentrations.	56
Figure 5.21: Results from toxin assays.	57
Figure 5.22: Adsorption capacity of crystalline material from chemical functionalized ferritin variants.	59
Figure 5.23: Comparison of adsorption capacity for the crystalline and non-crystalline adsorbents.	61
Figure 5.24: Fluorophore-loaded ferritin variants.	63
Figure 5.25: Ftn ^(neg) -SNARF pH sensing capability.	64
Figure 5.26: Sequence alignment for Ftn ^(neg) -Ap variants.	66
Figure 5.27: Structure and surface potential of Ftn ^(neg) mutations on the inner surface.	67
Figure 5.28: Variations in toxin binding affinity in amino acid composition during modeling generations.	70
Figure 5.29: Positions of ligands, mutations and surface potential of the designed Ftn ^(neg) -docking variants.	72
Figure 5.30: Chemical environment of toxins in docking models.	73
Figure 5.31: Ftn ^(neg) with all binding sites combined.	75
Figure 5.32: Design of a binding pocket.	76
Figure 5.33: 3-fold channel redesign.	78
Figure 5.34: Ferritin model with removed ferroxidase site and iron transport pathways.	79
Figure 5.35: SEC chromatogram, SDS-PAGE and TEM images for Ftn ^(neg) -Ap variants.	81
Figure 5.36: Crystals of Ftn ^(neg) variants with increased density of hydrophobic amino acids.	83
Figure 5.37: SEC and SDS-PAGE of Ftn ^(neg) -docking variants.	85
Figure 5.38: Crystals of Ftn ^(neg) -dock variants.	86
Figure 5.39: Expression and purification of Ftn ^(neg) with toxin binding pocket.	87

List of figures

Figure 5.40: Characterization of ferritin with expanded pores.	88
Figure 5.41: Ftn ^(neg) -Ap4-3A crystals.	89
Figure 5.42: Batch crystallization of Ftn ^(neg) -Ap variants.	90
Figure 5.43: Biocompatibility assay for redesigned Ftn ^(neg) -dock43.	92
Figure 5.44: Adsorption capacity for crystalline and non-crystalline materials based on Ftn ^(neg) -Ap variants.	95
Figure 5.45: Adsorption capacity of Ftn ^(neg) -Ap4 variant with expanded pore.	97
Figure 5.46: Adsorption capacity of crystalline material based on Ftn ^(neg) variants with toxin binding sites.	98
Figure 5.47: Comparison of adsorption capacity of crystalline and non-crystalline material based on Ftn ^(neg) variants with toxin binding sites.	100
Figure 5.48: Recyclability of protein-based adsorbents.	102
Figure 5.49: Protein-protein docking for interface A1 of Ftn ^(neg) and Ftn ^(pos)	105
Figure 5.50: Interface energy in binary protein crystal.	106
Figure 5.51: Interface energy in binary protein crystal.	107
Figure 10.1: Distances between reactive cysteine sites.	154
Figure 10.2: Ion-exchange chromatography (IEC) of Ftn ^(neg) -Cysteine variants.	155
Figure 10.3: ESI-MS spectra of Ftn ^(neg) -1xCys and 2xCys.	156
Figure 10.4: Size exclusion chromatogram of Ftn ^(neg) -3xCys after Phe coupling in presence of TCEP.	157
Figure 10.5: ESI-MS spectra of Ftn ^(neg) -4xCys functionalized with Phe ligand.	158
Figure 10.6: Ftn ^(neg) -4xCys functionalization with 20 eq. C10 ligand at 50% EtOH in reaction mixture.	159
Figure 10.7: Ftn ^(neg) -4xCys functionalization with 20 eq. C10 ligand at 80% EtOH in reaction mixture.	159
Figure 10.8: Ftn ^(neg) -4xCys functionalization with 40 eq. C10 ligand at 80% EtOH in reaction mixture.	160
Figure 10.9: ESI-MS spectrum of Ftn ^(neg) -3xCys variants.	161
Figure 10.10 ESI-MS spectrum of Ftn ^(neg) -4xCys variants.	162
Figure 10.11: Size of solvent channel.	165

List of figures

Figure 10.12: Standard deviation for PBUT measurement.....	166
Figure 10.13: Reduction of toxin concentration for different crosslinking techniques.	167
Figure 10.14: Possible ligands for functionalization of Ftn ^(neg) -Cys.....	168
Figure 10.15: Binding energy as determined by Rosetta for Ftn ^(neg) -Ap variants.	169
Figure 10.16: Influence of single amino acids on the overall dG _{separated}	170
Figure 10.17: IEC chromatogram of ferritin variants with increased hydrophobic side chain density.....	171
Figure 10.18: Surface potential of ferritin variants with increased density of hydrophobic residues.....	172
Figure 10.19: IEC chromatogram for Ftn ^(neg) -docking variants.	173
Figure 10.20: ESI-MS spectra for Ftn ^(neg) -Ap variants.	175
Figure 10.21: ESI-MS spectra for Ftn ^(neg) -docking variants.	176
Figure 10.22: Electron density omit maps for Ftn ^(neg) -Ap4.....	179
Figure 10.23: Electron density omit maps for Ftn ^(neg) -Ap7.....	180
Figure 10.24: Electron density omit maps for Ftn ^(neg) -Ap16.....	181
Figure 10.25: Electron density omit maps for Ftn ^(neg) -dock03.....	184
Figure 10.26: Electron density omit maps for Ftn ^(neg) -dock23.....	185
Figure 10.27: Electron density omit maps for Ftn ^(neg) -dock43.....	186
Figure 10.28: Improved binding site of Ftn ^(neg) -dock43.....	188

