Supramolecular control over protein assembly

Citation for published version (APA):

DOI:
10.6100/IR715593

Document status and date:
Published: 01/01/2011

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Supramolecular control over protein assembly

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus, prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op maandag 19 september 2011 om 16.00 uur

door

Dana Anniela Uhlenheuer

geboren te Dortmund, Duitsland
Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. L. Brunsveld

Cover Design: D. A. Uhlenheuer, K. Petkau
Printing: Wöhrmann Print Service, Zutphen

A catalogue record is available from the Eindhoven University of Technology Library
# Table of contents

## Chapter 1

**Supramolecular chemical biology – controlling proteins with supramolecular chemistry**

1. Introduction ................................................................................................................. 2
2. Supramolecular chemistry for biology in water ......................................................... 5
3. Controlling proteins with host-guest chemistry ......................................................... 7
   3.1 Protein surface recognition ...................................................................................... 8
   3.2 Controlling protein assembly in solution with synthetic elements ..... 11
   3.3 Supramolecular protein immobilization ................................................................. 15
   3.4 Conclusion .............................................................................................................. 17
4. Aim and outline of this thesis ..................................................................................... 18
5. References .................................................................................................................. 19

## Chapter 2

**Modulation of protein dimerization by a supramolecular host-guest system**

1. Introduction ................................................................................................................ 24
2. Protein design and synthesis ....................................................................................... 26
3. Supramolecular induced protein dimerization ............................................................ 28
4. Selective covalent locking of supramolecularly induced protein heterodimers ............... 36
5. Conclusions ................................................................................................................ 40
6. Experimental part ...................................................................................................... 42
7. References .................................................................................................................. 45

## Chapter 3

**An improved host-guest system for supramolecular protein dimerization**

1. Introduction ................................................................................................................ 48
2. Synthesis of a mono-functionalized cysteamine-cyclodextrin ...................................... 50
3. Ligation of cysteamine-cyclodextrin to mYFP ............................................................. 53
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>CB[8] induced heterodimerization of methylviologen and naphthalene functionalized proteins</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Design and synthesis</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>CB[8]-induced protein heterodimerization</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Conclusions and outlook</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Experimental part</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>Labeling of SNAP-tag fusion proteins with supramolecular ligands</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Synthesis of benzylguanine conjugates</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>2.1 Synthesis of reactive benzylguanines</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>2.2 Conjugation of lithocholic acid</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>2.3 Conjugation to β-cyclodextrin</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>2.4 Conjugation to bipyridine-discotic</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Labeling of AGT- fusion proteins with supramolecular elements</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>3.1 Purification of AGT fusion proteins</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>3.2 Labeling with host-guest molecules</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>3.3 Labeling with a supramolecular polymer</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Conclusions and outlook</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Experimental part</td>
<td>100</td>
</tr>
</tbody>
</table>
Chapter 6 .................................................................................................................. 113

Supramolecular immobilization of SNAP-tag fusion proteins ................. 113

1 Introduction ............................................................................................................ 114

2 Synthesis of modified proteins ......................................................................... 116
   2.1 Synthesis of adamantane and ferrocene BG-conjugates ......................... 116
   2.2 Ligation to proteins ...................................................................................... 118

3 Immobilization of adamantane proteins on CD-surfaces ......................... 119
   3.1 Surface plasmon resonance measurements ............................................ 119
   3.2 Protein patterns .......................................................................................... 121

4 Interaction of adamantane proteins with CD-vesicles ............................ 122

5 Immobilization of ferrocene proteins on CB[7] surfaces .......................... 124

6 Conclusions and outlook .................................................................................... 126

7 Experimental part ............................................................................................... 127

8 References ........................................................................................................... 134

Summary .................................................................................................................. 137

Curriculum Vitae .................................................................................................... 141

Acknowledgements ................................................................................................ 143
Chapter 1

Supramolecular chemical biology – controlling proteins with supramolecular chemistry

Supramolecular chemistry has primarily found its inspiration in biological molecules, such as proteins and lipids, and their interactions. Currently the supramolecular assembly of designed compounds can be controlled to a great extent. This provides the opportunity to combine these synthetic supramolecular elements with biomolecules for the study of biological phenomena. In this chapter requirements for the application of supramolecular elements to proteins are discussed using examples from the recent literature. Applications of supramolecular structures for the inhibition of protein-protein interactions and the use of host-guest chemistry for controlled protein assembly are described. The combination of bionanotechnology with synthetic supramolecular chemistry, for example the immobilization of proteins on surfaces, is briefly discussed as well. This framework of supramolecular chemistry applied to biology, directed the formulation of this thesis ‘supramolecular control over protein assembly’.

The main part of this chapter has been published: D. A. Uhlenheuer, K. Petkau, L. Brunsveld, Combining supramolecular chemistry with biology, Chem. Soc. Rev. 2010, 39, 2817-2826.
Chapter 1

1 Introduction

Supramolecular chemistry is the study of non-covalent interactions in and between molecules, and the resulting multimolecular complexes. Supramolecular chemistry is referred to as the chemistry that goes beyond the covalent bond, beyond the individual molecule\(^1\). Natural molecules, such as proteins, oligonucleotides, lipids and their multimolecular complexes, have been the major source of inspiration for supramolecular chemists. Design and synthesis of novel multimolecular supramolecular architectures of similar complexity and functionality is the dream of many supramolecular scientists. Evidently, synthetic supramolecular systems were initially rather small and composed of relatively simple building blocks\(^2\). However, the increased knowledge over intermolecular interactions and molecular recognition has culminated in an impressive control over molecular self-assembly\(^3\). Supramolecular synthesis of multimolecular architectures with diverse shapes, compositions, and functionalities is now possible in a wide-range of conditions in solution and the solid state.

Even though the initial inspiration for supramolecular chemistry came from biology, most of the current applications of supramolecular chemistry can be found in the materials sciences. Supramolecular concepts have enabled unprecedented control over materials organization and properties. This has yielded for example supramolecular switches, electronics, and polymers\(^4\). Remarkably, most of these supramolecular materials display their specific properties in the solid or gel state or in organic solvents. This is in contrast with the biological systems from which their design initially was derived. Proteins, lipids, and sugars assemble and unfold their specific properties in water. The application of supramolecular chemistry to study biology, supramolecular chemical biology, thus requires synthetic systems that assemble in water or biological media.

The physical-chemical characteristics of water are completely different from those of organic solvents. In water, isolated polar secondary interactions, such as hydrogen bonds, are interfered with the solvent and require a shielding hydrophobic environment, such as in the interior of a protein, to become fully functional. However, hydrogen bonds do not require shielding to enhance the properties of synthetic polymers in organic solvents or in the solid state\(^4\). Water thus imposes specific design criteria on the application of polar secondary interactions in supramolecular systems\(^5\). Many different supramolecular architectures that
assemble under dilute conditions in water or buffered media have been designed and synthesised in recent years. As a result, supramolecular chemistry has become an attractive approach to address self-assembly in biological systems and has returned to its alma mater, biology.

Interactions between molecules in biological systems are supramolecular interactions. The localization, interactions, and functions of biomolecules such as proteins and lipids are modulated in an environment where supramolecular assembly determines the biological processes. Gaining control over these supramolecular interactions is the key for understanding and targeting (misregulated) biological processes in diseases. The most successful approach to obtain control over these processes is the modulation of specific biomolecules with small molecules. So-called target-based drug discovery is in essence supramolecular chemistry. Drugs bind to proteins via supramolecular host-guest interactions and therewith influence the functioning or conformation of the protein.

Apart from individual host-guest interactions, as addressed in medicinal chemistry, biological systems also feature more complex supramolecular interactions, for example when cells interact. Such recognition events are characterized by multiple interactions acting simultaneously, the so-called phenomenon of polyvalency. Synthetic polymers displaying multiple bio-active ligands have found applications as synthetic scaffolds for inhibition of biological polyvalency events. Small molecule drugs and polyvalent polymers have shown the power of designing novel synthetic systems for studying and manipulating biology. Nevertheless, when considering the size and well-defined superstructure of biological systems such as protein complexes and cell membranes, synthetic systems featuring equal complexity are still missing. Both for the small molecules and for the polymers a significant gap exists between the synthetic compounds on the one hand and the biological targets on the other hand. The synthetic compounds used to date lack the specific supramolecular characteristics, such as adaptivity to the environment and formation of well-defined superstructures, of the biological molecules they are targeting. Self-assembling synthetic molecules, supramolecular architectures, could bridge this gap (Figure 1).
Figure 1: Bridging the gap. Synthetic supramolecular architectures feature sizes, environment adaptivity, and higher order superstructures analogous to biological systems such as proteins and membranes and therefore could bridge the molecular gap between synthetic molecules and polymers on the one hand and biological structures on the other hand.

Until recently, well-defined synthetic supramolecular systems have found little application for the investigation and modulation of biological systems. This is highly intriguing, as synthetic supramolecular systems are capable of self-organization into multi-molecular assemblies that resemble, in both size and function, biomolecules and assemblies such as proteins and membranes\textsuperscript{8,9}. Supramolecular assemblies and biomolecules feature similar characteristics, such as dynamics, reversibility, topology, and polyvalency. The combination of supramolecular chemistry with biology thus offers a wealth of new possibilities to study and influence biological processes. In this introduction examples from recent literature will be discussed which illustrate this principle. First the design criteria for the application of supramolecular molecules to biological systems in aqueous media will be discussed. Subsequently, the combination of synthetic supramolecular systems with biology is discussed. Host-guest systems are highlighted as discrete, well-defined interactions to control protein folding, dimerization, and immobilisation (Figure 2).
2 Supramolecular chemistry for biology in water

The concepts of molecular recognition and self organization in synthetic supramolecular architectures rely on inspiration from natural systems. Concomitantly, the design, synthesis, and study of supramolecular assemblies in water are intriguing goals\(^5,10\). Synthetic systems featuring similar levels of ingenuity and complexity as encountered in natural systems are a far-standing aim. Nevertheless, significant progress has been made in the field of supramolecular chemistry in water starting from the earlier highly innovative work on micelles and vesicles\(^11\) to recently developed receptors and self-assembling systems\(^5,10\). The main driving force for self-assembly in water of most supramolecular architectures, either biological or synthetic, is the hydrophobic effect. Additionally, strengthening and directing polar interactions such as hydrogen bonding, ion-ion, and ion-dipole interactions can be applied. It is important to note that self-assembly processes in aqueous solutions depend on the concentration and type of salts\(^12\). Salt effects are not only important for self-assembly systems based on ion-ion interactions, but also of influence on self-assembly processes driven by hydrophobic interactions depending on the type of salt\(^13\). Most of the supramolecular systems applied to biological targets feature a strong hydrophobic assembly component, combined with structuring polar interactions. Overall, for the application of synthetic supramolecular chemistry to biology three important requirements can be defined:

A) **Supramolecular systems have to assemble in buffered biological media with interaction strengths in the regime of biomolecules;**

Supramolecular interactions in biological systems typically occur in the \(\mu\)M to \(p\)M regime. Analogously, synthetic supramolecular building blocks should feature similar

![Conceptual application of synthetic supramolecular chemistry to biology. Discrete small host-guest assemblies allow recognizing and modulating interactions between specific biomolecules.](image)
interaction strengths in biological media. In the end, the actual required affinity is
determined by the biological interaction that is to be studied and depends for
example on the concentration of the biomolecules to be targeted and the
applicability of polyvalency. In case a supramolecular assembly displays multiple
epitopes, the size and epitope density of the supramolecular aggregate will
determine affinity.

B) Supramolecular systems have to assemble in a bio-orthogonal and selective
manner;

Unselective interactions of synthetic supramolecular elements with biological matter
should be avoided. Assembly processes based purely on a single type of interaction,
such as ionic or hydrophobic interactions, will feature high tendency for unselective
interactions with biological matter. In order to achieve both a strong supramolecular
assembly and sufficient selectivity, supramolecular recognition motifs are preferably
based on two or more different secondary interactions. Typically, a combination of
hydrophobic interactions, accounting for sufficient assembly affinity, with polar
structuring elements such as ion-dipole or hydrogen bonding interactions is required.

C) Supramolecular systems have to feature the possibility for bioconjugation;

Conjugation of biological ligands to synthetic supramolecular elements is typically
required for applications to biological systems. The supramolecular building blocks
thus require molecular handles that can be modified with biomarkers. Additionally,
the conjugation of biological ligands to the supramolecular system should not disturb
the self-assembly process.

Supramolecular systems operative in water can roughly be divided into two types of
architectures, based on their structural characteristics. There are supramolecular
building blocks that aggregate in water in multi-molecular micellar or vesicular
assemblies\textsuperscript{8,11}. Environmental changes can typically result in changes in size and
composition of these assemblies. These multimolecular architectures feature great
potential as polyvalent scaffolds for the interaction with biological surfaces, like cells,
and will not be discussed here. Here, the focus is on small supramolecular host-guest
systems (Figure 2). These well-defined molecular systems provide ideal platforms for
the recognition and assembly of specific protein complexes.
3 Controlling proteins with host-guest chemistry

Of all biomolecules, proteins are probably the most captivating, especially from a molecular and structural perspective. Proteins are involved in virtually all biological processes and their supramolecular modulation, for example with small molecules, is the basis for most drugs. Typically, proteins work in concert via so-called protein-protein interactions. Controlling protein-protein interactions is currently one of the biggest challenges in the life sciences as this would allow for detailed investigations of their functioning and for new types of drugs. However, most proteins and their interactions are difficult to regulate with a small molecule. On top of that, for many proteins and protein-protein interactions structural information is absent and their regulation by post-translational modifications is frequently unknown. Also, many protein interactions are transient, not allowing clear cut on-off regulation. New and diverse molecular strategies are therefore required to selectively control proteins and their interactions. Supramolecular host-guest systems can constitute such a new molecular strategy for the selective and reversible control of proteins and their interactions.

Synthetic host-guest assemblies are typically well-defined in size and assembled from a limited number of molecular components. The assembly characteristics of host-guest systems, such as affinity, can typically be tuned via chemical modifications of the individual components. Examples of host-guest systems featuring high affinity, the possibility for bioconjugation and bioorthogonal assembly, in line with the three important design requirements formulated in the previous section, are given in Figure 3.

![Supramolecular Host-Guest Interactions and Recognition Motifs](image)

**Figure 3:** Supramolecular host-guest interactions and recognition motifs, which are functional in water and their respective association constants: a) and b) strong host-guest interactions of different synthetically attractive guest molecules with cucurbiturils of different sizes\textsuperscript{15,16}; c) strong binding via optimal hydrophobic recognition of lithocholic acid by \( \beta \)-cyclodextrin\textsuperscript{17}; d) inclusion of a specific peptide motif in a self-assembled supramolecular host\textsuperscript{18}. 

Cucurbiturils, for example, are donut-shaped, symmetric host molecules of different ring sizes. They are highly attractive supramolecular host systems for biological applications, because of their high affinity to a wide range of synthetic recognition motifs in water (Figure 3a-b). Cucurbit[7]uril typically recognizes hydrophobic elements with a quaternary amine with high affinity in the nano- to picomolar regime. The high affinity is caused by hydrophobic interactions combined with optimal fit, and the interaction of the quaternary ion with the electronegative carbonyl rim (Figure 3a)\(^{15}\). Cucurbit[8]uril is a somewhat larger donut-shaped family member that can host two guest molecules simultaneously (Figure 3b). Interestingly, cucurbit[8]uril can for example simultaneously bind two N-terminal peptide motifs\(^{16}\), which opens up the possibility to recognize and bind two proteins simultaneously. Cyclodextrins are sugar-based non-symmetrical cone-shaped host molecules that also recognize a variety of hydrophobic guests in water. The recognition of lithocholic acid by β-cyclodextrin (Figure 3c) arises around 1 μM concentrations in water and is accounted for by the optimal geometry of the steroid for the cyclodextrin cavity, providing a high selectivity over other steroid scaffolds\(^{17}\). The design and synthesis of larger self-assembled supramolecular host molecules opens up the possibility to engineer molecular receptors that recognize specific peptide motifs (Figure 3d) and possibly complete proteins\(^{18}\). Synthetic supramolecular host systems can also be designed that recognize specific amino acids, peptides, protein motifs or protein patches with significant affinity (Figure 4a)\(^{19}\). Such supramolecular protein element binders provide entries for controlling protein functioning and for inhibition of protein-protein interactions. Examples include the selective recognition of amino acids by donut-shaped molecules such as cyclodextrins\(^{20}\) and cucurbiturils\(^{21}\) and the recognition of specific protein elements by synthetic receptors\(^{22}\). Such small host-guest systems have found for example application for the purification of chemically modified proteins via affinity chromatography over an immobilized cyclodextrin matrix\(^{23}\). The recognition of specifically charged protein patches or cofactors by synthetic supramolecular receptors offers new concepts for the modulation and study of protein activity. The recognition of proteins by small synthetic receptors will be illustrated in the following with a few examples.

### 3.1 Protein surface recognition

Calix[n]arenes are a class of macrocyclic organic host molecules which can complex a large variety of different molecules depending on ring size and substitution. In their function as host molecules they have been used to complex biologically relevant
small molecules such as amino acids and small peptides. A different application uses calix[n]arenes as multivalent scaffolds. Functionalization with sugar ligands for example, yields glycalixarenes which bind to lectines with high affinities. Here the emphasis is on calixarenes which have been designed to bind to specific parts on protein surfaces. Mendoza et al. have used a calixarene scaffold decorated with cationic ligands to target the tetramerization domain of p53 and of mutant p53-R337H. The tumor suppressor protein p53 is a transcription factor that is involved in the control of the cell cycle and an important target protein in cancer therapy. Active p53 is a homotetramer consisting of four monomers with 393 residues each. The tetramerization domain consists of four chains which fold into two dimers that dimerize to form the tetramerization domain. In the mutant p53-R337H the saltbridge between Arg337 and Asp352 is lost which leads to destabilization of the tetramer. Based on their work with a tetraguanidinium oligomer as binder of the p53 tetramerization domain the group of Mendoza designed a calix[4]arene with four guanidinoethylresidues at the upper rim and hydrophobic loops at the lower rim (Figure 4a). The cationic residues on the upper rim interact with negatively charged glutamates on p53 tetramer and the hydrophobic loops fit into the hydrophobic pocket of the tetramer thereby stabilizing the mutated p53. DSC analysis proved that the calixarene depicted in Figure 4a stabilizes the mutant p53-R337H and ESI-MS showed the noncovalent formation of a 2 to 1 ligand/protein complex (Kd as estimated by NMR-experiments 10^-4 M).

The group of Hamilton also used calix[4]arenes to build receptors to target protein surfaces and inhibit protein-protein interactions. They started with a calix[4]arene carrying four negatively charged cyclopeptides GDGD arranged around the hydrophobic cavity of the calixarene. The peptide loops were introduced to interact with lysine residues on the protein surface. Activity tests showed that the synthetic receptor was indeed able to inhibit the activity of cytochrome c. A similar design but with a different cyclopeptide proved to be an efficient inhibitor of the interaction of the platelet derived growth factor PDGF and the tyrosine receptor kinase PDGFR. PDGF plays an important role in various cellular processes such as cell proliferation, apoptosis and angiogenesis. An antagonist of PDGF is therefore a potential antiangiogenic and anticancer agent. Based on structure activity relationship studies, the four cyclopeptides on the calixarene scaffold were replaced by simple acid functionalized isophthalate groups. These receptors were still effective PDGF antagonists. To gain deeper insight into the interaction of the synthetic receptor a library of differently substituted calixarenes was synthesized varying the number of
isophthalates, the linker and alkylation of the lower rim hydroxygroups. They found that three isophthalate groups were enough to keep high binding affinity ($K_d$(tris(isophthalate))= 21 nM) and activity in cellular assays. A similar approach starting from a calix[4]arene scaffold varying alkylation on the lower rim and changing substituents on the upper rim lead to a tetrabutoxy-calix[4]arene that showed activity against both human immunodeficiency virus (HIV) and hepatitis C virus (HCV). This shows that binding of a calixarene receptor to a protein can effectively inhibit protein function in a cellular assay.

Apart from calixarenes, also other scaffolds were studied as synthetic protein receptors. Hamilton and others have been using porphyrins for the surface recognition of proteins such as cytochrome c and VEGF. Porphyrins have a fourfold symmetry as the calix[4]arenes described before. The core can be functionalized with different ligands to give a family of different binders which can be screened for their affinity to the target. Substituted tetracarboxyphenyl porphyrins have been studied in the group of Hamilton as binders of cytochrome c taking advantage of the intrinsic fluorescence of the scaffold. The surface of cytochrome c consists of hydrophobic regions as well as an array of basic amino acids. Therefore the design of the best binder with a $K_d$ of 20 nM was based on hydrophobic and acidic modifications (Figure 4b). Using CD-spectroscopy they showed that a nonfluorescent Cu porphyrin dimer (Figure 4c) can induce the unfolding of cytochrome c under physiologically relevant conditions and facilitate enzymatic proteolysis. Later they found that catalytic amounts of porphyrin are enough to achieve accelerated, effective proteolytic

Figure 4: Supramolecular scaffolds that recognize proteins. a) recognition of negatively charged protein patches by a synthetic calix[4]arene ligand; b), c) porphyrin receptors that bind to the surface of cytochrome c with a high affinity and induce unfolding of the protein; d) Two Ru(II)tris(bipyridine) complexes that differ in the geometric arrangement of the ligands and therefore bind to cytochrome c with different affinity.
digestion\textsuperscript{35}. Substituted porphyrins are a suitable structure if the protein surface consists of hydrophobic and charged regions which is the case for cytochrome c but has also been used to disrupt protein-protein interaction. Hamilton et al. chose the interaction of the receptor tyrosine kinase KIT and its ligand, stem cell factor (SCF), as a target for a small library of modified meso-tetrakis (4-carboxyphenyl)porphyrin\textsuperscript{36}. The inhibitors were first screened in an ELISA assay using the extracellular KIT receptor and labelled SCF. The best inhibitors were carrying four negative groups which were essential for binding. These were further used in a cell based assay to investigate their ability to inhibit KIT phosphorylation. For one of the inhibitors no KIT phosphorylation was observed showing that this class of supramolecular scaffolds allows for designing efficient protein binding agents. Following up on their work on the tetraphenylporphyrins Hamilton and coworkers used selfassembling G-quadruplexes as recognition element for cytochrome c\textsuperscript{37} and developed this further as a fluorescence detection system for proteins\textsuperscript{38}. Other scaffolds have been used for the recognition of protein surfaces based on the recognition of hydrophobic and charged areas such as anthracene derived receptors\textsuperscript{39}, host-guest complexes\textsuperscript{40} and metal bipyridine complexes\textsuperscript{41}. Wilson and coworkers studied the selective recognition of cytochrome c by a fluorescent Ru(II)tris(bipyridine) complex. This scaffold allows for easy detection of binding using fluorescence quenching or changes in fluorescence anisotropy. They could show that their receptor had a higher affinity for cytochrome c ($K_d= 2$ nM) over lysozyme ($K_d= 270$ nM) with similar charge state and surface composition\textsuperscript{42}. Further work revealed that the recognition of cytochrome c depends on the geometric arrangement of the ligands around the metal core (Figure 4d) showing that the scaffolds are binding specifically to cytochrome c by interacting with surface exposed lysines\textsuperscript{43}. These examples demonstrate that synthetic supramolecular protein receptors might become a new study tool to inhibit and control protein-protein interactions. Important challenges that need to be addressed in this respect include, showing that these concepts can be transferred to different proteins with high selectivity in biologically relevant media.

### 3.2 Controlling protein assembly in solution with synthetic elements

In the following, the usage of two or more synthetic supramolecular elements for the control over protein assembly and function will be discussed. A beautiful first example deals with the assembly of two peptides for functional DNA binding, mediated by a supramolecular host-guest system\textsuperscript{44}. Recognition events between biomolecules are critically dependent on the proper structural arrangement of the interacting partners. Recognition of specific DNA elements by proteins frequently
occurs via protein dimerization processes. Natural DNA recognizing proteins first dimerize via noncovalent interactions. It is therefore important to generate synthetic systems that equally assemble via noncovalent interactions to allow the equilibrium of dimer formation to be a regulating tool for obtaining DNA specificity. Synthetic oligopeptides were therefore modified with either β-cyclodextrin or an adamantyl group to form a supramolecular heterodimer that recognizes DNA (Figure 5a). Only the complex of the two supramolecular functionalized peptides is capable of strongly binding the DNA and this recognition can be inhibited by disrupting the supramolecular interaction with either free β-cyclodextrin or adamantane, proving the necessity for a heterodimer\textsuperscript{44}. The cooperative formation of the peptide dimer and subsequent DNA recognition can be critically tuned by the interaction affinity of the applied host-guest system with typical $K_d$\textsuperscript{s} of $1-3\cdot10^{-6}$ M\textsuperscript{45}. The prestabilization of the peptide dimer leads to strong dimer-DNA affinities of up to $K_d = 7\cdot10^{-13}$ M. Application of the supramolecular peptide dimer additionally allows DNA recognition with high sequence selectivity and over narrower ranges of peptide concentrations than an analogous covalent peptide dimer\textsuperscript{46}. Oligomerisation of the peptides via incorporation of a host β-cyclodextrin at one end of the peptide and a guest adamantyl at the other end further enhances the sequence selective DNA binding. This is explained with the low affinity for nonspecific DNA sequences of the monomer peptide which is in equilibrium with an intramolecular inclusion complex\textsuperscript{45}. These supramolecular peptides bind multiple direct-repeat DNA sequences with positive cooperativity and pave the way to artificial transcription factors.

Attachment of supramolecular host-guest elements to complete proteins can be used to effectively induce protein heterodimerisation (Figure 5b)\textsuperscript{47}. Fluorescent proteins can be site-selectively modified with the lithocholic acid–β-cyclodextrin host-guest system,\textsuperscript{17} which features about a ten-fold higher affinity than the previously mentioned adamantane–β-cyclodextrin system. The host-guest system directs the formation of fluorescent protein heterodimers via the selective hetero-assembly of the host-guest system. Addition of an excess of either the lithocholic acid guest molecule or the β-cyclodextrin host molecule in their free form dismantles the protein dimer, via competition with the supramolecular protein elements. This supramolecular modulation of protein dimerization can not only be performed in buffered solution, but is also operative in cells. The application of an appropriately strong and selective supramolecular interaction thus allows for bioorthogonal protein assembly in a diverse set of biological media.
Figure 5: Supramolecular control over protein properties and function via attached host-guest elements. 

a) Supramolecular dimerization of two peptides enables strong and selective DNA recognition\(^4^4\); b) site-selective incorporation of host-guest elements in two fluorescent proteins allows inducing selective protein heterodimerisation\(^4^7\); c) host-molecule induced homodimerization of two proteins with genetically encoded N-terminal peptide tags\(^4^8\); d) an intramolecular supramolecular interaction inhibits the reconstitution of a split protein; supramolecular blockage of this interaction unfolds the peptide and enables GFP reconstitution\(^5^2\).

Another system based on host-guest chemistry can be used to induce homodimerization of genetically engineered proteins. Urbach et al found that cucurbit[8]uril can simultaneously bind to two N-terminal FGG peptides forming an inclusion complex (\(K_{\text{ter}} \approx 2 \times 10^{11} \text{ M}^{-2}\)\(^1^6\). Based on that finding Dung Dang in our group expressed two fluorescent proteins with an N-terminal FGG peptide motif (Figure 5c)\(^4^8\). Upon addition of the host cucurbit[8]uril the proteins formed stable...
homodimers that could be separated by size exclusion chromatography. After addition of a different guest molecule for CB[8], methylviologen, the protein dimer was disrupted which could be studied following FRET between the fluorescent proteins. Control proteins expressed with an MGG motif did not show any unspecific dimerization showing that this strategy is selective for FGG tagged proteins. It is therefore a promising tool to be used in biological environment as for example for the dimerization of membrane receptors.

Scherman et al. used a ternary cucurbit[8]uril complex for the formation of a supramolecular protein-PEG conjugate. They modified both BSA and poly(ethylene glycol) with the guest molecules viologen or naphthalene. Formation of a ternary host-guest complex could be observed after adding the naphthalene carrying BSA to the mixture of viologen and CB[8] by NMR and ITC. Using different combinations of the guest molecules on the protein or the PEG proved that the complex forms specifically only in presence of all three components of the ternary complex. This modular approach might give easy access to a variety of different polymer bioconjugates starting from relatively simple building blocks. Taking cucurbituril chemistry one step further from buffered solution towards cellular applications, Kim et al recently reported on the use of host-guest chemistry to isolate ferrocene-labeled membrane proteins from cell lysates. In analogy to pull down experiments for the isolation of biotinylated proteins from cell lysates with streptavidin coated beads, they modified sepharose beads with cucurbit[7]uril and showed that proteins which were functionalized with a ferrocene derivative selectively bind to these beads. They further labelled cells with reactive ferrocene derivatives and could then isolate proteins from the membrane by incubating the cell lysates with the beads. This is an example for a synthetic host-guest system that has comparable selectivity and affinity to the often used streptavidin-biotin pair and therefore might find various applications in biochemistry.

Supramolecular host-guest systems appended to protein elements can also be used to control the functional reassembly of a protein. So-called split proteins can be divided into two halves, whose reassembly reinstates a functional protein. This feature makes split proteins excellent tools for biological studies. In analogy with the previously discussed peptide and protein heterodimerisation systems, a β-cyclodextrin based host-guest system has been used to control the functional reassembly of two split-Green Fluorescent Protein (GFP) fragments (Figure 5d). A β-cyclodextrin and a coumarin unit were individually attached at each side of the small
3.3 Supramolecular protein immobilization

Another field where supramolecular chemical biology starts to be applied is the field of protein immobilization. Different strategies have been developed to immobilize functional proteins in a defined orientation to form structured surfaces that can be applied in the fields of bioanalytics and biomedicine. Surfaces modified with different proteins or peptides are further of interest for the study of cell growth and cell differentiation. The interaction of cells with synthetic surfaces is an important research area as it affects many applications such as tissue engineering, medical implanted materials and cellular assays. It is therefore of great interest to build surfaces where different biological ligands can be immobilized allowing control over distribution, density and orientation while preventing unspecific binding of ligands. Supramolecular chemistry can be an excellent tool that meets these requirements and offers the advantage of being reversible and often switchable giving access to dynamic surfaces that can mimic the natural environment in a better way than covalent chemistry does. There are already examples in the literature for protein immobilization based on recognition motifs such as NiNTA-His-tagged proteins, biotin-streptavidin and DNA-tagged proteins. Here the focus is on proteins that have been modified with a synthetic supramolecular element and then immobilized via specific host-guest interactions. The very strong ferrocene-cucurbit[7]uril interaction (Figure 3a), for example, has been used to immobilize the enzyme glucose oxidase on gold substrates (Figure 6a right). The cucurbit[7]uril was functionalized with alkanethiolate spacers and immobilized on gold surfaces. The glucose oxidase enzyme was randomly labelled (approximately 19 times on average) with ferrocenemethylammonium. The supramolecular protein immobilisation allows the protein to retain its correct fold and catalytic function. These protein monolayers can subsequently be used as glucose sensors.
Orientation controlled supramolecular protein immobilisation was shown using large protein complexes (Figure 6a left)\textsuperscript{58}. The light harvesting LH2 antenna complex was genetically engineered to contain cysteine residues at the end of the C-terminus of the \( \alpha \) polypeptide chain to enable a topological controlled modification. These cysteines were reacted with an iodoacetyl modified adamantyl derivative, to yield an average modification degree of three adamantyl guest molecules. The polyvalency of the adamantyl LH2 proteins significantly increases the affinity for \( \beta \)-cyclodextrin coated glass substrates. The proteins can be supramolecularly immobilized via simple incubation with the protein solution, or via nanoimprint lithography to achieve high spatial control over the protein immobilisation. In the same group \( \beta \)-cyclodextrin monolayers have been investigated as platforms for antibody recognition and in the next step for cell immobilization\textsuperscript{59}. They used a bivalent adamantane with a biotin functionality to immobilize streptavidin on the surface. Biotinylated protein G bound to the streptavidin surface and was recognized by a monoclonal antibody. On these surfaces lymphocytes were seeded and due to the immobilization of the monoclonal antibody via the biotinylated protein G, the specificity of the cell adhesion was increased compared to nonspecifically immobilized antibody. This example shows that supramolecularly immobilized proteins can form a platform for cell immobilization.

**Figure 6:** a) Site-specific\textsuperscript{58} or random\textsuperscript{57} modification of proteins with supramolecular ligands allows supramolecular protein immobilisation; b) Attachment of multiple adamantanes induces assembly of enzymes on supramolecular cyclodextrin polymersomes\textsuperscript{61}.

Finally, it is even possible to immobilize proteins on a surface in a controlled fashion utilizing a single host-guest interaction. A single ferrocene-cucurbit[7]uril recognition motif is sufficiently strong to generate stable homogenous protein monolayers\textsuperscript{60}. C-terminal incorporation of the ferrocene in a fluorescent protein allows for printing of
the protein on cucurbit[7]uril monolayers. The combination of supramolecular chemistry with protein immobilisation opens up the possibility to generate complex protein monolayers whose arrangement is dictated by the host-guest motifs present in the proteins and on the surface.

Analogous to the supramolecular immobilisation of proteins on surfaces, self-assembled vesicles can also be decorated with proteins via host-guest chemistry (Figure 6b)\textsuperscript{61}. Giant amphiphiles consisting of polystyrene end-capped with permethylated β-cyclodextrin form vesicular structures (polymersomes) on which the enzyme horseradish peroxidase, modified with adamantane groups, can be immobilized via the host-guest interaction. The adamantane groups were attached to the protein via long poly(ethylene glycol) spacers to facilitate polyvalent binding to the polymersome surface. The guest conjugated protein binds strongly to the polymersomes and is still catalytically active. The polymersomes can be washed and filtered whereby they retain binding to the protein.

3.4 Conclusion

Synthetic supramolecular host-guest systems are thus very well suited as bioorthogonal elements to control protein assembly and dimerization as well as to immobilise proteins on surfaces in a controlled and reversible manner. The attachment of synthetic supramolecular elements to proteins or the recognition of specific peptide motifs by supramolecular host molecules offers a platform to manipulate the properties of these proteins. Additionally, the reversible character of the supramolecular elements allows switching the protein assembly and function. Depending on the protein characteristics (such as valency of the assembly) and the specific requirements of the biological system (such as concentration regime of interest and extra- or intracellular environment), the characteristics of a supramolecular system can be chosen to allow an optimal interplay with the biological system. Control over membrane protein dimerization, transcription factor assembly, and enzyme complexes are only a few interesting protein complexes for which molecular control via novel chemical approaches is required. Another fruitful application of the combination of host-guest chemistry with proteins lies in the field of bionanotechnology. The generation of reversible, patterned protein surfaces could find applications in biosensors and tissue engineering.
4  Aim and outline of this thesis

The aim of this thesis is the application of host-guest chemistry for the assembly of proteins, both in solution and on surfaces. This includes the functionalization of different host-guest molecules, the subsequent modification of proteins with these molecules and finally the study of the protein assembly in solution and on surfaces. In chapters 2, 3, and 4 proteins were modified using expressed protein ligation which required expression of fluorescent proteins as thioesters and their ligation to supramolecular ligands functionalized with an N-terminal cysteine. The controlled heterodimerization of these supramolecular modified proteins was then investigated in solution using fluorescence spectroscopy studies. In these three chapters, different host-guest complexes were investigated for controlled protein assembly. The formation of A-B complexes was used in chapter 2 and 3 for the dimerization of two proteins. Upon mixing of the host- and guest-modified proteins, a protein heterodimer is formed which can be disrupted by addition of a competitor (Figure 7a). The host-guest complex between β-cyclodextrin and lithocholic acid was applied in chapter 2 to study the interplay between the intrinsic protein affinity and the supramolecular host-guest interaction using two different variants of fluorescent proteins. The introduction of a cysteine for expressed protein ligation allows for covalent locking of supramolecular protein dimers by forming disulfide bridges in an oxidizing environment.

Figure 7: Two different strategies to induce supramolecular protein heterodimers; a) A-B host-guest complex which can be reversed by addition of a competitor, b) A-B-C ternary host-guest complex which can be induced by addition of guest C.
Chapter 3 deals with the application of an improved host-guest complex between a modified cyclodextrin and lithocholic acid for the dimerization of two fluorescent proteins. In chapter 4, the formation of a ternary complex consisting of A, B and C is applied, where A and B are attached to the proteins of interest and C is the host-molecule which complexes A and B (Figure 7b). This system allows the controlled formation of a protein dimer in solution upon addition of an external ligand. Two fluorescent proteins were functionalized with the two host molecules methoxynaphthalene and methylviologen which allows for controlled protein assembly after addition of cucurbit[8]uril that forms a ternary complex with these two host-molecules.

For the modification of the proteins in chapters 2 to 4, expressed protein ligation was used. This technique allows for site selective modification of purified proteins at the C-terminus. However, it is not compatible with intracellular labelling or labelling of proteins in cell lysates. To take the concept of supramolecular host-guest chemistry for protein assembly a step further than working with purified proteins, an enzymatic labelling technique, the SNAP-tag labelling was introduced in chapter 5. SNAP-tag labelling requires expression of the protein of interest as alkylguaninetransferase fusion protein and functionalization of the supramolecular element with a benzylguanine moiety. Different supramolecular elements, including a supramolecular polymer were attached to fluorescent proteins studying the SNAP-tag as an alternative to expressed protein ligation. In chapter 6 fluorescent proteins were labelled via the SNAP-tag with ferrocene and adamantane and these proteins were then immobilized on cyclodextrin surfaces and vesicles.

5 References

Chapter 1

Chapter 2

Modulation of protein dimerization by a supramolecular host-guest system

Two sets of cyan (CFP) and yellow (YFP) fluorescent proteins, monomeric analogues and analogues with a weak affinity for dimerization, were functionalized with supramolecular host-guest molecules via expressed protein ligation. The host-guest elements induce selective assembly of the monomeric variants in a supramolecular heterodimer. For the second set of analogues, the supramolecular host-guest system acts in cooperation with the intrinsic affinity between the two proteins, resulting in the induction of selective protein-protein hetero-dimerization at more dilute concentration. Additionally, the supramolecular host-guest system allows locking of the two proteins in a covalent heterodimer via the facilitated and selective formation of a reversible disulfide linkage. For the monomeric analogues this results in a strong increase of the energy transfer between the proteins. The protein hetero dimerization can be reversed in a stepwise fashion. The trajectory of the disassembly process differs for the monomeric and dimerizing set of proteins. The results highlight that supramolecular elements connected to proteins can both be used to facilitate the interaction between two proteins without intrinsic affinity, and to stabilize weak protein-protein interactions at concentrations below those determined by the actual affinity of the proteins alone. The subsequent covalent linkage between the proteins generates a stable protein dimer as a single species. The action of the supramolecular elements in concert with the proteins thus allows the generation of protein architectures with specific properties and composition.

Chapter 2

1 Introduction

Protein-dimerization and aggregation are intensively studied phenomena. Understanding and modulating protein-protein interactions provides entries to study their associated diseases on the molecular level\(^1\). Signal transduction\(^2\), protein aggregation\(^3\), receptor clustering\(^4\), and transcription factor assembly\(^5\) are all processes in which protein-protein interactions and controlled assembly to higher order aggregates play a decisive role. Detailed molecular understanding and control of these interactions is hampered, for example, because structural information is absent, or because these interactions are regulated by unknown mechanisms and are latent under normal conditions\(^6\). Chemical biology approaches have been developed to selectively induce or inhibit specific protein-protein interactions with small molecules\(^7\). Chemical inducers of dimerization have, for example, provided versatile tools to modulate protein assembly and dimerization under the control of a small molecule\(^8\). Nevertheless, new and diverse synthetic strategies are required to selectively modulate protein-protein interactions and to assemble multi-protein complexes in a controlled, hierarchical fashion.

Supramolecular chemistry approaches have generated a vast variety of multicomponent architectures by virtue of non-covalent interactions\(^9\). Examples such as supramolecular switches\(^10\), electronics\(^11\), and polymers\(^12\) have shown that synthetic self-assembling systems allow control of material properties and functions and are a valuable contribution to the field of materials engineering. Many of the supramolecular examples in the materials field have been inspired by biological structures. Biological assemblies and molecules, such as cell-membranes, proteins, and polynucleotides, have been an important source of inspiration for supramolecular chemists to generate synthetic systems with analogous shape, size, or function\(^13\). The biological structures from which inspiration is derived all assemble under aqueous conditions. Water, however, is different from all other solvents and imposes specific requirements on self-assembling systems. Isolated secondary interactions of the polar type, such as for example hydrogen bonding, might drive self-assembly in organic solvents, but typically are not operative in water. In combination with a hydrophobic environment such polar interactions can be used to generate complex multicomponent assemblies with a well-defined and stable architecture in water\(^14\). As a result, supramolecular chemistry in water has come full circle and has become an attractive approach to apply to biological systems\(^15-18\). Nevertheless, only a limited set of initial examples can be found in the recent
Modulation of protein dimerization by a supramolecular host-guest system

literature, in which synthetic supramolecular systems have been used to modulate interactions between proteins. Vesicles were decorated with β-cyclodextrin to immobilize adamantane-modified enzymes via the host-guest interaction\textsuperscript{16}. A β-cyclodextrin based host-guest system has been used to control the reassembly of two split-Green Fluorescent Protein (GFP) fragments and reinstate its function\textsuperscript{17}. We have used the interaction between β-cyclodextrin and lithocholic acid to induce the assembly of two fluorescent proteins in solution and in the cell\textsuperscript{18}. Here we show in detail how the introduction of a supramolecular host-guest system in two sets of fluorescent protein variants, differing in their intrinsic affinity for dimerization, enables and modulates the selective formation of protein heterodimers (Figure 1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Schematic representation of supramolecular induced protein heterodimerization and covalent locking of two sets of fluorescent protein variants. Monomeric fluorescent proteins are induced to heterodimerize by the host-guest systems, resulting in the occurrence of energy transfer between the two proteins. Covalent locking of the complex through a selective disulfide bridge enhances the amount of energy transfer. Dimerizing fluorescent protein variants show a significantly stronger affinity for dimerization because of the additional interactions of the proteins in the heterodimer.}
\end{figure}

The host-guest system attached to the C-termini of two so-called monomeric fluorescent proteins, which have no intrinsic affinity for dimerization, induces a selective assembly of a protein heterodimer. When the same host-guest elements
are combined with two proteins that feature a weak affinity for unselective
dimerization, the supramolecular system induces a selective protein-protein
heterodimerization, at concentrations below those determined by the latent
interaction of the proteins. Furthermore, we show that the supramolecular induced
protein dimers can be selectively locked in a covalent heterodimer, by virtue of a
cysteine residue introduced by the supramolecular element in each protein. The
supramolecular elements thus provide control over protein dimerization in interplay
with the protein characteristics.

2 Protein design and synthesis
Fluorescent proteins are highly valuable tools to study the function of other proteins,
both in vitro and in vivo\(^\text{19}\). A large spectrum of different variants is available to the
scientific community, and their spectral characteristics make them perfect model
proteins in biology. Classical fluorescent proteins are known to feature a weak
affinity towards dimerization at elevated concentrations\(^\text{20}\). This intrinsic affinity could
lead to artefacts in the study of protein-protein interactions and therefore variants
have been developed that exhibit differing dimerization affinities, due to specific
point-mutations at the periphery of the protein. So-called monomeric Yellow
Fluorescent Protein and Cyan Fluorescent Protein (mYFP and mCFP) variants carry
the mutation A206K preventing their dimerization\(^\text{21}\). Furthermore, fluorescent
proteins with an increased tendency to dimerize were developed for enhanced
properties in Förster Resonance Energy Transfer (FRET)-based sensors\(^\text{22}\). These so-
called dimerizing variants (dYFP and dCFP) contain two mutations (S208F and V224L),
which increase their tendency to dimerize. These monomeric and dimerizing
fluorescent proteins thus differ in their intrinsic affinity, which makes them ideal
models to study protein dimerization. Based on these two variants we therefore
generated two sets of CFP and YFP proteins conjugated to supramolecular host-guest
elements. The monomeric fluorescent proteins were envisioned to be model proteins
for dimerization completely under control of external supramolecular elements. The
dimerizing fluorescent proteins represent a model system for proteins featuring a
weak, non-selective affinity for dimerization. This protein dimerization is stabilized
and made selective for heterodimerization by the external supramolecular elements.