14 List of tables

Table 2.1: Renal replacement therapies.	5
Table 2.2: Terms in REF15 energy function	24
Table 5.1: Theoretical and measured molecular mass of Ftn ^(neg) -Cys variants.....	34
Table 5.2: Theoretical and measured molecular mass for Ftn ^(neg) -3xCys derivates.....	40
Table 5.3: Theoretical and measured molecular mass for Ftn ^(neg) -4xCys derivates and their diameter according to DLS measurements.....	40
Table 5.4 Rosetta score of Ftn ^(neg) variants with additional hydrophobic residues.....	66
Table 5.5: Molecular mass of Ftn ^(neg) -Ap variants.....	82
Table 5.6: Molecular mass of Ftn ^(neg) -docking variants.....	85
Table 8.1. MRM parameters	120
Table 8.2: Sequence of HPLC program.....	120
Table 8.3. Sequence of primers used for QuikChange™ PCR protocol.....	122
Table 10.1: Solvent accessible surface area of amino acids at Ftn ^(neg) inner surface.....	153
Table 10.2: Solvent-accessible surface area.....	154
Table 10.3: Conductivity of buffer at protein elution in IEC.....	156
Table 10.4: Comparisons of UV absorption of Phe functionalized Ftn ^(neg) cysteine variants.....	157
Table 10.5: Summary of crystallographic data.....	163
Table 10.6: Removed and added negative charged amino acids in redesigned Ftn ^(neg) variants.....	173
Table 10.7: Overview variants for protein-protein docking.....	177
Table 10.8: Summary of crystallographic data.....	177
Table 10.9: Summary of crystallographic data.....	182
Table 10.10. Highest capacities for respective toxin	187

Acknowledgment – Danksagung

Zuerst möchte ich mich bei meinem Doktorvater Prof. Tobias Beck für die Möglichkeit an diesem aufregenden Projekt zu arbeiten bedanken. Besonders hervorheben möchte ich dabei die zahlreichen Diskussionen, wissenschaftlichen Freiheiten und großartige Unterstützung, ohne die die vorliegende Dissertation nicht hätte entstehen können.

Prof. Wolfgang Parak danke ich für die freundliche Übernahme des Gutachtes für die Dissertation und Prof. Torda und Dr. Riedner als Mitglieder der Prüfungskommission.

Prof. Joachim Jankowski und Dr. Setareh Orth-Alampour vom Uniklinikum der RWTH Aachen möchte ich für die freundliche Kooperation, die wissenschaftlichen Diskussionen und die Durchführung der Biokompatibilitätsbestimmung danken.

Bei den ehemaligen Mitgliedern unserer Arbeitsgruppe Dr. Matthias Künzle, Dr. Marcel Lach und Dr. Made Budiarta möchte ich für die exzellente Einarbeitungen in meinen frühen Jahren als Doktorand bedanken. Weiterhin möchte ich mich bei allen aktuellen und ehemaligen Mitgliedern der Arbeitsgruppe Michael Rütten, Dr. Brandon Seychell, Laurin Lang, Varnika Yadav und Dr. Rafiga Masmaliyeva für die freundschaftliche Arbeitsatmosphäre, die großartige Unterstützung und die zahlreichen wissenschaftlichen Diskussionen bedanken. Außerdem möchte ich mich bei meinen Forschungsstudenten Konstantin Sopeniak, Maria Brinn und Henrike Wagler, die mit ihren Arbeiten direkt oder indirekt zum Gelingen dieser Dissertation beigetragen haben, bedanken.

Mein Dank geht an Dr. Maria Riedner und ihren Mitarbeitern Gaby Graack und Dany Gellert für die ESI-MS Messungen sowie die Hilfe und Einarbeitung an dem HPLC-MS System, das für die Toxin Bestimmung benötigt wurde. Weiterhin möchte ich mich bei Stefan Werner für die elektronenmikroskopischen Untersuchungen bedanken. Auch bei Dr. Thomas Orgis und Dr. Hinnerk Stüben möchte ich für den großartigen Support im Rahmen der Arbeiten auf dem HPC-cluster Hummel bedanken.

Weiterhin möchte ich mich bei Michael Rütten, Christoph Schwarz, Dr. Rafiga Masmaliyeva und Varnika Yadav für das Korrekturlesen dieser Dissertation bedanken.

Abschließend möchte ich mich besonders bei meinen Eltern und meiner Lebensgefährtin Jasmin Schneider bedanken, dir mir in allen guten und schweren Phasen meines Lebens mit Unterstützung, Verständnis und Hilfe zur Seite standen und stehen.

Eidesstattliche Erklärung:

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

13.08.2022

Hendrik Böhler

Hendrik Böhler

Curriculum Vitae

Personal

Name:	Hendrik Böhler
Date of Birth:	12.12.1991
Place of Birth:	Gelsenkirchen
Nationality:	German

Education

01/2020 – 12/2022	Doctoral studies Institute Physical Chemistry University Hamburg
01/2019 – 12/2019	Doctoral studies Institute Inorganic Chemistry RWTH Aachen
10/2016 – 10/2018	Master of Science in Chemistry Universität Duisburg-Essen
10/2011 – 09/2016	Bachelor of Science in Chemistry Universität Duisburg-Essen
06/2002 – 07/2011	Allgemeine Hochschulreife Anette von Droste-Hülshoff Gymnasium