We selected lithocholic acid and β-cyclodextrin as the supramolecular elements to be
connected to the fluorescent proteins for the manipulation of their assembly. This
host-guest system undergoes a strong, uni-directional heterodimerization in water
Modulation of protein dimerization by a supramolecular host-guest system

with a dissociation constant, \( K_d \), in the submicromolar range\(^{23} \), and has proven effective for the assembly of proteins in solution and in a cellular context\(^{18} \). The dimerization affinity of this supramolecular system is significantly higher than the dimerization affinity of any of the fluorescent proteins under study. This ensures that for the study of the protein dimerization with these supramolecular elements, experimental conditions can be chosen at which the intrinsic dimerization of the bare proteins is negligible in comparison with that of the host-guest pair (see below). This is especially relevant for the dimerizing proteins dYFP and dCFP as it ensures that any protein-protein dimerization event is not due solely to the intrinsic affinity of the proteins toward each other. \( \beta \)-Cyclodextrin and lithocholic acid were each connected via a small linker to a cysteine to enable C-terminal ligation to the fluorescent proteins (Figure 2). The thus introduced cysteines, close to the supramolecular host-guest pair, were envisioned to be used for the selective covalent locking of a supramolecular induced protein heterodimer, by tuning their oxidation state.

**Figure 2:** Design and synthesis of two sets of fluorescent protein variants (mCFP-mYFP and dCFP-dYFP) with no or a weak intrinsic affinity. The proteins are modified with supramolecular host-guest elements (lithocholic acid and \( \beta \)-cyclodextrin) via expressed protein ligation. The synthetic approach additionally introduces a cysteine directly after the supramolecular element at the site of ligation, which can be used for the formation of a disulfide bridge in the supramolecular induced protein dimer.

We used expressed protein ligation (EPL)\(^{24} \) for the site-selective incorporation of the supramolecular elements at the C-terminus of the proteins. The monomeric or dimerizing CFP and YFP were each fused with an intein domain, followed by a chitin binding domain. To enhance the protein expression levels in *E. coli* and to facilitate the purification of the proteins, an additional His-tag was N-terminally fused with the
CFP derivatives and a Strep-tag was N-terminally fused with the YFP derivatives. In the CFP-constructs an additional tobacco etch virus (TEV) protease cleavage site was introduced close to the C-terminus of the protein to allow enzymatic cleavage of the inserted supramolecular element. The resulting fusion proteins were expressed in E. coli, purified using chitin agarose beads and cleaved from the intein by sodium 2-sulfanyl ethane sulfonate (MESNA) to generate C-terminal thioesters (Figure 2). These protein thioesters were stored in phosphate buffer and expression yields typically ranged from about 40 mg (CFPs) to about 15 mg (YFPs) of fluorescent protein thioester per litre of bacterial culture.

The protein thioesters were connected to the above mentioned cysteine modified β-cyclodextrin and lithocholic acid via a native chemical ligation reaction (Figure 2). The different physicochemical nature of the β-cyclodextrin and lithocholic acid elements required two different ligation and purification strategies. For the conjugation of β-cyclodextrin to the YFP thioesters, the protein and cyclodextrin were mixed in a phosphate buffer containing 20 mM TCEP in a 1:20 ratio. After overnight incubation on a rotating wheel at room temperature, the reactions were typically complete and the excess of β-cyclodextrin and reducing agent were removed, resulting in the pure mYCD and dYCD constructs. Due to the limited solubility of the lithocholic acid-cysteine conjugate in water, the ligation of this supramolecular element to the CFP constructs was performed in the presence of 20 % methanol. The lithocholic acid and CFP thioester were mixed to yield a protein:lithocholic acid ratio of about 1:4 and incubated on a rotating wheel overnight at room temperature. Unligated CFP thioester could be separated from the ligated protein by an extraction with the detergent Triton X-114. After warming the reaction mixture with the detergent to 37°C, a separation of the detergent-rich phase containing the ligated protein and the water phase occurred, enabling the isolation of the pure conjugated proteins mCLA and dCLA by separation of the two phases. The purity and integrity of all proteins was established by mass spectrometry and SDS-PAGE.

3 Supramolecular induced protein dimerization

The dimerization of the two sets of supramolecular conjugated fluorescent protein pairs was investigated using fluorescence spectroscopy. All fluorescence spectra, including studies on YFP proteins alone, were recorded by using an excitation wavelength of 410 nm, which was optimized for CFP excitation with very limited YFP excitation. The non-conjugated protein thioesters mY, dY, mC, and dC (Figure 2) were
used as controls when appropriate. Fluorescence spectra were recorded of the individual proteins or their mixtures in a buffered solution that contained dithiothreitol (DTT) to ensure a reducing medium. To observe a possible interaction between the protein pairs, resulting in FRET, emission spectra of 1:1 mixtures of the YFP and CFP proteins were recorded. In the case of the occurrence of energy transfer between the two proteins, a decrease of the fluorescence maximum of CFP (donor) at 475 nm and an increase of the fluorescence maximum of the YFP (acceptor) at 527 nm, compared to the single proteins, are expected to be observed. The fluorescence spectra of all proteins individually and 1:1 mixtures of the YFP and CFP variants are shown in Figure 3 for the monomeric and dimerizing proteins respectively. The fluorescence spectra of the individual proteins mCLA and dCLA, as well as mYCD and dYCD are similar for the monomeric and dimerizing variants. Also the mixtures of mC-mYCD and dC-dYCD are similar for both variants and are a mere summation of the individual components. Mixtures of mCLA-mYCD and of dCLA-dYCD, however show the occurrence of energy transfer. For the monomeric pair, a small decrease in intensity of the mCFP maximum at 475 nm and the increase of the mYFP maximum at 527 nm occurs, compared to the mC-mYCD mixture, which lacks one supramolecular element. In case of the dimerizing proteins a very strong decrease of the dCFP fluorescence and a strong increase of the dYFP fluorescence could be observed, compared to the dC-dYCD mixture. The ratio of the intensities of the YFP and CFP maxima (Y/C ratio) was determined for all possible mixtures of monomeric and dimerizing proteins (Table 1). For all control experiments in which either one or two of the fluorescent proteins were lacking a supramolecular element, the ratios feature the same value of 0.51 for the monomeric proteins and around 0.56 for the dimerizing proteins. The modification of only one of the proteins with a supramolecular element clearly does not result in an increase of the Y/C ratio. This proves that a supramolecular element alone does not induce (non-specific) interactions. The Y/C ratio determined for the mCLA-mYCD complex (0.63) is higher than that of the corresponding monomeric controls (0.51). The dCLA-dYCD complex has a very high value for the same ratio (2.32).
Figure 3: Fluorescence spectra corrected for concentration of the set of monomeric fluorescent proteins (top) and dimeric fluorescent proteins (bottom): 1 µM mCLA or dCLA (blue); 1 µM mYCD or dYCD (black) and 1 : 1 mixtures of mC-mYCD or dC-dYCD (red); mCLA-mYCD or dCLA-dYCD (green) (concentration of the single components = 0.5 µM); spectra recorded using an excitation wavelength of 410 nm, measured in phosphate buffer (pH 7.5, 4 mM DTT).

For both sets of proteins energy transfer is only observed when both CFP and YFP feature a supramolecular element. Energy transfer between CFP and YFP can only occur when they are in molecular proximity. The supramolecular elements therefore must be inducing the proximity of the two proteins via their host-guest dimerization. The amount of energy transferred for the dimerizing proteins is significantly more than for the monomeric proteins. It can therefore be concluded that either more of the dCLA-dYCD complex is formed compared to the mCLA-mYCD complex, or that the energy transfer is stronger in the dCLA-dYCD complex than in the mCLA-mYCD complex. In general it is clear that there is an interplay between the protein pairs and the supramolecular elements, resulting in a different amount of transferred energy between the monomeric and dimerizing variants.
Table 1: Y/C Ratios ($I_{527\text{nm}}/I_{475\text{nm}}$) measured for mixtures of monomeric and dimerizing CFP and YFP proteins after incubation for 1h at 20°C (concentration of each protein = 0.5 μM); spectra recorded using an excitation wavelength of 410 nm, measured in phosphate buffer (pH 7.5, 4 mM DTT).

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Y/C Ratio</th>
<th>Proteins</th>
<th>Y/C Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>mY-mC</td>
<td>0.51</td>
<td>dY-dC</td>
<td>0.56</td>
</tr>
<tr>
<td>mCLA-mY</td>
<td>0.51</td>
<td>dY-dCLA</td>
<td>0.54</td>
</tr>
<tr>
<td>mC-mYCD</td>
<td>0.51</td>
<td>dYCD-dC</td>
<td>0.57</td>
</tr>
<tr>
<td>mCLA-mYCD</td>
<td>0.63</td>
<td>dYCD-dCLA</td>
<td>2.32</td>
</tr>
</tbody>
</table>

To quantitatively assess and gain deeper insight into the supramolecular induced protein dimerization, fluorescence lifetime measurements were performed. Fluorescence decay curves were recorded of both monomeric and dimerizing sets of CFP analogues, pure as well as mixed with their respective YFP analogues, all at a final concentration of 1 μM. The samples were excited at 364 nm to ensure the exclusive excitation of CFP and the decay of the fluorescence intensity of CFP was monitored at 475 nm. All samples for which at least one component lacked the supramolecular binding motif (nonbinding pairs), exhibited the double exponential decay characteristic for unmodified CFP\textsuperscript{25}. Fitting the recorded decay curves yielded lifetime components of $\tau_1 \approx 1.5$ ns and $\tau_2 \approx 3.8$ ns for both the dimerizing as well as the monomeric set of samples (Table 2). The absence of any further, faster decay component provides evidence that for CFP and YFP lacking the host-guest pair no energy transfer is taking place.

The picture drastically changes for pairs of CFP and YFP both containing their respective supramolecular element. A third, much shorter decay component ($\tau_Q$) appears for the dimerizing pair and, to a lesser extent, for the monomeric pair, unseen in all other samples. Energy transfer from a FRET donor to a FRET acceptor acts as an additional decay channel for the FRET donor, resulting in the observed shorter lifetime. Thus, clearly, the host-guest interaction results in the dimerization of the CFP and YFP and therefore in energy transfer from CFP to YFP. The $\tau_Q$ component is very short and approaches the limit of time resolution of the setup, indicative of very high energy transfer efficiency. The coexistence of the three lifetime components of $\tau_1$, $\tau_2$, and $\tau_Q$ indicates the parallel presence of CFP-YFP pairs and free CFP in the solution. Since we have determined the decay characteristics for
uncoupled CFP before (Table 2), we fitted the decay curves of the binding pairs using
3 exponentials, keeping two components fixed at the values for $\tau_1$ and $\tau_2$ of the
respective mCLA and dCLA. The fast, quenched decay component $\tau_Q$ could thus be
determined for both sets of proteins (Table 2). The $\tau_Q$ of the dimerizing set of
proteins (dCLA-dYCD) was determined as $\sim$300 ps contributing with 12% relative
amplitude, while the $\tau_Q$ of the monomeric set (mCLA-mYCD) contributed with only
5% relative amplitude and was even shorter, $< 200$ ps, and approached the time
resolution limit of the setup. The different relative contributions of $\tau_Q$ to the overall
signal for the monomeric and the dimerizing protein pairs show that the affinity to
heterodimerize is much larger for the dimerizing set of proteins than for the
monomeric set. The data also show that the monomeric set of proteins yields a decay
component $\tau_Q$ with a much shorter lifetime than for the dimerizing set. This
potentially indicates that, whereas the amount of protein dimer is higher for the
dimerizing variants, a more efficient energy transfer is occurring in the monomeric
set of proteins. In contrast to the dimerizing variants, the orientation between the
monomeric proteins is not fixed. This flexibility of the monomeric proteins may
account for the formation of a supramolecular protein dimer with an overall
orientation that features more efficient energy transfer than for the dimerizing
proteins.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\tau_Q$ (ns)</th>
<th>$\tau_1$ (ns)</th>
<th>$\tau_2$ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mC</td>
<td>-</td>
<td>1.4</td>
<td>3.9</td>
</tr>
<tr>
<td>mCLA</td>
<td>-</td>
<td><strong>1.4</strong></td>
<td><strong>3.8</strong></td>
</tr>
<tr>
<td>mC-mY</td>
<td>-</td>
<td>1.3</td>
<td>3.8</td>
</tr>
<tr>
<td>mC-mYCD</td>
<td>-</td>
<td>1.5</td>
<td>3.9</td>
</tr>
<tr>
<td>mCLA-mY</td>
<td>-</td>
<td>1.3</td>
<td>3.8</td>
</tr>
<tr>
<td>mCLA-mYCD</td>
<td>$&lt;0.2$</td>
<td><strong>1.4</strong></td>
<td><strong>3.8</strong></td>
</tr>
<tr>
<td>dC</td>
<td>-</td>
<td>1.6</td>
<td>3.9</td>
</tr>
<tr>
<td>dC-mY</td>
<td>-</td>
<td><strong>1.5</strong></td>
<td><strong>3.7</strong></td>
</tr>
<tr>
<td>dC-dY</td>
<td>-</td>
<td>1.5</td>
<td>4.0</td>
</tr>
<tr>
<td>dC-dYCD</td>
<td>-</td>
<td>1.6</td>
<td>3.9</td>
</tr>
<tr>
<td>dCLA</td>
<td>-</td>
<td>1.6</td>
<td>3.9</td>
</tr>
<tr>
<td>dCLA-dY</td>
<td>-</td>
<td>1.6</td>
<td>3.9</td>
</tr>
<tr>
<td>dCLA-dYCD</td>
<td>0.3</td>
<td><strong>1.5</strong></td>
<td><strong>3.7</strong></td>
</tr>
</tbody>
</table>

In order to compare the dimerization affinities of both the monomeric and the
dimerizing variants, titration studies were performed to determine the dissociation
constants ($K_d$) of the complexes. This was done by titrating mYCD and dYCD to 0.2 μM solutions of mCLA and dCLA, respectively. Upon addition of increasing amounts of YFP, increasing amounts of complex are formed. At any point during the titration the solutions will contain differing amounts of free CFP and YFP components as well as formed CFP-YFP complexes. As can be seen for the monomeric set (Figure 4a) and for the dimerizing set (Figure 4b), the intensity of CFP fluorescence decreases upon addition of YFP, while YFP fluorescence increases. It can be concluded that as more YFP is added, the population of CFP molecules for which transfer of energy is occurring is increasing, indicative of increasing concentrations of heterodimer forming in the mixture.

\[
\Delta I = I_0 - I = \frac{\Delta I_{\text{max}} \cdot c(\text{acceptor})}{K_d + c(\text{acceptor})}
\]  

(1)

using the fluorescence intensity of the donor CFP in the absence ($I_0$) and presence of acceptor YFP ($I$) (Figures 4c and 4d). In order to determine the $K_d$ for each set of proteins, equation (1) was used to fit the titration curves assuming first order kinetics. This yields a $K_d$ for the dCLA-dYCD complex of 0.4 ± 0.1 μM and for the mCLA-mYCD of 4.3 ± 0.2 μM. The dissociation constant of the dimerizing variants is
nearly ten times smaller than that of the monomeric variants. The two dimerizing proteins thus exhibit a significantly stronger binding affinity, even though the supramolecular elements attached to the two sets of proteins are the same. The interaction of lithocholic acid with \( \beta \)-cyclodextrin in water, as measured by ITC, is reported to have a dissociation constant of 1.2 \( \mu \)M\(^{23} \). The \( K_d \)-value of the monomeric proteins modified by the host-guest system (4.3 \( \mu \)M) thus shows that the interaction of these two proteins is governed by the supramolecular elements attached to the proteins. The higher affinity observed for the dimerizing set (0.4 \( \mu \)M) shows that there is an effect additional to the supramolecular interaction contributing to an increased binding affinity. This is most probably due to a binding event of the supramolecular host-guest pair in conjunction with the intrinsic dimerization affinity of the proteins.

Since the dimerization of the proteins in both hetero-complexes is governed by a supramolecular recognition motif, dissociation should be inducible by inhibition of the supramolecular \( \beta \)-cyclodextrin-lithocholic acid host-guest interaction. In order to test this, unfunctionalized \( \beta \)-cyclodextrin was added to the mCLA-mYCD and dCLA-dYCD complexes. The free \( \beta \)-cyclodextrin host acts as competitive binder for the lithocholic acid conjugated to the CFP (Figure 5). For both protein complexes the addition of \( \beta \)-cyclodextrin results in a decrease of the YFP fluorescence, which, together with the recovery of the CFP fluorescence, shows that the interaction between the proteins is indeed of a reversible nature and can be modulated with small supramolecular moieties.

![Figure 5: Fluorescence spectra of 0.5 \( \mu \)M solutions of a supramolecular protein complex before and after addition of 2 mM \( \beta \)-cyclodextrin; mCLA-mYCD before (blue), mCLA-mYCD after (black), dCLA-dYCD before (green), dCLA-dYCD after (red) (concentration of both proteins= 0.5 \( \mu \)M); spectra recorded at an excitation wavelength of 410 nm, measured in phosphate buffer, (pH 7.5, 4 mM DTT).](image)
A TEV-cleavage site incorporated in the C-terminus of the CFPs makes it possible to disassemble the protein heterodimers via an enzymatic peptide cleavage reaction. Treatment of the assembled protein complexes mCLA-mYCD and dCLA-dYCD with TEV protease results in the cleavage of a short C-terminal peptide sequence with the lithocholic acid from the CFPs, thereby releasing the CFPs from the complex. For this, the mCLA-mYCD and dCLA-dYCD complexes were incubated for 1h at 20°C with TEV-protease. This resulted in a decrease of the energy transfer in the samples (Figure 6). This proves that the TEV cleavage sites within the protein dimers are accessible for the enzyme and that cleavage of this site dismantles the protein complex. The observed lowering of the energy transfer is especially relevant for the dimerizing proteins, as it shows that the enhanced energy transfer observed for the dCLA-dYCD complex compared to the mCLA-mYCD complex is indeed due to the supramolecular interaction, and not to unspecific interactions of the proteins. The observed energy transfer has almost completely disappeared for mCLA-mYCD mixture, but has only partly been lowered for the dCLA-dYCD mixture during the same amount of time. Apparently, the stronger dimerization interaction between the dimerizing proteins makes the complex less accessible and reactive towards the TEV protease.

![Figure 6: Fluorescence spectra of solutions of 1 µM protein complex before and after 1h incubation with AcTEV™-protease (10 U TEV per 10 µg CFP); mCLA-mYCD before (blue), mCLA-mYCD after (black), dCLAd-dYCD before (green), dCLA-dYCD after (red); (concentration of both proteins= 1 µM); spectra recorded using an excitation wavelength of 410 nm, measured in phosphate buffer (pH 7.5, 4 mM DTT).](image)
4 Selective covalent locking of supramolecularly induced protein heterodimers

Our strategy for generation of the supramolecularly tagged fluorescent proteins via EPL resulted in the introduction of a cysteine with a free thiol side-chain in each protein, close to the host-guest interaction site. These thiol functionalities were planned for the subsequent covalent capture of the generated protein heterodimers via selective oxidation to a disulfide. The supramolecular assembly would thus guarantee the selective formation of CFP-YFP disulfide bridged protein heterodimers. We first investigated whether it was possible to covalently capture the complex between lithocholic acid and β-cyclodextrin without protein (Scheme 1). A mixture of the cysteine conjugated β-cyclodextrin and the cysteine conjugated lithocholic acid in phosphate-buffer was incubated for 4 hours under oxidizing conditions. The resulting reaction product was analyzed by MALDI-TOF mass spectrometry (Figure 7). The mass spectrum features three strong signals corresponding to both individual compounds and the selectively formed disulfide-bridged hetero-complex. Addition of DTT to the covalent captured dimer resulted in the disappearance of the signal of the hetero-complex in the spectra. The cysteine modified lithocholic acid and β-cyclodextrin can thus form an inclusion complex and lock this complex into a covalent dimer under oxidizing conditions. By changing the redox state of the buffer, the covalent locking can again be removed.

Scheme 1: Formation of a covalent heterodimer between lithocholic acid and β-cyclodextrin via reversible oxidation of the specifically introduced cysteines.
Modulation of protein dimerization by a supramolecular host-guest system

**Figure 7:** MALDI-TOF-spectra of disulfide cross-linked lithocholic acid/β-cyclodextrin heterodimer before (a) and after treatment with DTT (b).

The covalent locking was also studied for the proteins. First covalently locked protein heterodimers were generated for both the mCLA-mYCD and the dCLA-dYCD protein pairs by equilibrating the mixtures under ambient atmosphere, enabling oxidation of the introduced cysteines to a disulfide. Both protein pairs showed strong energy transfer (Figure 8). When these locked protein pairs were treated with β-cyclodextrin no decrease of the energy transfer could be observed in the fluorescence spectra. When subsequently DTT was added to reduce the disulfide bridge and unlock the covalent complex, a strong decrease of the energy transfer could be observed in both samples. These results show that covalent locking of the proteins into a stable disulfide bridged heterodimer indeed occurs. After covalent locking, the supramolecular host-guest interaction is not required anymore for protein dimerization.
Figure 8: Fluorescence spectra of 0.5 µM preformed covalently locked oxidized protein complex before (blue) and after adding β-cyclodextrin (2 mM; red), followed by DTT (4 mM; green) for a) mCLA-mYCD and b) dCLA-dYCD. The solutions were prepared in phosphate buffer, pH 7.5 and incubated at 4°C for 2 days before measurement. Spectra were recorded using an excitation wavelength of 410 nm.

To further investigate the interplay of the proteins and the supramolecular host-guest system, the stepwise disassembly of the covalently locked protein dimers with β-cyclodextrin and DTT was carried out in reverse order. First the covalently trapped complexes were reduced with DTT, and then the supramolecular host-guest interaction was inhibited with β-cyclodextrin (Figure 9). In this experiment a clear distinction in the disassembly process between the mCLA-mYCD and the dCLA-dYCD protein pairs could be observed. After reduction of the disulfide bridge, no change was observed for the dimerizing proteins, whereas for the monomeric proteins, the energy transfer decreased significantly. Subsequent treatment with β-cyclodextrin then further decreased energy transfer of the monomeric variants and also perturbed the energy transfer of the dimerizing variants.

Figure 9: Fluorescence spectra of 0.5 µM preformed covalently locked oxidized protein complex before (blue) and after adding DTT (4 mM; red), followed by β-cyclodextrin (2 mM, green) for a) mCLA-mYCD and b) dCLA-dYCD. The solutions were prepared in phosphate buffer, pH 7.5 and incubated at 4°C for 2 days before measurement. Spectra were recorded using an excitation wavelength of 410 nm.
These results clearly show that for the monomeric proteins the supramolecular elements induce the assembly of a heterodimer, but a significant amount of transferred energy only occurs, when the complex subsequently gets covalently locked. The dimerizing proteins are already strongly associated via the supramolecular host-guest pair, in conjunction with the interaction of the protein domains. Covalent locking of the dCLA-dYCD complex does not increase the energy transfer further.

A covalent complex generated via disulfide formation should still be amendable for inhibition of energy transfer by enzymatic cleavage of the TEV-cleavage site. After incubating 1:1 mixtures of both the mCLA-mYCD and dCLA-dYCD covalently locked complexes with TEV-protease for 4 hours, a decrease in energy transfer was observed for both pairs of proteins (Figure 10). This experiment also proves that the covalent complex indeed resulted from the selectively introduced disulfide and that no other covalent bonds between the proteins were formed. The observed energy transfer has almost completely disappeared for the mCLA-mYCD pair, but has only partly been lowered for the dCLA-dYCD pair during the same amount of time, consistent with studies on the non-covalently locked dimers (Figure 7). Apparently, also in the covalently locked form, the dimerizing proteins are more tightly packed than their monomeric counterparts, making the heterodimer less accessible for the protease.

![Figure 10](image_url)

**Figure 10:** a) Fluorescence spectra of solutions of 0.5 µM disulfide oxidized protein complex before and after 4h incubation with TEV-protease (20 U TEV per 5 µg CFP); mCLA-mYCD before (blue), mCLA-mYCD after (black), dCLA-dYCD before (green), dCLA-dYCD after (red); (phosphate buffer, pH 7.5; excitation wavelength 410 nm). b) Model of the covalent complex and its cleavage with TEV protease, illustrated for the dimerizing variants dCLA and dYCD.

Based on the above results the assembly process of the monomeric and dimerizing fluorescent proteins into selective heterodimers by virtue of the supramolecular host-guest interaction becomes clear (Figure 11). Supramolecular host-guest
elements attached to the monomeric proteins selectively dimerize mCFP with mYFP proteins. The proteins are brought to dimerize via the supramolecular host-guest interaction, but a specific protein-protein interaction is not induced. A subsequent covalent locking of the complex by oxidation of the synthetically introduced cysteines stabilizes the dimer and probably brings the proteins in somewhat closer proximity. The dimerizing proteins are brought into proximity via the same host-guest interaction, which allows the intrinsic affinity between these proteins to come into play and additionally stabilize the heterodimer. This results in a 10 fold lower $K_d$ of the dCLA-dYCD pair versus the mCLA-mYCD pair. The bivalent interaction of the host-guest system together with the intrinsic protein dimerization affinity results in a strong and selective formation of a heterodimeric protein-protein interaction. Covalent locking of this complex is possible, however, does not result in changes in the energy transfer.

![Supramolecular Protein Dimerization of Monomeric Proteins](image)

**Figure 11**: Schematic model for the hierarchical assembly / disassembly process of the mCLA-mYCD and the dCLA-dYCD protein pairs. Depending on the dimerization characteristics of the protein, the supramolecular host-guest pair either induces selective assembly of a protein dimer via the supramolecular elements alone, or induces the formation of a more stable protein heterodimer by virtue of the interplay with the protein dimerization surface.

## 5 Conclusions
The dimerization of proteins in biological systems can be the result of diverse processes\(^2,6\). Gaining control over such protein interactions would allow the
controlled formation of multi-component protein complexes at specific locations in or outside the cell and represents a great challenge. We have developed a supramolecular host-guest system that provides a novel supramolecular approach to controlling these processes. Two sets of fluorescent CFP and YFP variants, monomeric and weakly dimerizing analogues, were used as model systems for protein dimerization. These proteins were functionalised with the supramolecular lithocholic acid - β-cyclodextrin host-guest system, via expressed protein ligation. The supramolecular elements induce a selective assembly of the monomeric variants in a protein heterodimer. When attached to the dimerizing variants, the supramolecular host-guest system acts in cooperation with the intrinsic affinity between the two proteins and enables the selective formation of protein heterodimers at concentrations much lower than those determined by the actual affinity of the proteins alone. This bivalent interaction results in the induction of a selective protein hetero-dimerization featuring very strong energy transfer. Both the monomeric and dimerizing supramolecular induced protein heterodimers subsequently can be covalently locked in a stable heterodimer via a selectively introduced and reversible disulfide linkage. This results in an increase of the energy transfer between the monomeric proteins. The two sets of proteins undergo two distinct hierarchical assembly processes, which are the result of the interplay of the supramolecular host-guest systems with the proteins’ affinity for dimerization.

The attachment of synthetic supramolecular elements to two different proteins allows controlling the interactions between these two proteins. For proteins that do not feature an intrinsic interaction, the host-guest system can be used to facilitate a hetero-specific molecular level dimerization of these proteins. This type of interaction mode opens up the possibility to generate designed protein assemblies under control of supramolecular interactions and could for example prove valuable for the specific colocalization of proteins at specific compartments in- or outside of the cell. For proteins that feature a weak intrinsic dimerization affinity, supramolecular host-guest elements can induce the selective stabilization of the latent protein-protein interaction. The host-guest system acts in cooperation with the intrinsic protein affinity and can be used to induce the protein-protein interactions at concentrations well below the normal dimerization affinity. This interaction mode provides a tool to selectively induce and control weak protein-protein interactions. The host-guest system thus could allow studying of for example protein-protein interactions for which the biological regulation mechanisms are unknown. The combination of supramolecular chemistry with proteins provides a
novel strategy for the control over protein dimerization and in our view offers significant potential to assemble protein complexes in a controlled hierarchical fashion.

6 Experimental part

Plasmids were constructed by Dr. Hoang Nguyen. Fluorescence lifetime measurements were done together with Dr. Christian Blum and Dr. Dorothee Wasserberg.

6.1.1 Synthesis

Synthesis of cysteine-conjugated lithocholic acid and cysteine-conjugated β-cyclodextrin was performed as described\(^{18}\).

6.1.2 MALDI-spectrometry

A 4 mM solution of cysteine-conjugated lithocholic acid (LA-Cys) and cysteine-conjugated β-cyclodextrin (CD-Cys) was prepared in buffer (50 mM Na-Pi, 30 mM NaCl, pH 7.5) and incubated at room temperature for 4 h. The sample was desalted and analyzed by MALDI-TOF: m/z: LA-Cys calculated: 797.5 [M+H\(^+\)], found: 797.4 [M+H\(^+\)], 819.4 [M+Na\(^+\)]; CD-Cys calculated: 1280.4 [M+H\(^+\)], found: 1280.0 [M+H\(^+\)], 1302.0 [M+Na\(^+\)]; LA-Cys-S-S-CD-Cys calculated: 2096.9 [M+Na\(^+\)], found: 2096.7 [M+Na\(^+\)]. DTT was then added to a final concentration of 5 mM and after 30 min the mixture was analyzed again by MALDI-TOF: m/z: LA-Cys: 797.5 [M+H\(^+\)], 819.6 [M+Na\(^+\)]; CD-Cys: 1302.2 [M+Na\(^+\)].

6.1.3 Spectroscopy

Samples for spectrophotometric measurements were prepared under ambient conditions in quartz cuvettes of 0.5 cm pathlength and 300 µL minimal volume. Samples for experiments performed under oxidizing conditions were prepared in phosphate buffer (50 mM Na-Pi, 30 mM NaCl, pH 7.5) according to the concentrations given in the text. If the experiment was done under reducing conditions, DTT (4 mM) was added to the buffer. The concentration of the proteins was measured and adjusted by UV/Vis-spectroscopy using the absorbance at 435 nm (A\(_{435}\)) and a molar extinction coefficient of 32 500 M\(^{-1}\)cm\(^{-1}\) for cyan fluorescent proteins (CFP)\(^{26}\) and using A\(_{515}\) and a molar extinction coefficient of 84 000 M\(^{-1}\)cm\(^{-1}\) for yellow fluorescent proteins (YFP)\(^{27}\). UV/Vis spectra were recorded on a JASCO, V-550 spectrophotometer, fluorescence data were recorded on a JASCO FP-6500
fluorimeter equipped with a water-cooled Peltier thermostatted cell holder. UV/Vis data were recorded under ambient conditions and at room temperature. All steady-state fluorescence data were recorded using an excitation wavelength of 410 nm and at 20°C but otherwise under ambient conditions. All experiments were repeated at least once. Measurement parameters and sample concentrations were kept constant (unless indicated otherwise) to make the data comparable. Time-resolved fluorescence data were obtained at an emission wavelength of 475 nm and a time range of 50 ns using a HORIBA Jobin Yvon FluoroMax-4 equipped with a pulsed NanoLED excitation source from IBH (FWHM = 1.3 ns, \( \lambda = 364 \) nm) and a TC-SPC (time correlated – single photon counting) detections system (FluoroHub, HORIBA Jobin Yvon) based on time to amplitude conversion (TAC). The fluorescence time-resolved data were fitted using the DAS6 Decay Analysis Software package from HORIBA Jobin Yvon. A machine response function was recorded on strongly scattering silica oxide particles in water using identical settings as for the samples; only the emission wavelength was adjusted. Deconvolution analysis was carried out with 2 or 3 exponentials. Fluorescence lifetime data were recorded under ambient conditions and at room temperature. Fluorescence lifetime measurements were repeated at least twice. Measurement parameters and sample concentrations were kept constant (unless indicated otherwise) to make the data comparable.

6.1.4 Construction of plasmids

To enhance protein expression levels of ECFP and EYFP in *E. coli* and to facilitate the purification steps, a His-tag was fused N-terminal to the ECFP derivatives and a strep-tag was fused N-terminally to the EYFP derivatives. The hybridized-forms of the two complementary oligonucleotides ONO31, 5’- tatggaagcagcaccatccacctcacatccacctatcc-3’ and ONO32, 5’- tagaatgt tgtatagtgtatgtatgtatgtatgtggctgcttc-3’ encoding for His-tag and ONO47, 5’- tatggaagcagcaccatccacctcacatccacctatcc-3’ and ONO48, 5’- tagcagaccttggctgccatggcttcagctgc-3’ encoding for Strep-tag were introduced into pTW1N-ECFP and pTW1N-EYFP at the *NdeI* restriction site, resulting in plasmids pH7T47 and pH4T82 respectively. In order to generate the monomeric constructs, plasmids pH7T47 and pH4T82 were amplified using primers ON097, 5’- gagcaccagctgcagcaccatccacctcacatccacctatcc-3’ and ON098, 5’- gggtcttggctgcgtggctgtgtgtgtcg-3’, by which the mutation A206K was introduced, resulting in plasmids pH4T76 (His-mCFP) and pH4T86 (Strep-mYFP). Similarly, plasmids pH4T83 (His-dCFP) and pH4T84 (Strep-dYFP) carrying the S208F and V224L mutations, as dimerizing variants were constructed using primers ON099, 5’- caagcagaaagcgcagcaccatccacctcacatccacctatcc-3’ and ON100, 5’- gaccatgtgtgtgcgcgttctctggtggggtcttggg-43
aacaggccggactgggtgc-3’ with pHT477 as template; and ON099 and ON102, 5’-gaccatgtgatccggttcgtttggtctttgaacaggccggactgtagc-5’ with pHT482 as template, respectively.

### 6.1.5 Protein expression and purification

The resulting plasmids were transformed into E. coli Rosetta 2 (DE3) cells, and transformants were selected on ampicillin (125 mg L\(^{-1}\)) and chloramphenicol (34 mg L\(^{-1}\)) agar plates. A single colony was inoculated into LB medium (5 mL) containing ampicillin (125 mg L\(^{-1}\)) and chloramphenicol (34 mg L\(^{-1}\)), and the culture was grown overnight at 37 °C. This preculture was used to seed fresh LB medium (1 L) containing antibiotics, and the culture was incubated at 37 °C until the absorbance at 600 nm (OD600) reached 0.7-0.9. After cooling down, IPTG was added to a final concentration of 0.5 mM, and overnight (or 6 h) induction was performed at 20 °C. Cells were harvested by centrifugation (5000 g, 20 min, 4 °C). The pellet was frozen at -80°C. After defreezing it was resuspended in lysis buffer (25 mM Na-phosphate, pH 7.5, 0.5 M NaCl, 1 mM PMSF, 0.5 mM EDTA), and cells were lysed by sonication. Triton X-100 was added to a final concentration of 1%. The lysate was cleared by centrifugation (20000 rpm, 30 min, 4 °C). Columns (EconoPac, Biorad, 20 mL) containing 10 mL of chitin beads (New England Biolabs) were equilibrated with washing buffer (25 mM Na-Phosphate, 0.5 M NaCl, 0.5 mM EDTA). The supernatant was applied to the columns at room temperature (22 °C). The beads were washed with 10-15 column volumes of washing buffer. Cleavage of the fusion proteins was induced by suspending the beads in cleavage buffer (25 mM Na-Phosphate, 500 mM NaCl, 0.5 mM EDTA, 0.5 M MESNA) and overnight incubation at 18 °C. The supernatant was collected and passed over a small column to separate from the non cleaved fusion proteins. The pooled fractions were concentrated and the buffer exchanged three times to storage buffer (25 mM Na-Phosphate, 50 mM NaCl, 0.1 mM EDTA, 20 mM MESNA) and concentrated to 10 mg mL\(^{-1}\). The protein was shock frozen in multiple aliquots using liquid nitrogen and stored at -80 °C. Yields typically ranged from about 40 mg (CFPs) to about 15 mg (YFPs) of fluorescent protein thioester per liter of bacterial culture.

### 6.1.6 Protein ligations

**YFP-thioesters + CD-Cys:** A solution of CD-Cys (8 mM) in ligation buffer (30 mM Na-Phosphate pH 7.5, 50 mM NaCl degassed by bubbling Argon through the buffer) was prepared. TCEP (400 mM pH 7 in H\(_2\)O) was added in order to reduce all the cysteines. To the protein-solution that was stored at -80°C, a solution of MESNA (2.5 M stock in
ligation buffer) was added to a final concentration of 0.5 M. After 30 min the two solutions were mixed in a 1 to 1 ratio and then incubated on a rotating wheel overnight at room temperature. The mixture was analysed by SDS-PAGE or MALDI-TOF spectrometry. If the ligation had not gone to completion additional CD-Cys was added to the solution and then incubated for additional 5 h. The reaction mixture was centrifuged (1 min, 13000 rpm) to remove any precipitate. The excess of CD-Cys and MESNA could be removed by exchanging the buffer several times using small centrifugal filters (Millipore) with a nominal molecular weight limit of 10 kDa. The pure cyclodextrin ligated YFP was stored in Na-Phosphate buffer at -80°C.

**CFP-thioesters + LA-Cys:** A suspension of LA-Cys (2.5 mM) in MeOH was prepared and a solution of TCEP (400 mM in H₂O or ligation buffer, pH 7) was added to a final concentration of 75 mM, resulting in dissolving of the LA-Cys. To the protein solution that was stored at -80°C, a solution of MESNA (2.5 M in ligation buffer) was added to a concentration of 0.25 M MESNA. The LA-Cys solution and the protein solution were mixed to about 3-4 eq LA-Cys for 1 eq of CFP thioester. The solution was incubated on a rotating wheel overnight at room temperature. The reaction was controlled by SDS-PAGE or MALDI. If the ligation had gone to completion, the reaction mixture was centrifuged (1 min, 13000 rpm) to remove any precipitate. Then the excess LA-Cys and MESNA could be removed by exchanging the buffer several times using small centrifugal filters (Millipore) with a nominal molecular weight limit of 10 kDa. The LA-ligated CFP was stored in Na-Phosphate buffer at -80°C. If the ligation did not go to completion, it was possible to separate ligated and unligated protein by a Triton-extraction as described previously¹⁸.

7 References


12 Supramolecular Polymers, 2nd ed. (Ed.: A. Ciferri), CRC, Boca Raton, FL, **2005**.


Chapter 3

An improved host-guest system for supramolecular protein dimerization

The supramolecular host-guest complex between lithocholic acid and heptakis-[6-deoxy-6-(2-aminoethylsulfanyl)]-β-cyclodextrin (cysteamine-cyclodextrin) was applied for the dimerization of two fluorescent proteins. A yellow fluorescent protein was ligated via expressed protein ligation to a cysteine-functionalized cysteamine-cyclodextrin. First, the synthesis and purification of the mono-functionalized cysteamine-cyclodextrin are described. The second part deals with the ligation of the host-molecule to monomeric yellow fluorescent protein (YFP) and subsequent purification of the ligation product. The efficiency of the new host-molecule for the formation of a supramolecular protein dimer was investigated using fluorescence spectroscopy. The complexes of monomeric cyan fluorescent protein (CFP) ligated to lithocholic acid with both, cysteamine-cyclodextrin-YFP and β-cyclodextrin-YFP were compared. A higher FRET-effect was observed in the case of the cysteamine-cyclodextrin, and first $K_d$-measurements suggested that the affinity of this complex is higher than that based on the bare β-cyclodextrin.
1 Introduction

The application of supramolecular host-guest chemistry for the controlled assembly of proteins has recently drawn the attention of different research groups and areas\(^1\). The choice of host-guest systems for controlling proteins is dictated by the molecular properties of the supramolecular system. First, the supramolecular complexes typically should have interaction strengths that allow formation at low concentrations (i.e. \(\mu\text{M} - \text{nM}\)) in aqueous media. Secondly, it should be possible to functionalize the host and guest molecules with a reactive group for attachment to the protein of interest. There are only a few systems known that meet both criteria. Certain complexes based on cucurbituril host-guest complexes, for example, show very high affinities in water\(^2\). The interaction between positively-charged adamantane derivatives and cucurbit[7]uril displays an affinity constant, \(K_a\), in buffered solution of \(10^{12}\) M\(^{-1}\)\(^2\). The binding of the larger derivative, cucurbit[8]uril, to these adamantane derivatives is weaker (\(K_a = 10^8 - 10^{10}\) M\(^{-1}\)) but the \(K_d\) is still in the nanomolar range. Because of these properties and their experience with the functionalization of cucurbiturils\(^3\), the group of Kim has already used cucurbituril complexes with ferrocene derivatives for protein immobilization\(^4\) and to capture labeled membrane proteins from cell lysates\(^5\). However, methods for selective mono-functionalization of cucurbiturils are not yet reliable compared to cyclodextrins and the purification of cucurbituril mixtures presents many challenges. Given that the mono-functionalization of the host molecule is an important requirement for the use in expressed protein ligation, cyclodextrins constitute a highly attractive class of host-molecules because they are watersoluble and can be readily mono-functionalized. Dissociation constants of complexes of cyclodextrins with organic guest molecules are in the millimolar to micromolar range for non-functionalized cyclodextrins\(^6\). The synthesis of cyclodextrin dimers results in receptors with even higher affinities in the submicromolar range\(^7,8\). The affinity for cholesterol, for example, increased 200-300 times when a cyclodextrin dimer was applied\(^9\). The introduction of a functional linker that is, for example, photocleavable allows for the “on/off” switching of the interaction\(^10\). In general the synthesis of cyclodextrin dimers is straightforward; however the attachment of a cysteine for ligation to a protein is more difficult and requires the design and optimization of linker molecules. Another possibility to increase the affinity of cyclodextrins to specific guests is via modified cyclodextrins with extended lipophilic cavities. A \(\gamma\)-cyclodextrin derivative – per-substituted at the 6-position with thioethers – forms an extremely stable complex with the steroidal neuromuscular blocker rocuronium bromide (\(K_a = 10^7\text{M}^{-1}\))\(^11\). However, in this case
the guest molecule is difficult to synthesize and therefore not very attractive for protein ligation.

In the previous chapter, the host-guest complex between lithocholic acid and \( \beta \)-cyclodextrin has been investigated for the formation of both, covalent and reversible, protein heterodimers in buffered solution. The host-guest complex between lithocholic acid and \( \beta \)-cyclodextrin meets all the requirements for the application in protein assembly displaying a \( K_d \) in the range of 1 \( \mu \)M\(^{12} \). Recently, Nitz et al. have published the formation of an inclusion complex between lithocholic acid and heptakis-[6-deoxy-6-(2-aminoethylsulfanyl)]-\( \beta \)-cyclodextrin (cysteamine-cyclodextrin) that exhibits a \( K_a \) of \( 5.5 \times 10^7 \) M\(^{-1} \) (\( K_d = 0.02 \) \( \mu \)M) in buffered solution at pH 7.5\(^{13} \) (Figure 1).

![Figure 1: Supramolecular protein dimerization utilizing a complex of lithocholic acid and cysteamine-modified \( \beta \)-cyclodextrin.](image)

The increase in affinity is thought to be caused by the enlarged hydrophobic cavity of the cyclodextrin derivative 2 compared to \( \beta \)-cyclodextrin. In NMR-experiments, Nitz et al. found the thiol ethers to be oriented over the hydrophobic cavity of the
cycloextrin cavity\textsuperscript{13}. This would allow them to interact with hydrophobic guests like lithocholic acid. Additionally, the thiol ethers of 2 are more hydrophobic than the primary hydroxyl groups of normal cycloextrin. Therefore the hydrophobic cavity of 2 is larger than that of unmodified β-cycloextrin allowing the formation of more stable inclusion complexes with hydrophobic guest molecules. Applying this strong complex for protein assembly would in principle allow the dimerization of proteins at lower concentrations (Figure 1) than with the β-cycloextrin-based system.

As the lithocholic acid has already been conjugated to fluorescent proteins (chapter 2), a method for the conjugation of the cysteamine-cycloextrin to the protein had to be developed. The cysteamine-cycloextrin has been studied as an inhibitor of the anthrax toxin\textsuperscript{14}, as a scaffold for glycocusters\textsuperscript{19}, as a camptothecin receptor\textsuperscript{15} and for the formation of a self-assembling fluorescent receptor\textsuperscript{16}. However, in all these examples symmetric derivatives of 2 have been used. For the ligation of the yellow fluorescent protein to the new cycloextrin variant via expressed protein ligation, the cysteamine-cycloextrin has to be mono-functionalized with a cysteine, which is described in the first part of this chapter. Subsequently, the ligation of the cysteine-cycloextrin conjugate to the protein and the purification of the ligation product based on charge differences between the ligated protein and unligated protein are discussed. Then, the new host-guest complex is applied to the dimerization of CFP and YFP which can be followed by measuring Förster resonance energy transfer (FRET) between the two fluorescent proteins. Complexes of proteins functionalized with either β-cycloextrin or cysteamine-cycloextrin are studied in parallel to investigate whether the expected increase in affinity can be observed for the new supramolecular protein dimer.

2 Synthesis of a mono-functionalized cysteamine-cycloextrin

For the ligation to protein thioesters via expressed protein ligation, heptakis-[6-deoxy-6-(2-aminoethylsufanyl)]-β-cycloextrin 2 had to be mono-functionalized with an N-terminal cysteine. The known hydrochloride salt of 2 can be prepared in two steps starting from commercially available 3\textsuperscript{13,15,17,19}. Therefore, 3 was first converted at the C6 position into a bromide in accordance with a literature procedure\textsuperscript{18}. The obtained per-6-bromo-β-cycloextrin was then treated with cysteamine hydrochloride under basic conditions to yield pure 4\textsuperscript{19} as is shown in Scheme 1. In the next step, 4 was reacted with N-Boc- and S-StBu- protected cysteine. The StBu
An improved host-guest system for supramolecular protein dimerization

protective group allows for the in situ deprotection immediately prior to the protein ligation reaction.

![Diagram](attachment:image.png)

**Scheme 1:** Synthesis of mono-functionalized aminoethylsulfanyl-β-CD in 4 steps; a) N-bromo-succinimide, PPh₃, DMF, 75 °C, 4h, 91%¹⁸, b) 2-aminoethanethiol hydrochloride, Et₃N in DMF, 48 h, rt, 89%¹⁹, c) BocCysStBu-OSu²⁰, Et₃N, DMSO, 18 h, rt, 13%, d) 95% TFA/H₂O, quantitative.

Unlike the mono-functionalization of β-cyclodextrin, which is a well studied reaction²¹, the mono-functionalization of derivatives of 4 has not been reported in the literature. It was envisaged that the selective modification of only one of the amine groups with a pre-activated cysteine derivative would be difficult, but that the slow addition of only one equivalent in dilute solution would mainly yield the desired product. The cysteine was therefore pre-activated as the hydroxysuccinimide ester²⁰.
of which one equivalent was slowly added drop-wise to a DMSO solution of 4 in the presence of excess triethylamine. As expected, the formation of mono-substituted product 5 could be observed by LC/MS, along with unreacted 2 and di- and tri-substituted cyclodextrin. Therefore, a purification method had to be developed to isolate the β-cyclodextrin bearing only one cysteine functionality. The purification took advantage of the difference in polarity of the various components. Using an optimized gradient on an analytical reversed phase HPLC system, it was possible to separate the desired mono-substituted product from the other components present, as depicted in Figure 2. The same method was applied for a preparative purification using an HPLC with a mass-triggered detector set on the charged species [M+3H]3+, [M+2H]2+ and [M+TFA+H]2+; the most abundant ions (Figure 2).

Figure 2: LC/MS chromatogram before purification, TIC, masses corresponding to mono-functionalized 5 (MONO), di- and tri functionalized cysteamine-cyclodextrin and the starting material (SM); below mass trace of the product peak at 9.7 min.
After successful isolation of mono-functionalized 5, deprotection of the Boc-group was performed in a quantitative manner using 95 % TFA/H₂O. The pure trifluoroacetate salt 6 could then be used for ligation studies with the YFP-thioester.

3 Ligation of cysteamine-cyclodextrin to mYFP

The cysteamine-cyclodextrin derivative 6 (CDNH₂) was to be ligated to the monomeric variant of YFP (mYFP) that had been used in the previous chapter (Scheme 2). The monomeric variant was chosen to allow for easy comparison of the new cyclodextrin variant with β-cyclodextrin (CD) as the affinity of this supramolecular protein dimer is determined mainly by the supramolecular host-guest interaction and not by an intrinsic protein-protein interaction²².

The mYFP-thioester was isolated as described in chapter 2. Cyclodextrin-cysteine 6 was first thiol-deprotected using an excess of TCEP in phosphate buffer and then added directly to the protein thioester. Ligation of the protein was performed overnight at room temperature using 18 equivalents of the cysteine compound 6. To remove the excess of ligand, the buffer was exchanged after the ligation to TrisHCl buffer, pH 8, containing 50 mM NaCl. This was the lowest salt concentration that did not lead to protein aggregation. In the LC/MS, small amounts of nonfunctionalized protein could be detected. For purification of the desired protein-conjugate, therefore, ion exchange chromatography was chosen. It was reasoned that the presence of the highly-charged cyclodextrin-cysteamine modification would facilitate sequestration from the starting protein thioester based on a difference in charge.
Chapter 3

Assuming the protein modification with 6 to be equal to the introduction of 6 lysine side chains, a difference in isoelectric point (pI) of 1 between the modified (pI= 6.8) and unmodified protein (pI= 5.7) was calculated.

![Figure 3](image)

**Figure 3:** SDS-PAGE, fractions of ion-exchange chromatography of YFPCDNH₂, pooled fractions 44-51.

A strong anion exchange resin was chosen for purification, eluting the protein with increasing concentrations of NaCl. As can be seen in Figure 3, the first fractions 44-51 contained only the modified protein which runs slightly higher on the gel than the unmodified protein thioester (mY). From fraction 53 on, the unmodified protein is coeluting as is evident from a second band which runs at the same height as the reference mY. In fractions 62-64 an impurity is co-eluting, which was already present in the starting thioester protein. The mass of this protein is between 50 kDa and 64 kDa which is most probably uncleaved intein-fusion protein. In Figure 4 the LC/MS spectra of the modified protein mYCDNH₂ before and after the ion-exchange chromatography are shown. After the purification step only one sharp peak is detected corresponding to the expected mass of the ligated protein, proving that the purification strategy was successful. The fact that separation of the cysteamine-cyclodextrin ligated protein from the unmodified starting material is possible, is an advantage over the previously used β-cyclodextrin system for which purification via chromatography was not possible.
An improved host-guest system for supramolecular protein dimerization

The spectra case Protein 4 Figure supramolecular free earlier ligation for monomeric direct earlier comparison MS and studies of excitation. To study the applicability of the new modified cyclodextrin 2 for supramolecular protein dimerization, the ligated protein mYCDNH₂ was mixed with monomeric CFP ligated to lithocholic acid (mCLA). The same protein had been used in earlier studies applying the normal cyclodextrin (Chapter 2). This thus allows for a direct comparison of the two cyclodextrin derivatives.

4 FRET-studies

Protein assembly was investigated using fluorescence spectroscopy. Fluorescence spectra were recorded using an excitation wavelength of 410 nm which is optimized for CFP excitation. To study the applicability of the new modified cyclodextrin 2 for supramolecular protein dimerization, the ligated protein mYCDNH₂ was mixed with monomeric CFP ligated to lithocholic acid (mCLA). The same protein had been used in earlier studies applying the normal cyclodextrin (Chapter 2). This thus allows for a direct comparison of the two cyclodextrin derivatives.

The FRET experiments were performed in phosphate buffer containing 1 mM TCEP. The TCEP was added to ensure that no disulfide bridges were formed between the free cysteines, which had been introduced into the protein construct during the ligation of the supramolecules. The two functionalized proteins were mixed in a 1 to 1 ratio and fluorescence emission spectra were recorded over one hour (Figure 5). In case a protein heterodimer was formed, FRET would occur, which is visible as an increase in YFP emission at 527 nm and a decrease in CFP emission at 475 nm. Indeed

Figure 4: LC/MS spectra of mYCDNH₂ before and after ion exchange chromatography (Total Ion Current signal). MS spectrum of the purified protein after ion exchange with its deconvoluted mass.
a strong FRET signal can be observed for the mixture containing both of the functionalized proteins: mCLA and mYCDNH₂. This strong FRET signal shows that a supramolecular protein hetero-dimer is formed by virtue of the host-guest elements. In the control experiment, mYCDNH₂ was mixed with nonfunctionalized monomeric CFP thioester protein. In this control experiment, no increase in FRET could be detected (Figure 5), showing that the cysteamine-cyclodextrin does not induce unspecific protein assembly.

During the time-course experiment shown in Figure 5, the signal at 527 nm was found to decrease slightly, while the emission of the control mixture remained constant. In the case of the previously studied system based on the β-cyclodextrin the signal has been stable over time. For those experiments, dithiothreitol (DTT) was used as a reducing agent (Chapter 2). To investigate whether the decrease of the
An improved host-guest system for supramolecular protein dimerization

signal might be caused by the modification of the cysteamine-cyclodextrin or due to the addition of TCEP to the solution, the supramolecular protein complex with the normal cyclodextrin was investigated in the same buffer containing TCEP (Figure 6). As expected, a small FRET-effect was observed for the mCLA-mYCD complex. In contrast to the mYCDNH₂ system, however, the FRET-effect was stable over the 1 h time-course of the measurement. These experiments were performed in small volume cuvettes. When the experiments were repeated in larger volume cuvettes, the spectra were more stable over time. During binding studies in capillaries using the modified cyclodextrin binding of the protein to the glass was observed. These findings suggested that the cysteamine-cyclodextrin increases the probability of the protein interacting with the glass surface leading to denaturation and a decrease in the fluorescence signal. Therefore, all later studies were performed in large volume cuvettes. Interestingly, the direct comparison of the two supramolecular protein dimers reveals a significantly higher FRET for the complex with mYCDNH₂ (Figure 5) compared to mYCD (Figure 6). This finding is in agreement with the design of the host molecule 6 which is expected to show significantly higher binding affinity for lithocholic acid compared with β-cyclodextrin. FRET is not only dependent on the distance of the chromophores, but also on their orientation. It is therefore possible that part of the observed high FRET is caused by a better orientation of the two proteins for energy transfer in the complex with mYCDNH₂. A different orientation could be caused by an additional interaction between the positively charged cysteamine-cyclodextrin and charges on the surface of the protein.

![Diagram](image)

**Figure 6:** Fluorescence emission spectra of 0.5 µM mCLA and 0.5 µM mYCD in buffer containing 1 mM TCEP. Time course experiment over 1 h, one spectrum recorded every 6 min; red graph corresponds to the last measurement.
Titration experiments were performed to compare the affinity of the modified proteins by measuring the $K_d$ with both YFP-conjugates, featuring either the normal cyclodextrin or the cysteamine-cyclodextrin. A defined quantity of mCLA was titrated with increasing amounts of the YFP-conjugate. Upon addition of the cyclodextrin-tagged YFP, protein heterodimerization occurs leading to energy transfer, which results in a decrease in the CFP emission and an increase in the YFP emission. The decrease of the donor emission at 475 nm was monitored during this experiment, while the change in donor intensity, $\Delta I$, was plotted versus the concentration of the YFP component (Figure 7). The following formula was used to fit the data points, obtaining a $K_d$ for both systems in TCEP buffer: 

$$\Delta I = I_0 - I = \frac{\Delta I_{\text{max}} c(\text{acceptor})}{K_d + c(\text{acceptor})}$$

![Figure 7: Titration of 0.05 µM mCLA with mYCDNH2 (triangle) and mYCD (square) in phosphate buffer with 1 mM TCEP and the corresponding $K_d$ values with error of the fit.](image)

Comparison between the two complexes identifies an at least 3-fold lower $K_d$ for the modified cysteamine-cyclodextrin compared with the normal cyclodextrin. By comparison, the published data for the two complexes reports a greater than 10-fold increase in affinity for lithocholic acid in favour of the cysteamine-cyclodextrin\textsuperscript{12,13}. However, there is no data available on mono-functionalized cysteamine-cyclodextrins especially concerning their binding affinity towards hydrophobic guests. It is possible, therefore, that through desymmetrization of the cysteamine-cyclodextrin, the conformation is changed so that the lithocholic acid can no longer fit as optimally as in the case of the symmetrical molecule. Alternatively, even though the lithocholic acid – cysteamine-cyclodextrin interaction is significantly stronger, the positively charged periphery of the cyclodextrin might lead to a repulsion of the protein...
lithocholic acid-CFP construct. Also, the experimental set-up, requiring a specific concentration of fluorescent protein for sufficient signal, might limit the determination of high $K_d$'s. Finally, the performed $K_d$ experiments still have to be repeated in detail. It turned out that the results are influenced by the type of reducing agent used in the buffer. For example, the $K_d$ measured for the mYCD – mCLA complex in the TCEP buffer was found to be lower than the $K_d$ measured before in buffer containing DTT. This is in line with the observed FRET effect for the 1:1 protein mixture (Figure 6) which is slightly higher than has been previously observed (Chapter 2). It seems that DTT is binding with a low affinity to $\beta$-cyclodextrin. Nevertheless, both a significantly stronger FRET effect and a decrease in the $K_d$ was observed for the lithocholic acid - cysteamine-cyclodextrin system. This strongly indicates that the supramolecular host-guest complex of lithocholic acid and cysteamine-cyclodextrin is more efficient for the dimerization of two proteins.

5 Conclusions and outlook

In this chapter, the supramolecular host-guest complex between lithocholic acid and cysteamine-cyclodextrin was applied for the dimerization of two fluorescent proteins. First, the synthesis and isolation of a mono-functionalized cysteamine-cyclodextrin was described. Even though a selective mono-functionalization of the cysteamine-cyclodextrin with cysteine turned out not to be feasible, the purification strategy using reversed phase HPLC permitted the successful isolation of the mono-cysteine conjugate. The pure cysteine-compound could be used for protein ligation. Due to the introduction of the cysteamine-cyclodextrin, the $pI$ of the ligated protein differed from that of the non-functionalized protein which allowed for the purification of the ligation mixture via anion-exchange chromatography. This is a significant advantage over the standard $\beta$-cyclodextrin-system, described in the previous chapter where it was not possible to sequester non-functionalized protein (for example, due to thioester hydrolysis) in the case of incomplete ligation. The option to purify the ligation mixture allows for applying this modification to different proteins obtaining pure, well-defined compounds for biochemical characterization.

In the final part, the first results of subsequent protein dimerization studies were discussed. The new cyclodextrin variant seemed to be more prone to aggregate or to interact with the glass of the cuvettes used than the $\beta$-cyclodextrin. However, studies of 1:1 mixtures of host- and guest-functionalized proteins showed a higher FRET in the case of the cysteamine-cyclodextrin functionalization compared to the normal
cyclodextrin. Preliminary K_d measurements suggest that the affinity of the complex bearing the cysteamine-cyclodextrin modification is indeed higher. However, the increase in affinity compared with the normal cyclodextrin was significantly lower than expected, if one compares these values to the published K_d values of the small molecules. Here, more studies must be performed, e.g. to investigate the influence of mono-functionalization on the binding capability of cysteamine-cyclodextrin toward lithocholic acid and other guest molecules. A first attempt in that direction was performed measuring thermophoresis\textsuperscript{23} with the small molecules and the lithocholic acid functionalized protein. Microscale thermophoresis is a technique that determines binding affinities by measuring the movement of molecules along a temperature gradient\textsuperscript{24}. These measurements did not lead to conclusive results mainly due to binding to the glass capillaries at high concentrations of cysteamine-cyclodextrin. The next attempt could focus on isothermal calorimetry studies of the protein complex, which would yield a more reliable K_d of the system. This would support our hypothesis that the cysteamine-cyclodextrin-lithocholic acid complex can function as a system for supramolecular protein dimerization at sub-micromolar concentrations.

6 Experimental part

6.1.1 Synthesis

Cyclodextrin compounds 5 and 6 were synthesized by Dr. Lech-Gustav Milroy.

For analytical LC/MS of proteins, reversed phase chromatography was performed using a C4-column 150x2.00 mm (Jupiter). A linear gradient was applied starting from 5 % CH_3CN/H_2O to 70 % CH_3CN over 10 min, followed by 2 min at 70 % CH_3CN and a final equilibration step to 5 % CH_3CN/H_2O until 15 min. Both solvents contained 0.1 % formic acid.

Mono-cysteine(Boc,StBu)-cysteamine-cyclodextrin 5

The cyclodextrin salt 4\textsuperscript{19} (430 mg, 0.238 mmol) was pre-weighed in an oven-dried 100 mL round-bottomed flask fitted with a magnetic stirrer (overnight at 130 °C) and dissolved in dry DMSO (8.5 mL) to form a colorless solution. With stirring, dry triethylamine (1.25 mL, 8.97 mmol, 38 eq with respect to cyclodextrin salt) was next added via plastic syringe and the reaction left to stir for 5 min (solution remains colorless). BocCysStBu-OSu\textsuperscript{25}, (96.7 mg, 0.238 mmol, 1.0 eq.), was pre-weighed in a
separate dry 25 mL round-bottomed flask (overnight at 130 °C) and dissolved in 8.5 mL of dry DMSO with manual swirling. The amino acid solution was then added drop-wise at room temperature to the stirred cyclodextrin solution using a motorized syringe pump over 4 h while maintaining a positive argon pressure. On completion of the drop-wise addition, the reaction was stirred at room temperature for 21 h 30 min under a positive argon pressure. The DMSO was then removed by heating at 45 °C (water bath) under high vacuum (oil pump, 1-2 mbar) over 6 h to form a clear faint-orange foam/viscous oil. Absolute EtOH was added (20 mL) and the crude material scratched from the wall of the flask to produce a pale-colored suspension, which was agitated further for 1 h using magnetic stirring, then carefully filtered via suction filtration and washed thoroughly with absolute EtOH (6 x 5 mL). The resulting crude skin-colored solid was transferred via spatula to a 100 mL round-bottomed flask, dissolved in 10 mL millipore water and lyophilized overnight to afford 396 mg of a crude off-white fluffy lyophilate.

The crude mono-functionalized cyclodextrin (253.4 mg; the remainder was lost during optimization of purification conditions) was purified by reverse-phase HPLC. A fresh 100 mL stock solution of 2% CH₃CN / 0.1% TFA / 98 % Millipore H₂O was first prepared and left to stir for 1 h. The crude material was then divided into 12 x 20 mg and 1 x 13.4 mg sub-samples, which were weighed in glass vials. Each sub-sample was dissolved in the CH₃CN/TFA/H₂O stock solution and filtered (first dissolve solid in 0.5 mL solvent, filter and then wash sample vessel and filter with a further 0.5 mL affording approx. 0.8 mL filtered volume; make up to 1 mL by adding a further 0.2 mL) to make a 1 mL volume ready for manual injection onto the RP-column. The samples were then purified according to the following method: Mass-triggered RP-HPLC, Alltima C18 column, 5 mL fractions; Elution conditions (CH₃CN/H₂O/0.1% TFA): 0-1 min, isocratic 2% CH₃CN; 1-10 min, linear gradient, 2-10% CH₃CN; 10-11 min, linear gradient, 10-70% CH₃CN; 11-13 min, isocratic, 70% CH₃CN; 13-14 min, linear gradient, 70-2% CH₃CN; Collect mass peaks: 613.7 [M+3H]³⁺, 920.6 [M+2H]²⁺ & 976.7 [M+TFA+H]²⁺. Mass-triggered collection between 9.5 and 10.5 min (see Figure 2). The procedure was repeated for all 13 samples (253.4 mg crude in total, see above) and the collected fractions combined and lyophilized to afford the mono-functionalized cyclodextrin, 5 (50.1 mg, 13%, white powder). MALDI-TOF C₆₈H₁₂₆N₈O₃₁S₉ calculated 1838.6, found [M+H]+ 1839.2. ¹H-NMR (400 MHz, D₂O) δ= 5.20-5.11 (m, CD), 4.38-4.32 (m, Cys), 4.20-2.80 (m), 1.53-1.40 (Boc, tBu). NMR trace is shown in Figure 8.
Mono-cysteine(StBu)-cysteamine-cyclodextrin 6

Cyclodextrin 5 was pre-weighed in a 10 mL round-bottomed flask fitted with a magnetic stirrer. 50 μL Millipore grade H₂O was added, followed by 950 μL TFA. The reaction mixture was stirred at rt for 2 h 30 min then diluted with 2 mL toluene and evaporated at 35 °C. The co-evaporation was repeated twice more followed by continuous drying for 2 h at rt, <1 mbar. The resulting solid white paste was re-dissolved in 2 mL Millipore grade H₂O and lyophilized overnight to deliver the crude cyclodextrin-derived TFA salt, 6 (10.3 mg, quantitative, white lyophilate), in sufficient purity (¹H-NMR, Figure 8) for use in subsequent protein ligation experiments. MALDI-TOF C₆₃H₁₁₈N₈O₂₉S₉ calculated 1738.6 found [M+H]⁺ 1739.6, [M+Na]⁺ 1761.5. ¹H-NMR (400 MHz, D₂O) δ= 5.18-5.09 (m, CD), 4.28-4.23 (m, Cys), 4.02-2.85 (m), 1.48-1.37 (tBu).

Figure 8: ¹H-NMR spectra of cyclodextrin compounds 5 and 6 compared to the spectrum of the symmetric cysteamine-CD.
6.1.2 Ligation to mYFP
An 8 mM solution of 6 in ligation buffer (25 mM Na-Pi, pH 7.5, 50 mM NaCl, degassed by bubbling Ar through the buffer) with 40 mM TCEP was prepared and incubated for 30 min at room temperature to remove the StBu protection group of 6. This solution was added to the protein thioester (final concentration 210 µM) together with 4-mercapto-phenylacetic acid (30 mM) obtaining a final concentration of 2.7 mM of 6. After incubation overnight at room temperature approximately 60 % (estimation from LC/MS) of the protein had reacted with the compound 6 and the remaining 40 % were still present in the thioester form. To completely modify the protein, another 5 equivalents of deprotected cyclodextrin compound were added to the protein and incubated for 5 hours. The excess of ligand, TCEP and MPAA were then removed by exchanging the buffer (20 mM TrisHCl pH 8, 50 mM NaCl) several times using centrifugal filters with a nominal molecular weight cutoff of 10 kDa. The protein solution was loaded manually onto a 1 mL HiTrap Q HP column (Q-sepharose – CH$_2$N$^+$ (CH$_3$)$_3$) and then inserted into a Aekta Prime FPLC system. After washing with 20 column volumes of buffer A (20 mM TrisHCl pH 8, 50 mM NaCl), a linear gradient with a maximal concentration of 50 % of buffer B (20 mM TrisHCl pH 8, 1 M NaCl) was applied (flow rate 1 mL/min, fraction size 0.2 mL). Elution of the protein started at a concentration of approximately 20 % B. The fractions were analysed by SDS-PAGE and LC/MS. The pure fractions were pooled and after buffer exchange to buffer A, the protein was stored in aliquots at a concentration of 6 mg/mL at -80 °C. In total 1.5 mg (60 %) of mYCDNH$_2$ were isolated. LC/MS rt= 7.8, $C_{1352}H_{2080}N_{346}O_{415}S_{17}$ calculated M= 30366, found 30369.

6.1.3 Fluorescence spectroscopy
All samples for fluorescence spectroscopy measurements were prepared under ambient conditions in quartz cuvettes. Samples were prepared in sodium phosphate buffer (25 mM sodium phosphate, 50 mM NaCl, pH 7.5) with 1 mM TCEP or 4 mM DTT. The concentration of the proteins was determined by UV/Vis spectroscopy on a NanoDrop ND-1000 spectrophotometer using the absorbance at 435 nm and $\varepsilon_{435}=32500$ M$^{-1}$cm$^{-1}$ for CFP and the absorbance at 515 nm and $\varepsilon_{515}=84000$ M$^{-1}$cm$^{-1}$ for YFP. Fluorescence data were recorded on a Varian Cary Eclipse photoluminescence spectrometer. Fluorescence emission spectra were recorded at an excitation wavelength of 410 nm. Samples of 1:1 mixtures of proteins were measured in small cuvettes using 400 µL solution and samples for titration were measured in bigger
cuvettes with 2 mL volume. Titrations were performed by adding small volumes of YFP-component to the solution with mCLA diluting it by maximal 1.5 %.

7 References

Chapter 4

CB[8] induced heterodimerization of methylviologen and naphthalene functionalized proteins

The formation of a ternary host-guest complex between cucurbit[8]uril as host, and methylviologen and methoxynaphthalene as guest elements is applied for the controlled supramolecular dimerization of two sets of fluorescent proteins. Enhanced variants, eCFP and eYFP, as well as dimerizing variants of these proteins, dCFP and dYFP, have been modified via expressed protein ligation with the supramolecular guest molecules. For these functionalized proteins, cucurbit[8]uril acts as a supramolecular inducer of protein heterodimerization. The protein assembly can be followed by measuring Förster resonance energy transfer from the CFP to the YFP. This supramolecular protein system with intrinsic fluorescence properties allows for molecular elucidation of the protein assembly process. Specifically, the detection of nonspecific protein aggregation versus specific supramolecular protein heterodimerization can be visualized, in contrast to techniques that only observe the formation of the ternary host-guest complex. The study of the dimerizing protein variants shows that the modification with methylviologen can lead to nonspecific protein aggregation in the presence of hydrophobic hot spots on the protein surface. However, cucurbit[8]uril shields the methylviologen and inhibits these unspecific interactions, subsequently allowing specific heterodimerization of different proteins.

1 Introduction
The application of synthetic supramolecular host-guest systems to biomacromolecules is very attractive for the generation of for example small, well-defined protein assemblies such as protein homo- and heterodimers. In the previous chapter, the use of host-guest chemistry to control the heterodimerization of two proteins has been demonstrated using the complex formed between a β-cyclodextrin and a lithocholic acid element, appended to two different proteins. To take the concept of supramolecular control over protein assembly via host-guest chemistry a step further, cucurbituril, as alternative host molecule, is being investigated in this chapter.

The cucurbiturils (CBs) are a series of synthetic, pumpkin-shaped molecules made from glycoluril monomers. These molecules recognize small guest molecules with high affinity (up to $10^{12} \text{ M}^{-1}$) and selectivity over other members of their own family due to their varying cavity and portal sizes. For a strong interaction with a CB host, the guest molecule(s) should contain a hydrophobic segment to fit inside the CB cavity and a positively chargeable group to interact with the carbonyl oxygens on the rim of the CB. The most commonly studied cucurbiturils are CB[6], CB[7] and CB[8], numbered as per number of glycoluril monomers the CB consists of. Whilst CB[6] and CB[7] typically complex a single guest molecule, the wider cavity of CB[8] allows it to strongly bind either one large or two smaller guest molecules which allows for example for applications based on self-sorting. The two-molecule binding system has been a focus of investigation in recent years. CB[8] is known to form stable 1:1:1 ternary complexes with an electron deficient supramolecular guest molecule such as methylviologen (MV) and an appropriate corresponding electron rich guest such as alkoxynaphthalene (Np) via a charge transfer complex of the guests inside the CB[8] cavity. One reason for the high interest in these systems is the potential redox switchability of the interaction between CB[8] and MV. CB[8] binds one MV dication and a second electron rich guest such as dihydroxynaphthalene. Upon reduction of the viologen to the radical cation, the stoichiometry changes and it binds to two MV molecules, releasing the electron rich guest. This concept has been applied for building molecular machines, reversible complexation on surfaces, the formation of redox-controllable vesicles, and the functionalization of gold nanoparticles. Complexes of CB[8] and MV have also been applied for recognition of small biological molecules such as amino acids and peptides and have recently been used for the reversible conjugation of polyethylene glycol to modified BSA.
CB[8] also forms stable 1 : 2 complexes with small molecules such as the aromatic amino acids tryptophan and phenylalanine. Urbach et al. have studied the recognition of two N-terminal aromatic peptides inside the CB[8] cavity. These complexes have a remarkably high affinity with a $K_{\text{ter}}$ of up to $10^{11}$ M$^{-2}$ for the complex with two FGG peptides$^{19}$. Expanding on this example, Dung Dang in our group has incorporated N-terminal FGG motifs into fluorescent proteins. For these genetically modified proteins, CB[8] acts as a supramolecular inducer of protein homodimerization$^{20}$. However, for the controlled formation of protein heterodimers, alternative approaches are required. Here the use of the ternary complex between two different guest, methylviologen and methoxynaphthalene (Np), and CB[8] is applied for the selective heterodimerization of two differently colored fluorescent proteins. Furthermore, the interplay of the supramolecular host-guest system with the proteins is investigated. Two sets of fluorescent protein FRET pairs allowed establishing the specific characteristics of the CB[8]-MV-Np system with respect to protein assembly and unspecific protein aggregation.

2 Design and synthesis

Scheme 1: Supramolecular induced protein dimerization by formation of a ternary complex between methoxy-naphthol and methylviologen appended proteins and CB[8].
Methoxynaphthol (Np) and methylviologen (MV) were selected as the supramolecular guest pair to form a ternary complex with CB[8] (Scheme 1). These two guest molecules were chosen as it was known that MV forms strong charge transfer complexes with naphthalene derivatives inside the CB[8] cavity with association constants up to $K_a= 10^{12}$ M$^{-2}$. Rauwald et al. compared the binding of different electron rich compounds to the complex of CB[8] and MV$^{11}$. They found that the affinities for different naphthalene derivatives such as 2-hydroxynaphthalene, 2,6-dihydroxynaphthalene and 6-methoxy-naphthalene-2-ol are comparable, varying from 6.1 x 10$^5$ M$^{-1}$ for naphthol to 5.0 x 10$^5$ M$^{-1}$ for methoxynaphthol. For the site-selective ligation of the small molecule to the protein via expressed protein ligation, the compound has to be monofunctionalized with a cysteine and to be soluble in buffered solution. Therefore the methoxynaphthol (Np) was chosen over the naphthol as it was envisaged to have better solubility under the protein ligation conditions. Compared to the dihydroxynaphthalene, the methoxyderivative was easier to selectively functionalize with one cysteine group and was therefore selected for the ligation to the protein and supramolecular protein assembly.

Two sets of cyan and yellow fluorescent protein (CFP and YFP) FRET pairs were selected as the model system for the dimerization experiments. These protein variants differ in their intrinsic tendency to dimerize due to hydrophobic mutations. The normal enhanced variants of CFP and YFP (eCFP and eYFP) only exhibit a very weak intrinsic affinity whereas the dimerizing variants (dCFP and dYFP) carry point mutations which increase their intrinsic affinity for dimerization$^{22}$. In a previous study, these protein pairs have already proven to be good model systems for studying the interplay of supramolecular-induced dimerization and protein assembly$^2$. The protein assembly process was studied using fluorescence spectroscopy monitoring Förster resonance energy transfer (FRET) from CFP to YFP. FRET only occurs when the two different proteins are in close proximity, typically via the formation of a supramolecular heterodimer, and therefore allows the assembly to be followed on the protein level, in contrast to, for example, optical techniques that visualize the host-guest assembly in the CB[8] cavity$^{18}$. This specific feature of the design allows the interplay of the supramolecular elements with the protein assembly process to be studied as well as screening of potential, unspecific interactions of the host-guest system with the proteins.

The conjugation of the proteins to the supramolecular guest elements was performed via expressed protein ligation. For this, the proteins were expressed and
CB[8] induced heterodimerization of methylviologen and naphthalene functionalized proteins

purified using the intein based pTWIN system from NEB, which allows isolation of the proteins as C-terminal thioesters\textsuperscript{23}. The proteins were additionally equipped with His- or Strep-tags for further purification steps, either before or after the ligation reaction. For the ligation to the protein thioesters, the small molecules must first be functionalized with an \textit{N}-terminal cysteine. The synthesis of the cysteine-modified viologen derivative 2 (Scheme 2, left) is based on the synthesis of the unsymmetric aminoethylviologen salt which was reported in literature\textsuperscript{24}. This is synthesized by slow addition of methyliodide to commercially available 4,4’-bipyridine to give the methylviologen, which is further reacted with 3-bromopropylamine hydrobromide to obtain the desired starting material. This was then coupled to a Boc- and StBu protected cysteine-succinimide ester\textsuperscript{25} in acetonitrile. Coupling of the succinimide ester was preferred over direct coupling to the unactivated cysteine using HBTU or comparable coupling reagents, because of the very low solubility of the starting material in DMF\textsuperscript{26}. Coupling in DMF is possible, but requires prior ion exchange as described in the literature\textsuperscript{24}. The cysteine conjugate was purified by preparative HPLC and subsequently the Boc group was removed with TFA/DCM to yield 2.

Scheme 2: Synthesis of cysteine-modified methylviologen (left) and methoxynaphthalene (right).

![Scheme 2: Synthesis of cysteine-modified methylviologen (left) and methoxynaphthalene (right).](attachment:image-url)
The naphthalene-based host molecule was synthesized in a short, 3-step synthesis from commercially available 6-methoxy-naphthalen-2-ol. First, an alcohol spacer was attached to the phenolic functionality of the naphthalene using 6-bromo-hexan-1-ol and subsequently the thus introduced primary alcohol group was coupled to a protected cysteine derivative. After purification via preparative HPLC the resulting compound was Boc-deprotected to give 5.

Scheme 3: Ligations of methylviologen (2) to dYFP- and eYFP-thioester (upper part) and methoxynaphthalene (5) to dCFP- and eCFP-thioester.

Deprotection of the cysteine thiol functionalities of 2 and 5, protected as StBu disulfide, was performed immediately prior to protein ligation by treatment of the compounds with TCEP. The as such deprotected compounds were then directly ligated to the purified thioesters of eCFP, eYFP, dCFP and dYFP (Scheme 3). For the ligation of the viologen, complete conversion of the protein thioester needed to be achieved, as separation of the non-ligated protein from the ligation product turned
out to be problematic. The conversion was checked by LC/MS. After optimization of the reaction conditions by varying the thiol additive in the ligation reaction (MESNA, mercaptopHENylacetic acid, thiophenol) the YFP-thioester variants were reacted with an excess of 2 in the presence of thiophenol for 12 hours to give complete conversion. The naphthalene-based host molecule 5 showed limited solubility in the ligation buffer. The ligations of 5 to the CFP variants were therefore performed in the presence of detergent (TritonX-114) to enhance the solubility and reaction speed. This reaction could not be driven to full completion and non-reacted protein-thioester was subsequently removed via a Triton extraction, a technique which allows separation of proteins with a hydrophobic tag from unmodified proteins27. All proteins were analysed by mass spectrometry and SDS-PAGE, confirming integrity and purity.

3 CB[8]-induced protein heterodimerization

The CB[8]-controlled protein assembly of the two protein FRET pairs was studied using fluorescence spectroscopy. Fluorescence spectra were recorded using an excitation wavelength of 410 nm, which is optimized for selective CFP excitation. The supramolecular-induced protein heterodimers were prepared by addition of CB[8] to the specific methylviologen modified YFP, followed by the addition of the naphthalene-modified CFP. A complete set of experiments, including all possible reference samples, was thus made for both protein pairs (Figure 1 to 3).

Figure 1: Fluorescence spectra of normal enhanced fluorescent protein pairs, normalized at the CFP emission maximum. Protein concentration 1 µM each, CB[8] concentration 10 µM.
The normal variants of the fluorescent proteins were first investigated (Figure 1). As envisioned in the design of the system, CB[8] indeed selectively induces the heterodimerization of MV-eYFP with Np-eCFP. This becomes clearly apparent from the increase in the YFP emission at 527 nm (blue line) which indicates the occurrence of FRET, upon the addition of CB[8]. This results in a YFP/CFP ratio (I(527 nm)/I(475 nm)) of 0.79, in contrast to a ratio of 0.51 for the non-assembled proteins. This CB[8]-induced energy transfer between the proteins is only observed in the presence of all three supramolecular components, allowing the formation of the ternary complex. In all other cases, such as for protein pairs lacking either one or two of the supramolecular units, or in the absence of CB[8], no FRET can be observed (Figure 2, left). This demonstrates that the chosen supramolecular host-guest system is specific and selectively induces supramolecular heteroassembly of two proteins.

In order to exclude that CB[8] acts as a inducer of homodimerization for the modified fluorescent proteins, homo-FRET measurements were performed with eCFPNp and eYFP. Homodimerization of two fluorescent proteins leads to a decrease of fluorescence anisotropy due to intermolecular energy transfer. This technique has been used before to study the cucurbit[8]uril-induced homodimerization of FGG-mYFP. In the case of FGG-modified proteins, the anisotropy value decreases from 0.32 without CB[8] down to 0.27 depending on the concentration of CB[8]. Here, no changes in anisotropy upon addition of CB[8] were detected for the fluorescent proteins modified with naphthalene (Np-eCFP: 0.31 and Np-eCFP CB[8]: 0.32) and methyl-viologen (MV-eYFP: 0.32 and MV-eYFP CB[8]: 0.33). This shows that no homodimerization upon addition of CB[8] occurs.
The other pair of proteins under study, **MV-dYFP** and **Np-dCFP**, carries specific point mutations\(^{22}\) which increase their intrinsic affinity for dimerization (both homo- and heterodimerization) due to increased hydrophobic interactions. In general, protein-protein interactions are frequently mediated to a significant extent by hydrophobic and charged regions on their surfaces\(^{29}\), also allowing for recognition with designed supramolecular receptors\(^{30}\). As such, knowledge on the interplay of the supramolecular elements with these protein motifs is highly important. The weakly dimerizing fluorescent model proteins **MV-dYFP** and **Np-dCFP** are an ideal system to study this. Control experiments performed on the **Np-dCFP** with unfunctionalized dYFP and CB[8] showed that these supramolecular elements do not lead to unspecific binding or protein interactions (Figure 2, right). In contrast, experiments with **MV-dYFP** showed that the methylviologen can lead to unspecific protein assembly of the weakly dimerizing proteins (Figure 3, green and black lines). Protein mixtures of **MV-dYFP** with either **Np-dCFP** or unmodified dCFP feature an increase of the YFP emission in the absence of CB[8]. Interestingly, however, upon the addition of CB[8] the unspecific protein assembly induced by the methylviologen is inhibited, as becomes apparent from the absence of YFP emission (Figure 3 red line). The CB[8] thus shields the methylviologen from making unspecific interactions with hydrophobic protein surfaces. As a result, in the **dYFP-MV - Np-dCFP** mixture (Figure 3 blue line), a specific CB[8] mediated energy transfer is observed, resulting from the supramolecular protein dimer. A comparison of the FRET ratio I(527nm)/I(475nm) of the supramolecular complex consisting of the dimerizing variants (0.69) to the one of the enhanced variants (0.79) shows a higher FRET for the proteins with a weaker intrinsic dimerizing affinity. This finding is opposite to what was observed before with the supramolecular host-guest complex between β-cyclodextrin and lithocholic acid where a significantly higher FRET ratio was observed for the dimerizing variants\(^2\). One possible explanation for this finding might be the affinity of the complex between CB[8] and MV for phenylalanine (K\(_a\)= 5 \times 10^3 M\(^{-1}\))\(^{16}\). The dimerizing variants of the fluorescent proteins contain the mutation S208F. It would therefore be possible that an intramolecular complex between the MV x CB[8] x phenylalanine prevents CFPNp binding to a certain extent. Also an intermolecular complex between two proteins via this mechanism could be envisioned but the affinity to phenylalanine in that case is significantly lower than the affinity to the naphthalene derivative. Another explanation might be that the proteins have a different orientation in the complex consisting of the dimerizing proteins which can lead to less efficient energy transfer.
Figure 3: Fluorescence spectra of dimerizing fluorescent protein pairs, normalized at the CFP emission maximum. Protein concentration 1 µM each, CB[8] concentration 10 µM.

The results described above show that the formation of a ternary supramolecular protein dimerization complex involves different molecular processes than the formation of the β-cyclodextrin and lithocholic acid host-guest complex, which leaves room for further studies. However, it was shown that the ternary system of CB[8] with MV and Np can definitely be successfully used for the formation of selective protein heterodimers of more hydrophobic proteins. In these cases the presence of CB[8] as host molecule is required to prevent MV-induced unspecific dimerization with hydrophobic protein surfaces. The formation of intramolecular complexes must be taken into account when studying such systems.

4 Conclusions and outlook

In conclusion, it was shown that CB[8] constitutes an attractive supramolecular inducer of protein heterodimerization using MV and Np as protein-appended host elements. The model system of fluorescent protein pairs allows specific visualization of the protein dimerization event instead of the formation of the supramolecular ternary complex only. This in turn allows for detection of potential unspecific interplay of the supramolecular elements with the proteins and provides molecular insights not attainable using experiments solely based on the observation of the charge-transfer complex. This is especially important as the modification with MV can lead to unspecific interactions in the absence of CB[8] and in the presence of the host molecule, potentially bind to aromatic amino acids on the surface of the protein. A variation of the linker length on the protein could modulate such intramolecular binding events. Keeping these findings in mind, this novel system for controlled
protein heteroassembly can be applied, for example, for targeting of N-terminal tryptophan motifs in proteins. N-terminal tryptophan has an affinity for the CB[8] x MV complex and would therefore be a good genetically introducible tag for protein heterodimerization. Furthermore this system is attractive for application in the area of bionanotechnology, for example, for the controlled immobilization of proteins on surfaces, as it potentially allows for obtaining electrochemically switchable protein platforms. In line with this, Jonkheijm and colleagues have recently shown that naphthol-modified fluorescent proteins can be immobilized on a surface functionalized with methylviologen by formation of a ternary complex with CB[8]. CB[8] controlled protein assembly thus constitutes an attractive platform for applications both in solution and on surfaces.

5 Experimental part

5.1.1 Construction of plasmids, protein expression and purification

Plasmids were constructed by Dr. Hoang Nguyen. To construct plasmids, pHT477 (His-ECFP-Intein-CBD-Strep) and pHT482 (Strep-EYFP-Intein-CBD-His), DNA encoding for Strep-tag and His-tag were introduced to the fusion protein at the C-terminus by applying a point mutagenesis method using two pairs of oligonucleotides for Strep-tag, ON061, 5’-gcaagcttgagccaccgcagttcggagaaagttgactgcaggaagggatccgg-3’ and ON062, 5’-cttttcgactgctgggtggctcgctttgaagctgccacaaggcaggaac-3’ and for His-tag, ON059, 5’-gaagccgagccaccatcaccatcactaccattgactgcaggaagggatccgg-3’ and ON060, 5’-atggtgatgtgtgatgtgtgatgtgtgctcgttttaagaagctgccacaaggcaggaac-3’ with pHT409 and pHT423 as templates, respectively. Plasmids, pHT409 (His-ECFP-Intein-CBD) and pHT423 (Strep-EYFP-Intein-CBD) were constructed by inserting two pairs of complementary oligonucleotides for His-tag, ON031, 5’-tatggaagccgagccaccatcaccatcactaccattc-3’ and ON032, 5’-tagaattgtgatgtgtgatgtgtgctcgttttaagaagctgccacaaggcaggaac-3’ and for Strep-tag, ON047, 5’-tagagccactttgcgggtgtgctcagttcaccata-3’ and ON048, 5’-tagagccactttgcgggtgtgctcagttcaccata-3’ into plasmids, pTWIN1-ECFP and pTWIN1-EYFP at Ndel. Plasmids pHT477 and pHT482 were transformed into E.coli Rosetta 2 (DE3) cells and transformants were selected on Chloramphenicol (34 mg/L) and Ampicillin (125 mg/L) containing Agar plates. Protein expression and purification were performed as described in chapter 2, leading to MESNA-protein thioesters of eCFP (from pHT477) and eYFP (from pHT482). LCMS: eYFP calculated: 28538, found: 28539; eCFP calculated: 29898, found: 29896. Expression and purification of the dimerizing variants, dYFP and dCFP has been described in chapter 2.
5.1.2 Chemical synthesis of cysteine-modified guest molecules

Synthesis and characterization of naphthalene derivatives 3, 4 and 5 has been carried out by Dr. Jacqui Young.

4-(1-{3-[2-amino-3-(tert-butyldisulfanyl)propanamido]propyl}pyridin-1-ium-4-yl)-1-methylpyridin-1-ium 1

In a round bottom flask, a solution of 1-(3-aminopropyl)-1'-methyl-[4,4'-bipyridine]-1,1'-diium\(^2\) (200 mg, 0.46 mmol) in 50 mL acetonitrile was prepared and DIPEA (90 mg, 0.69 mmol) was added. The hydroxysuccinimide ester of the Boc and StBu-protected cysteine\(^2\) (214 mg, 0.55 mmol) was added to the mixture and stirred for two hours at 40 °C. The solvent was evaporated and 100 mg of the crude mixture was dissolved in H\(_2\)O and purified via preparative HPLC to isolate 15 mg (estimated yield: 25%) of compound 1. \(^1\)H-NMR (400 MHz, CD\(_3\)OD) \(\delta\) (ppm): 1.34 (s, 9H, SC(CH\(_3\))\(_3\)), 1.47 (s, 9H, OC(CH\(_3\))\(_3\)), 2.31 (m, 2H, H2), 2.88-3.16 (m, 2H, CH\(_2\)S), 3.31-3.29 (m, 2H, H3), 4.24 (m, 1H, CH), 4.53 (s, 3H, CH\(_3\)), 4.77 (m, 2H, H1), 8.49-8.72 (m, 4H, ArCH), 9.23 (m, 4H, ArCH); LC/MS calculated for C\(_{26}\)H\(_{40}\)N\(_4\)O\(_3\)S\(_2\): 520.25 (M)+, found 519.3, 520.3 (M)+.

4-(1-{3-[2-amino-3-sulfanylpropanamido]propyl}pyridin-1-ium-4-yl)-1-methylpyridin-1-ium 2

Compound 1 (3 mg, 5 \(\mu\)mol) was dissolved in 1 mL of 30 % TFA in dichloromethane (v/v). The mixture was stirred for three hours. After co-evaporation with toluene, compound 2 was obtained as an orange solid in quantitative yield. \(^1\)H NMR (400 MHz, CD\(_3\)OD) \(\delta\) = 1.37 (s, 9H, SC(CH\(_3\))\(_3\)), 2.42 – 2.24 (m, 2H, H2), 3.29 - 3.09 (m, 2H, CH\(_2\)S), 3.35 - 3.50 (m, 2H, H3), 4.14 (m, 1H, CH), 4.53 (s, 3H, CH\(_3\)), 4.81 (m, 2H, H1), 8.55 –
8.78 (m, 4H, ArCH), 9.11 – 9.39 (m, 4H, ArCH). $^{13}$C NMR (100 MHz, CD$_3$OD) $\delta$ = 30.08, SC(CH$_3$)$_3$; 32.32, C2; 37.08, C3; 42.12, CH$_2$S; 54.01, CH; 60.67, C1; 127.85, 128.23, 147.36, 148.02, ArCH. LC/MS calculated for C$_{21}$H$_{32}$N$_4$OS$_2$: 420.20 (M)$^+$, found 420.1 (M)$^+$.

6-[(6-methoxy-2-naphthyl)oxy]hexan-1-ol 3

NaH (110 mg, 2.8 mmol) was added to a cold (0 °C) solution of 6-methoxy-naphthalen-2-ol (400 mg, 2.3 mmol) in DMF (6 mL) and stirred at 0 °C for 30 min then at room temperature for a further 30 min. A solution of 6-bromo-hexan-1-ol (250 µL, 1.9 mmol) in DMF (6 mL) was slowly added to the solution (over 15 min). The solution was stirred at room temperature under an argon atmosphere for 5 h then solvent removed in vacuo to give the crude product as a green solid. This was purified by column chromatography (SiO$_2$, ethyl acetate: cyclohexane 1:1) to give the pure product as a white solid (481 mg, 92 %). $^1$H NMR (CDCl$_3$, 400 MHz): 1.43-1.58 (m, 4H, H3, H4); 1.62 (m, 2H, H2); 1.85 (m, 2H, H5); 3.66 (t, $J$ = 6.6 Hz, 2H, H1); 3.89 (s, 3H, OCH$_3$); 4.05 (t, $J$ = 6.5 Hz, 2H, H6); 7.10 (m, 2H, H1', H5'); 7.13 (dt, $J$ = 8.8, 2.4 Hz, 2H, H3', H7'); 7.63 (dd, $J$ = 8.9, 5.2 Hz, 2H, H4', H8'). $^{13}$C NMR (CDCl$_3$, 100 MHz): 25.75, 26.14, C3, C4; 29.45, C5; 32.86 C2; 55.47, OCH$_3$; 63.03, C1; 68.09, C6; 106.27, 107.19, C1', C5'; 119.01, 119.38, C3', C7'; 128.25, 128.28, C4', C8'; 129.84, 129.97, C9', C10'; 155.73, 156.22, C2', C6'. LC/MS: m/z: 275.07 (M + H)$^+$. 

6[(6-methoxy-2-naphthyl)oxy]hexyl N-Boc-3-(t-butyl-disulfanyl)alaninate 4

Boc-Cys(StBu)-COOH (113 mg, 0.37 mmol) and EDC.HCl (70 mg, 0.37 mmol) were dissolved in DMF (4 mL) and stirred at room temperature for 15 min. This solution was then added to a stirring solution of 3 (25 mg, 0.09 mmol) in DMF (1 mL). DMAP (4.5 mg, 0.04 mmol) was added to the solution and the resulting solution was stirred
at room temperature under an argon balloon for 6.5 h. The solvent was removed in vacuo to give a yellow oil, which was redissolved in CH₂Cl₂. Water was added to the solution and the product was extracted twice with fresh CH₂Cl₂. The organic layers were combined, dried over Na₂SO₄, filtered and dried in vacuo to give the crude product as an orange oil. Following column chromatography (SiO₂, ethyl acetate: cyclohexane 1:4), the desired product 4 was obtained as a clear oil (31 mg, 60 %). ¹H NMR (CDCl₃, 400 MHz): 1.31 (s, 9H, StBu); 1.45 (s, 9H, C(CH₃)₃); 1.46-1.58 (m, 4H, H₃, H₄); 1.72 (m, 2H, H₂ or H₅); 1.85 (m, 2H, H₂ or H₅); 3.16 (m, 2H, CH₂S); 3.89 (s, 3H, OCH₃); 4.05 (t, J = 6.4 Hz, 2H, H₁ or H₆); 4.19 (t, J = 6.7 Hz, 2H, H₁ or H₆); 4.57 (m, 1H, CH); 5.38 (t, J = 7.9 Hz, 1H, NH); 7.09-7.13 (m, 4H, H₁', H₃', H₅', H₇'); 7.63 (dd, J = 8.9, 5.4 Hz, 2H, H₄', H₈'). ¹³C NMR (CDCl₃, 100 MHz): 25.95, 26.02, C₃, C₄; 28.54, C(CH₃)₃; 28.67, 29.39, C₂, C₅; 29.97, C(CH₃)₃; 43.18, CH₂S; 48.33, SC(CH₃)₃; 55.61, CH, 55.51, OCH₃; 65.92, 68.02, C₁, C₆; 80.26, OCH(C(CH₃)₃); 106.29, 107.21, C₁', C₅'; 119.05, 119.40, C₃', C₇'; 128.28, 128.31, C₄', C₈'; 129.88, 129.99, C₉', C₁₀'; 155.23, HNCOO; 155.73, 156.26, C₂', C₆'; 170.99, CH₂COO. LC/MS: calculated for C₂₉H₄₃NO₆S₂Na: 588.2424 (M+Na)⁺; found: 588.13 (M + Na)⁺.

6-[(6-methoxy-2-naphthyl)oxy]hexyl 3-(t-butyl-disulfanyl)alaninate 5

4 (14.8 mg, 26.2 µmol) was dissolved in a mixture of TFA:CH₂Cl₂ (1:4, 500 µL) and the solution stirred at room temperature for 3 h. The solvent was removed in vacuo to give the product as a brown oil (17.6 mg, quantitative). ¹H NMR (CDCl₃, 500 MHz): 1.32 (s, 9H, tBu); 1.43 (m, 2H, H₃ or H₄); 1.53 (m, 2H, H₃ or H₄); 1.72 (m, 2H, H₂ or H₅); 1.83 (m, 2H, H₂ or H₅); 3.20 (dd, J = 14.7, 7.2 Hz, 1H, CH₂S); 3.30 (dd, J = 14.6, 3.5 Hz, 1H, CH₂S); 3.89 (s, 3H, OCH₃); 4.03 (t, J = 6.3 Hz, 2H, H₁ or H₆); 4.25 (t, J = 6.6 Hz, 2H, H₁ or H₆); 4.42 (m, 1H, CH); 7.08-7.12 (m, 4H, H₁', H₃', H₅', H₇'); 7.62 (t, J = 8.4 Hz, 2H, H₄', H₈'); 7.70 (br, 3H, NH₃⁺). ¹³C NMR (CDCl₃, 125 MHz): 25.75, 25.95, C₃, C₄; 28.38, 29.33, C₂, C₅; 29.89, C(CH₃)₃; 39.10, CH₂S; 49.34, SC(CH₃)₃; 52.96, CH, 55.53, OCH₃; 67.67, 67.97, C₁, C₆; 106.35, 107.28, C₁', C₅'; 119.09, 119.37, C₃', C₇'; 128.33, C₄', C₈'; 129.94, 130.01, C₉', C₁₀'; 155.71, 156.30, C₂', C₆'; 168.16, C=O. LC/MS: calculated for C₂₉H₃₆NO₆S₂: 466.2080 (M+H)⁺; found: 466.20 (M + H)⁺.
5.1.3  Expressed protein ligation

Ligation of the naphthalene compound has been carried out by Dr. Jacqui Young and Dr. Hoang Nguyen.

**Np-eCFP and Np-dCFP:** Compound 5 (5 mg, 11 µmol) was dissolved in methanol (200 µL) and StBu-deprotected with tris(2-carboxyethyl)phosphine (TCEP) at mole ratios of 1:1.1 for 1 h at room temperature. TritonX-114 (100 µL of a 10% solution) and Tris.HCl buffer (pH = 8.5, 50 µL) were slowly added and the mixture allowed to sit at ambient temperature for 1.5 h. After this time, the mixture was slowly added to the CFP thioester (from above, 580 µL, 425 µM, ie. at a mole ratio of 1:45 (CFP: 5)). NaCl and MESNA were added to final concentrations of 500 mM and 200 mM respectively. The ligation mixture was kept at room temperature away from light overnight. Excess 5 was removed by size-exclusion chromatography and the remaining mixture of proteins was purified by triton exchange\(^2^7\). The triton was removed using a Ni-NTA-column and ligated protein eluted using imidazole. The buffer was exchanged to storage buffer (25 mM Sodium-Phosphate pH 7.5, 100 mM NaCl) and the protein was stored in small aliquots at low (25-40 µM) concentration at -80°C. The ligated protein was analysed by SDS-PAGE and LC/MS: Calculated for Np-dCFP: 30208, found: 30208; calculated for Np-eCFP: 30133, found 30131.

**MV-eYFP and MV-dYFP:** Compound 2 (10 mg, 15 µmol) was dissolved in a mixture of 60 µL methanol and 240 µL ligation buffer (sodium phosphate 25 mM, NaCl 50 mM, pH 7.5). For StBu deprotection, 60 µL of a 0.5 M solution of TCEP was added and the mixture was incubated for 30 min at room temperature. The mixture was then added to the YFP thioester (200 µM) at a molar ratio of 1:20 (YFP : 2) and thiophenol was added to a final concentration of 300 mM. The ligation mixture was incubated on a rotating wheel in the dark at room temperature overnight. The mixture was then centrifuged to remove any precipitate. To remove excess 2 and thiophenol, the buffer was exchanged to ligation buffer using centrifugal filters with a molecular weight cutoff of 10 kDa. LC/MS: MV-eYFP calculated: 28745, found: 28751; MV-dYFP calculated: 28828, found: 28822.
5.1.4 Fluorescence spectroscopy

All samples for fluorescence spectroscopy measurements were prepared under ambient conditions in quartz cuvettes. Samples were prepared in sodium phosphate buffer (25 mM sodium phosphate, 50 mM NaCl, pH 7.5) with 10 µM TCEP. The concentration of the proteins was determined by UV/Vis spectroscopy on a NanoDrop ND-1000 spectrophotometer using the absorbance at 435 nm and $\varepsilon_{435} = 32500$ M$^{-1}$cm$^{-1}$ for CFP and the absorbance at 515 nm and $\varepsilon_{515} = 84000$ M$^{-1}$cm$^{-1}$ for YFP. Fluorescence data were recorded on a Varian Cary Eclipse photoluminescence spectrometer. Fluorescence emission spectra were recorded at an excitation wavelength of 410 nm. Samples were prepared by premixing the YFP compound with a stock solution of CB[8] in water (50 µM) to a final concentration of 2 µM YFP and 20 µM CB[8]. This solution was mixed with a 2 µM solution of CFP compound to yield final protein concentrations of 1 µM each and after a short incubation time, the fluorescence spectrum was recorded. To ensure the data were comparable, all samples were prepared in the same way.

5.1.5 Fluorescence anisotropy measurements

Fluorescence anisotropy data were recorded on a Varian Cary Eclipse photoluminescence spectrometer. YFP was excited at 500 nm and emission was recorded from 524 nm to 532 nm. CFP was excited at 435 nm and emission was recorded from 472 nm to 478 nm. Anisotropy values were averaged over each wavelength.
6 References

Chapter 4

Chapter 5

Labeling of SNAP-tag fusion proteins with supramolecular ligands

Methods to selectively and efficiently label proteins with supramolecular host-guest molecules are crucial to establish supramolecular chemistry as a tool to control protein assembly. In the previous chapters the labeling of proteins was achieved using expressed protein ligation. In this chapter the use of the SNAP-tag is investigated for the labeling of fluorescent proteins with different supramolecular ligands. The SNAP-tag technology requires fusion of the protein of interest with the enzyme alkylguaninetransferase which reacts with molecules carrying an O6-benzylguanine modification. In the following, the conjugation of benzylguanine to lithocholic acid, β-cyclodextrin and a bipyridine discotic and their subsequent ligation to fluorescent proteins will be discussed. The chemical labeling of proteins with supramolecular elements via the SNAP-tag approach is shown to be rapid and selective, providing new chemical approaches to generate supramolecular protein constructs.
1 Introduction

Methods to label proteins with supramolecular ligands in a controlled manner are of great importance as can be seen from examples in the recent literature where proteins have been modified with supramolecular elements\(^1\). This protein labeling is often done via random tagging of lysine side chains, thus not providing site selective and controlled protein modification\(^2\). In the previous chapters the functionalization of proteins with different host or guest molecules via expressed protein ligation was discussed. This approach was shown to allow for selective modification at the C-terminus of a protein-thioester with almost any supramolecular element that is carrying an N-terminal cysteine. However there are certain limitations for the application of expressed protein ligation. For example the frequent need for thiol additives to the ligation mixtures limits its use to proteins that are stable in the presence of reducing agents. A further drawback is that often a large excess of labeling ligand and overnight reaction times are required to achieve complete protein functionalization. In this chapter the use of the SNAP-tag technology\(^3\) for the labeling with different supramolecular structures will be discussed as an alternative approach for the rapid, easy accessible and controllable labeling of proteins with supramolecular elements.

The SNAP-tag is a commercially available genetically encoded protein tag that has been developed in the lab of K. Johnsson for the labeling of proteins with different small molecules, both in buffered solution as well as in bacteria, yeast and mammalian cells\(^4\). The 21 kDa SNAP-tag is a mutant of the human DNA repair enzyme alkylguaninetransferase\(^5\). It has been engineered to react specifically with \(\text{O}^6\)-benzylguanine (BG) derivatives carrying a ligand of interest (see Scheme 1, top). The benzyl group of the BG derivative is transferred to a cysteine residue in the active site of the enzyme to covalently label the protein of interest with the ligand. The enzymatic labeling reaction has an efficiency of 100 % for the labeling of proteins in buffer and proceeds relatively fast\(^6\) with a rate of \(3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}\). \(^7\) Over the last years, the SNAP-tag has found wide-spread application in different research labs. The labeling of purified SNAP-tag fusion proteins has been used for example to label antibody fragments\(^8\) and to design a fluorescent sensor protein which allows determining changes of metabolite concentration by measuring FRET efficiency\(^9\).

\(^*\) The rate constant for a native chemical ligation between a peptide-alanyl-thioester and a cysteine peptide in the presence of 4-mercapto phenylacetic acid is \(8.7 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}\).
Purified fusion proteins have also been immobilized on different BG modified surfaces\textsuperscript{10,11}. These applications show the versatility and efficiency of the tag for labeling purified fusion proteins. However a great advantage of the SNAP-tag is the possibility to use it also for intracellular and cell surface labeling\textsuperscript{12}. The fusion of targeting sequences allows localization of the SNAP-tag protein in the cellular compartment of interest, which has been used for sensors that can detect varying concentrations of metal ions such as Zn\textsuperscript{2+} \textsuperscript{13} or Ca\textsuperscript{2+} \textsuperscript{14}. The addition of BG dimers to living cells leads to the formation of covalent protein dimers if the proteins are located in the same cellular compartment and can as such be used to probe spatial proximity of proteins in the cell\textsuperscript{15}. These examples show that the labeling of proteins with the SNAP-tag allows for applications in buffered solution, on surfaces as well as potentially in the cell. It is therefore an attractive method to use for the labeling of proteins with supramolecular ligands.

In this chapter the exploration of the SNAP-tag technology for the labeling of two fluorescent model proteins, eCFP and eYFP with three different supramolecular
elements (Scheme 1) will be described. These molecules have been chosen because of their different properties concerning solubility and aggregation behavior in buffered solution to explore the compatibility of the labeling reaction with different supramolecular structures. First the synthesis of the necessary BG conjugates will be shown and in the second part the labeling of the SNAP-tag fusion proteins with the supramolecular elements will be discussed.

2 Synthesis of benzylguanine conjugates

2.1 Synthesis of reactive benzylguanines

To apply the SNAP-tag labeling to supramolecular structures, the synthesis of $O^6$-benzylguanine-conjugates as depicted in Scheme 2, is required.

Scheme 2: Synthetic strategies for coupling different ligands to three benzylguanine building blocks.

For the conjugation of the supramolecular ligands to the benzylguanine moiety, three different approaches were chosen. The first strategy is based on azide-alkyne click chemistry and starts with benzylguanine 4 that can react with an azide carrying supramolecular ligand. The second strategy is based on amide coupling of $O^6$-(4-aminomethyl-benzyl)-guanine 5 to a carboxylic acid on the supramolecular ligand which has preferably been preactivated as NHS ester. The last method is based on
amide coupling as well but requires an amine group on the supramolecular ligand which can then be reacted with benzylguanaine 6 using coupling reagents like HBTU.

The synthesis of the benzylguanaine building blocks 4, 5, and 6 has been described in literature before. The synthesis of O₆-substituted guanines starts from 6-chloro-guanine 10 which is activated as quaternary amine to have a better leaving group for the following substitution with a benzyl alcohol such as 9 (Scheme 3). The literature procedure for compound 5 starts with activating the 6-chloro-guanine with 1-methylpyrrolidine, reporting yields between 66% and 79%. Here, a modified protocol was chosen using 1,4-diazabicyclo[2.2.2]octane (DABCO) which gave higher yields in this synthesis. The product 11 precipitates from solution and after washing with acetone it can be used without further purification. The fluorescence of the DABCO-purine allows for easy detection of formation and conversion of 11 by TLC, which was another reason to choose this activating method. For the synthesis of alkyne modified benzylguanaine 4 (Scheme 2) the literature protocol could be followed using the DABCO purine instead of the 1-methylpyrrolidine activated guanine. For the synthesis of aminomethyl-benzylguanaine 5 the literature route was slightly modified to the route depicted in Scheme 3.

![Scheme 3: Synthesis of aminobenzylguanaine 5.](image)

The aminomethyl-benzylalcohol 8 was synthesized starting from 4-cyano-benzaldehyde via reduction with 1M LiAlH₄ solution in THF. The crude product, containing residual salts, was used without further purification for the protection of the amine as trifluoroacetamide following a protocol from Keppler et al. For the
subsequent coupling of 9 to an activated-guanine, two methods were described in the literature: Either deprotonation of the alcohol using potassium tert-butoxide\textsuperscript{4} or sodium hydride in DMSO\textsuperscript{18}. Here, the reaction in DMSO using sodium hydride and dimethylaminopyridine (DMAP) as a catalyst worked best. The reaction was then quenched with a mixture of acetic acid in water which prevented hydrolysis of the trifluoroacetamide during the workup. The product 12 could be purified via normal phase column chromatography and was deprotected in the next step using methylamine yielding pure building block 5.\textsuperscript{18} This building block could then directly be used to couple to NHS-activated supramolecular ligands and as starting material for the synthesis of carboxylic acid functionalized benzylguanine building block 6. This was obtained by reacting 5 with glutaric anhydride\textsuperscript{15}. With these building blocks in hand, different supramolecular ligands could now be coupled via amide coupling or azide-alkyne click chemistry.

2.2 Conjugation of lithocholic acid

The first supramolecular element that was conjugated to a benzylguanine was lithocholic acid. Lithocholic acid is a supramolecular guest that forms hydrophobic inclusion complexes with β-cyclodextrin in water. Previously, it has been conjugated to fluorescent proteins using expressed protein ligation (Chapter 2). We chose it therefore as model compound for a hydrophobic guest molecule that would allow the properties of hydrophobic benzylguanine conjugates to be studied during the synthesis and the ligation to the fluorescent proteins.

From previous work it was known that modification at the C3-hydroxyl group does not affect the affinity to the host molecule β-cyclodextrin\textsuperscript{20}. Two different conjugates, 16 and 19, varying in the linker length between the steroid and the benzylguanine were synthesized (Scheme 4). In both cases N-hydroxysuccinimide ester derivatives of 1 were coupled to amine-benzylguanine 5. To avoid side reactions of the free carboxylic acid of lithocholic acid 1 during the synthesis, the carboxylic acid was protected as a methyl ester first (13). Then the succinimide ester of the C3-hydroxyl was formed using disuccinimidylcarbonate (DSC) in a mixture of triethylamine, chloroform and acetonitrile as described before\textsuperscript{21}. For the synthesis of the derivative without the linker, benzylguanine 5 was directly reacted with 14 in DMSO using triethylamine as base. The crude product was purified via column chromatography and in the last step deprotected using lithiumhydroxide in a mixture of water and methanol yielding lithocholic acid conjugate 16.
Analogously, the second lithocholic acid derivative 19 with a short ethyleneglycol spacer was prepared. Succinimide ester 14 was reacted with 4,7,10-trioxatridecane-1,13-diamine to compound 17 which was again activated as a succinimide ester using DSC and directly used in the next step. Coupling of the NHS-ester to aminomethyl-
benzylguanine 5 was performed in DMF in the presence of triethylamine followed by column chromatography to yield pure 18. The deprotection of the methylester was done as described for the previous lithocholic acid conjugate.

The successful synthesis and purification of the two lithocholic acid derivatives shows that coupling of benzylguanine 5 to activated NHS esters in DMSO can be applied for the conjugation of hydrophobic guest molecules. The choice of the high boiling solvent DMSO is required to dissolve the aminomethyl-benzylguanine 5 which is only poorly soluble in most organic solvents. Purification of the conjugates was possible via normal phase silica chromatography after adsorbing the crude products on silica gel or another inert carrier substance. This method should be adaptable for other hydrophobic guest molecules. In the next step, the behavior of the two different derivatives in the labeling of the proteins is to be investigated.

2.3 Conjugation to β-cyclodextrin

The second supramolecular element that was conjugated to benzylguanine was β-cyclodextrin. β-cyclodextrin is a supramolecular host-molecule that forms inclusion complexes with various hydrophobic small molecules in aqueous solution. The hydrophobic inclusion complex with lithocholic acid has already been investigated (chapter 2). Therefore it was chosen as a model compound for a water soluble host that was functionalized via azide-alkyne click chemistry with building block 4.

The synthesis of monofunctionalized β-cyclodextrin (CD) started from monotosylated CD 20 which was synthesized from CD 2 and tosylimidazole (Scheme 5). The mono-tosylated CD was then reacted with an excess of sodium azide to yield a mixture of CD 2 and mono-azido-β-cyclodextrin 21 following a literature procedure. The separation of the hydrolyzed starting material 2 and the product 21 was not successful and the mixture was directly used in the next step. For the click reaction between the azide-CD and the alkyne-benzylguanine 4, DMF was chosen as a solvent to dissolve the benzylguanaine. There are examples in the literature where CuI-catalyzed click-chemistry has been used to functionalize cyclodextrins in DMF. Copper sulfate was used here in stoichiometric amounts as CuII salts are known to associate with the secondary face of CD. After reacting for two weeks, formation of the product was observed in the MALDI-ToF spectrum and the reaction was stopped. Reasons for the slow conversion might be the low solubility of the reactants, complexation of the copper catalyst and the benzylguanaine as well as the impure starting material containing CD 2. The crude product was then isolated by
precipitation in acetone. Purification via normal phase silica column was not successful but the product could be isolated in pure form via reversed phase HPLC. Even though the reaction conditions and especially the synthesis of the starting material can still be optimized, an easy way of purifying BG-modified CD was thus found. Analysis of the obtained CD product by NMR turned out to be difficult, which is most probably due to partial complexation of the BG in the cavity of the CD. It should be noted that such supramolecular shielding of the BG might have an influence on the ligation to the fluorescent proteins which will be described later.

Scheme 5: Synthesis of a benzylguanine-cyclodextrin conjugate via click chemistry.

2.4 Conjugation to bipyridine-discotic

As a last example for the conjugation of BG to a supramolecular element, the discotic molecule 23, depicted in Scheme 6, was selected. The discotic monomers 3 are known to reversibly assemble into columnar supramolecular polymers at low concentrations in water. The attachment of ethylene glycol chains makes them suitable for biological applications as they ensure water solubility and reduce unspecific interaction with, for example, proteins. It has already been shown that proteins can assemble along these supramolecular columns in buffered solution. Therefore discotic molecules of structure 3 were chosen as a model system for a
supramolecular polymer that was to be labeled with proteins via the SNAP-tag technology.

Previous work on protein assembly was done with a $C_3$-symmetric discotic molecule, where each discotic carried three ligands for protein binding\textsuperscript{27}. Here, the $C_2$-symmetric discotic 23 carrying one amine group for functionalization was selected (Scheme 6). Reducing the number of protein ligation sites from three to one should facilitate the protein ligation and prevent formation of complex mixtures.

**Scheme 6: Synthesis of a benzylguanine discotic conjugate.**

Discotic 23 was synthesized by Katja Petkau\textsuperscript{28} in our group and could directly be coupled to benzylguanine building block 6 in dimethylacetamide (DMA) using HBTU and DIPEA (Scheme 6). The high boiling solvent DMA was necessary to dissolve compound 6. For a complete conversion of 23, an excess of BG 6 was used which resulted in a mixture of product 24 and benzylguanine building block 6. Purification of the mixture was necessary as both the starting material and the product, could react
in the subsequent labeling reaction with the protein. Making use of the difference in molecular weight, the product was separated from the BG starting material via size exclusion chromatography in DMF. This material could then be used for reacting proteins with a supramolecular polymer.

3 Labeling of AGT- fusion proteins with supramolecular elements

3.1 Purification of AGT fusion proteins

The fluorescent protein FRET pair consisting of eCFP and eYFP was chosen as a model system to study the applicability of the SNAP-tag technology for labeling proteins with supramolecular elements. Previously these proteins have been labeled with supramolecular host molecules and guest molecules via expressed protein ligation (Chapters 2-4). The use of the same proteins for the labeling with the SNAP-tag will thus allow for an easy comparison of the two labeling techniques. The SNAP-tag was fused to the C-terminus of the fluorescent proteins. Two purification tags, the Strep- and the His-tag were also introduced to allow for purification of both the expressed protein as well as the protein after the labeling reaction. Both constructs were designed and cloned by Dr. Hoang Nguyen. The proteins were expressed in *E. coli* BL21 and purified via the His-tag using Ni-NTA affinity chromatography with yields of typically 20 mg protein per liter of culture medium. Purity and integrity were confirmed by SDS-PAGE (Figure 1) and LC/MS.

3.2 Labeling with host-guest molecules

For the labeling of purified SNAP-tag fusion proteins, the proteins were incubated in phosphate buffer at 37 °C for 30-90 min with two to four equivalents of the BG-conjugate. The labeling worked at dilute protein concentrations of 1-20 μM. This was especially important for hydrophobic guest molecules, as labeling was possible despite the low solubility of the ligands. In Scheme 7 the labeling reaction of YFP-SNAP with lithocholic acid-BG 16 is depicted. The lithocholic acid derivative was dissolved in methanol and subsequently added to the protein solution. After 30 min incubation time, complete labeling of the proteins was observed and excess ligand 16 was removed by exchanging the buffer several times. The second lithocholic acid conjugate 19 was reacted analogously with a SNAP-fusion protein. The purity of the conjugates was confirmed by SDS-PAGE (Scheme 7) and LC/MS (Figure 1).
However, the mass of the observed ligated product was in each case between 16 and 32 Da too high. The same mass increase was observed for different ligands such as cyclodextrin 22 and other ligands, which will be described in the following chapter. This mass increase is most probably caused by one- (16 Da) or twofold (32 Da) oxidation of cysteine or methionine residues in the protein. The SNAP-tag contains several cysteines\textsuperscript{18} including the catalytically active cysteine which forms the thioether bond to the benzyllinker on the ligand of interest (Scheme 7). The concentration of DTT in the mixture was increased to prevent oxidation, but the higher masses were still observed in the LC/MS. Control ligations were performed where the protein was kept under the same conditions as during the ligations but without the BG derivative. The results showed that the protein itself is stable under these conditions. The difference between the protein before and after the ligation is the newly formed thioether bond that links the cysteine of the SNAP-tag to the benzyllinker of the label. An oxidation of this benzylic thioether is a very plausible explanation for the mass increase. However oxidation of this thioether should not have a negative influence on the behavior of the conjugate in host-guest experiments.

Overall it has been shown that the labeling of proteins with the hydrophobic guest molecule lithocholic acid works rapidly, efficient and completely in buffered solution. Compared to the previously used expressed protein ligation, this is an important improvement, as with the SNAP-tag technology the protein is rapidly and completely labeled and no detergent is necessary for dissolving the lithocholic acid reactant.
Labeling with the host-molecule cyclodextrin has also been performed successfully and complete ligation was achieved using a lower excess of cyclodextrin than was required for expressed protein ligation (see experimental section).

**Figure 1**: LC/MS of ligation of CFP-SNAP to lithocholic acid benzylguanine conjugate 19.

### 3.3 Labeling with a supramolecular polymer

To study the labeling of a supramolecular polymer with proteins, bipyridine based discotic molecules 3 and 24 were chosen. These discotics are known to form columnar supramolecular polymers at low concentrations in water. Previously, the assembly of proteins along a supramolecular wire of discotic molecules such as 3 was investigated using the binding of fluorescently labeled streptavidin to biotinylated discotics. These studies showed that the binding of proteins to the supramolecular polymer allows proteins to interact, i.e. show FRET. However, detailed investigations concerning the dynamic properties of the supramolecular polymer-protein system and the influence of multivalent binding of streptavidin are highly desirable. The monovalent and covalent modification of the discotics with one protein would allow for these types of investigations.

The results of the first ligation experiment with a 3-fold excess of BG-discotic 24 performed for 1 h at 37 °C are shown in Figure 2. The LC/MS shows that more than
50% of the protein has been functionalized but that the labeling did not go to completion. This is not surprising and actually anticipated since the stacking properties of the discotics lead to steric hindrance and concomitant to problems during the labeling reaction. Considering the 50 kDa protein mass in comparison with the 3.5 kDa supramolecular scaffold, the 50% reaction yield indicates a highly protein-decorated supramolecular polymer.

Figure 2: TIC MS chromatogram of the ligation of benzylguanine-discotic to YFP-SNAP.

To prevent problems due to steric hindrance and increase the distance between the proteins along the supramolecular wire, mixtures of reactive discotic 24 and inert discotic 3 were prepared. These mixtures should form self-assembled stacks with inert discotics as spacers between the reactive discotics. The concentration of 24 was kept constant and a 5 and 50 fold excess of the inert discotic were used in the following labeling reaction. Unfortunately, this did not lead to higher conversion of the protein. Therefore, further studies were performed to gain insight into the labeling of the supramolecular polymer using fluorescence spectroscopy.

Scheme 8: Labeling of YFP with mixtures of benzylguanine discotic 24 and inert discotic 3.
Labeling of SNAP-tag fusion proteins with supramolecular ligands

To follow the ligation reaction by fluorescence spectroscopy we took advantage of the intrinsic fluorescence of the discotic molecules upon stacking. If excited at 340 nm, the discotics emit between 500 and 560 nm. This makes them a suitable FRET-donor when combined with yellow fluorescent protein (Scheme 8).

![Fluorescence emission spectra](image)

**Figure 3:** Fluorescence emission spectra of 3 µM BGdiscotic 24 and 1 µM YFPSNAP (left) and 3 µM inert discotic 3 and 1 µM YFPSNAP (right) followed over night at 20 °C.

To follow the ligation, 3 µM solutions of discotic were mixed with 1 µM of YFPSNAP in a quartz cuvette and incubated at 20 °C. The change in fluorescence was followed over time as can be seen in Figure 3. The right spectrum shows the control experiment where inert discotic was mixed with YFPSNAP. No change of YFP emission at 527 nm can be observed. However, mixing the benzylguanine discotic with YFPSNAP (Figure 3, left), leads to a significant increase of the emission intensity at 527 nm over time. This shows that there is interaction between the discotics and the protein, which is a proof of a successful ligation reaction.

To study the influence of addition of inert discotic 3, SDS-PAGE was used to study the ligation efficiency. First, different mixtures of inert discotic 3 with increasing amounts of benzylguanine disc 24 were prepared. The overall disc concentration was kept constant at 15 µM to not change the emission of the discotics. As can be seen in both, the fluorescence spectra and the SDS-PAGE (Figure 4), the addition of only one equivalent of reactive disc did not lead to a significant amount of ligation. With increasing amount of benzylguanine disc more of the protein reacted. This experiment shows nicely that indeed the higher emission of YFP corresponds to a higher amount of ligated protein. Very small differences as between the last two mixtures cannot be visualized.
Figure 4: Fluorescence emission spectra of mixtures of benzylguanine discotic 24 and inert discotic 3 and corresponding SDS-PAGE. Time course experiments were performed at 20 °C in presence of 0.05% BSA for 20 h with 1 µM YFPSNAP and discotic concentration as indicated with a total concentration (3+24) of 15 µM.

Subsequently, experiments were performed with the same amount of benzylguanine disc, but increasing amounts of inert disc and analyzed by SDS-PAGE (Figure 5). In agreement with the previous results from the LC/MS, the addition of inert discotic did not improve the ligation efficiency.

Figure 5: SDS-PAGE of different mixtures of inert discotic 3 and benzylguanine discotic 24 for the ligation to YFP-SNAP. Ligation was performed at 37 °C overnight.

Overall, it was shown that the discotics can be covalently funcionalized with proteins via the SNAP-tag technology. However under the conditions studied the labeling does not go to completion. This might be optimized further by studying different
concentrations and ligation times. For further studies on the dynamics and self-assembling properties, it will be important to remove unfunctionalized proteins from the mixtures. Separation of the proteins by size is not possible as the difference in size is too small. An option for purification would be the use of benzylguanine functionalized beads, which are commercially available, that would react with the unfunctionalized proteins and then allow for the functionalized proteins to be collected in the flow through. Even though this is not an optimized approach, it would allow for more studies concerning the interplay of proteins and discotics. Defined mixtures of discotics with different fluorescent proteins could be prepared which would allow for a detailed study of the interplay of proteins and the supramolecular polymer.

4 Conclusions and outlook
In this chapter the synthesis of three supramolecular benzylguanine conjugates, featuring different chemical properties, was described. Purification strategies ranged from normal phase chromatography for the hydrophobic lithocholic acid over reversed phase HPLC for the hydrophilic β-cyclodextrin to size-exclusion chromatography for the bipyridine discotic. It was shown that the different compounds can be successfully ligated to the SNAP-tag fusion proteins. The efficiency and reaction speed for the host-guest molecules was higher than for the previously used expressed protein ligation (Chapters 2-4). Furthermore, the labeling of a supramolecular polymer was successfully performed. Further investigations will focus on the complete functionalization of the proteins and the purification of mixtures with functionalized and unfunctionalized proteins.

The available modified proteins can now be studied in different applications. First studies have been performed with lithocholic acid modified CFP-SNAP by mixing this protein with cyclodextrin modified YFP (Chapter 2, 3). A small FRET-effect was observed (see experimental section). It appears that the fusion of the fluorescent protein to the 20 kDa SNAP-tag increases the distance between the proteins in the host-guest complex significantly. This is disadvantageous for the study of host-guest complexation by monitoring FRET in solution. In these cases the modification via expressed protein ligation is considered to be superior. However, the labeling with the SNAP-tag offers more possible applications, such as immobilization on surfaces, labeling of proteins directly from cell lysates and eventually the labeling of SNAP-
fusion proteins in cells and localization of a supramolecular labeled protein at a specific cellular compartment.

5 Experimental part

General
For analytical LC/MS, reversed phase chromatography was done either using a C18 column 50 x 2.1 mm (GraceSmart) for small molecules or a C4-column 150 x 2.00 mm (Jupiter) for protein LC/MS. For small molecules, a linear gradient starting from 5 % CH$_3$CN/H$_2$O to 100 % CH$_3$CN over 10 min, followed by 1 min at 100 % CH$_3$CN and a final equilibration step to 5 % CH$_3$CN/H$_2$O until 15 min. For proteins, a linear gradient starting from 5 % CH$_3$CN/H$_2$O to 70 % CH$_3$CN over 10 min, followed by 2 min at 70 % CH$_3$CN and a final equilibration step to 5 % CH$_3$CN/H$_2$O until 15 min. Both solvents contained 0.1 % formic acid.

NMR-spectra of lithocholic acid were assigned in comparison with literature assignments based on the publication of Waterhous et al. $^{30}$

Scheme 9: Labeled structures of benzylguanine and lithocholic acid used for assigning NMR-spectra.

Thin-layer chromatography plates were obtained from Merck (Silica gel 60, F254). Compounds were visualized by UV-light (λ = 254 nm, 366 nm) or by staining in an iodide chamber or with anisaldehyde solution.

Lithocholic acid methyl ester 13

Synthesis of 13 was based on a published protocol.$^{31}$ 5 g (13.3 mmol) of lithocholic acid were dissolved in 25 mL of dry methanol. 130 μL of concentrated HCl were added and the mixture was heated under reflux for 5 h. The reaction was stirred at room temperature overnight and the white precipitate formed was isolated by filtration. The addition of water to the filtrate lead to formation of a white precipitate
which was isolated by filtration and identical to the first isolated solid. The combined solids were recrystallized from boiling methanol yielding 2.44 g (47%) of 13 as white crystals. $^1$H NMR (400 MHz, CDCl₃) δ = 3.69-3.58 (m, 4H, -COOCH₃, H₃), 2.42 – 2.17 (m, 2H, H2₃), 2.01 – 0.82 (m, 32H), 0.65 (s, 3H, H1₈). TLC in ethyl acetate / cyclohexane 1:1, $R_f$ = 0.75.

**Lithocholic acid succinimide ester 14**

Synthesis was based on a literature procedure$^{32}$. 375 mg (0.96 mmol) of lithocholic acid methyl ester were dissolved in 4 mL of a mixture of CHCl₃/CH₃CN/Et₃N 1:1:1.2 and 984 mg of disuccinimidylcarbonate (3.84 mmol) were added. The mixture was stirred at 45 °C for 5 hours. The mixture was diluted with 25 mL of chloroform and washed with water. The organic phase was dried over sodium sulfate and concentrated under reduced pressure. The raw product was purified by column chromatography using a gradient from ethyl acetate/cyclohexane 1:4 till 1:3 to obtain 228 mg (45%) of 14 as white residue. $^1$H NMR: (400 MHz, CDCl₃) δ = 4.78 – 4.63 (m, 1H, H3), 3.66 (s, 3H, -COOCH₃), 2.83 (s, 4H, succinimide), 2.44 – 2.14 (m, 2H, H2₃), 2.08 – 0.85 (m, 32H), 0.64 (s, 3H, H1₈). TLC in ethyl acetate / cyclohexane 1:2, $R_f$ = 0.6.

**4-(Aminomethyl)-benzyl alcohol 8**

To 50 mL of dry THF a 1M solution of LiAlH₄ in THF (0.030 mmol) was added. A solution of 2.6 g (0.020 mmol) cyanobenzaldehyde in dry THF was added dropwise to the LiAlH₄ containing solution. The darkgreen suspension was heated to reflux for one hour. TLC showed complete disappearance of the starting material. The mixture was cooled to 0 °C and 1.6 mL of water (0.09 mmol) were added slowly. The yellow mixture was stirred for 30 min at room temperature followed by addition of 11.5 mL of a 15% NaOH solution in water. After stirring for 30 min at room temperature, a clear solution with a white precipitate remained. The mixture was filtered and the precipitate washed with THF. The precipitate was discarded, the solutions were combined and the solvent removed under reduced pressure to yield 2.6 g (>100%) of a yellow solid which was used in the next step without further purification. $^1$H NMR (CDCl₃) δ = 7.31 (m, 4 H, Ar), 4.67 (s, 2 H, CH₂OH), 3.85 (s, 2 H, CH₂NH₂); in accordance with literature data$^{33}$ LC/MS: rt= 0.81 min, C₈H₁₁NO, calculated M= 137.08, found [M+H]$^+$ = 138.00
2,2,2-Trifluoro-N-(4-hydroxymethyl-benzyl)-acetamide 9

The crude product was used directly in a procedure published by Keppler et al. The solid obtained in the reaction before (2.6 g, 0.02 mmol) was dissolved in 20 mL dry methanol and 2.7 mL triethylamine (0.02 mmol) were added. Trifluoroacetic acid ethyl ester (2.86 mL, 0.024 mmol) was added dropwise and the reaction mixture was stirred for 2 h at room temperature. The mixture was diluted with 40 mL water and 40 mL of ethyl acetate. The aqueous layer was extracted with ethyl acetate and the combined organic phases were washed with brine and dried over MgSO₄. The solvent was removed under reduced pressure to give 4.3 g of the crude product which was dissolved in MeOH and adsorbed on Isolute (a modified form of diatomaceous earth with a high surface area that can absorb analytes from polar solvents) for column chromatography using a mixture of ethyl acetate / cyclohexane (1:2 till 1:1) to give 3.7 g (0.016 mmol) (79%) of 9. ¹H NMR (400 MHz, CDCl₃) δ = 7.45 – 7.27 (m, 4H), 6.51 (s, NH), 4.71 (d, J=5.9, 2H, CH₂OH), 4.54 (d, J=5.8, 2H, CH₂NH-), 1.67 (t, J=5.9, 1H, OH). TLC in ethyl acetate / cyclohexane 1:1, Rf= 0.3.

1-(2-amino-9H-purin-6-yl)-1γ⁵,4-diazabicyclo[2.2.2]octan-1-ylium chloride 11

Synthesis was performed similar to a literature protocol. A mixture of 1 g (5.9 mmol) 6-chloropurine and 1.98 g (17.7 mmol) DABCO in DMF was stirred at room temperature under Argon for 4h until TLC showed complete conversion (product fluoresces blue). The precipitate was isolated and washed with a small amount of DMF followed by acetone. The slightly yellow solid was dried under high vacuum and used without further purification. Yield: 1.6 g (5.7 mmol, 97%). ¹H NMR (400 MHz, D₂O) δ = 8.28 (s, 1H, H8'), 4.26 – 4.17 (m, 6H), 3.50 – 3.43 (m, 6H). LC/MS: rt=1.62 min, C₁₁H₁₆N₇⁺ calculated: M⁺= 246.15; found: M⁺= 246.1

N-(4-(2-Amino-9H-purin-6-yloxymethyl)-benzyl)-2,2,2-trifluoro-acetamide 12

The reaction was performed with variations from a procedure published by Keppler et al. To a solution of 250 mg (1.1 mmol) of 9 in 3 mL of dry DMSO 150 mg (3.7 mmol) of 60% NaH in oil were added in portions. The DABCO-purine (300 mg, 1.1 mmol) and 12 mg (0.1 mmol) dimethylaminopyridine were added and the mixture was stirred overnight at room temperature. A mixture of 300 µL water and 100 µL acetic acid was added to quench excess sodiumhydride. The solvent was removed in vacuo. The residue was suspended in methanol and absorbed on isolute for column chromatography starting with 5% methanol in dichloromethane increasing up to 10%
methanol. Yield: 240 mg (0.66 mmol, 60%). $^1$H NMR (400 MHz, DMSO) δ = 12.42 (s, 1H, H9'), 10.00 (s, 1H, -NH-CO-), 7.82 (s, 1H, H8'), 7.57-7.19 (m, 4H, Ar), 6.27 (s, 2H, -NH2), 5.47 (s, 2H, -CH2-O-), 4.40 (s, 2H, CH2-NH-) in agreement with Pauly et al.\textsuperscript{35}. Keppler et al.\textsuperscript{18} report the signal at 4.40 ppm as a doublet. LC/MS: rt= 4.3 min; $\text{C}_{15}\text{H}_{13}\text{F}_{3}\text{N}_{6}\text{O}_{2}$ calculated: M= 366.11; found: [M+H]$^+$= 367.08, [2M+H]$^+$= 732.83. TLC in 10% MeOH / CH2Cl2, Rf= 0.44.

**O$^6$-(4-Amino-methyl-benzyl)guanine 5**

Synthesis was done according to the procedure from Keppler et al.\textsuperscript{18}. 150 mg (0.41 mmol) of 12 was suspended in 2 mL dry methanol under argon atmosphere. After addition of 6 mL methylamine (33% in ethanol) a clear solution was obtained. After stirring overnight at room temperature, a white precipitate formed. The solvent was removed under reduced pressure. The deprotection was quantitative and the product was used without further purification. $^1$H NMR (400 MHz, DMSO) δ = 7.82 (s, 1H, H8'), 7.47 – 7.30 (m, 4H, Ar), 6.26 (s, 2H, NH2 guanine), 5.45 (s, 2H, Ar-CH2-O-), 3.71 (s, 2H, Ar-CH2-NH2). LCMS: rt= 0.9 min, $\text{C}_{13}\text{H}_{14}\text{N}_{6}\text{O}$, calculated M= 270.12, found: [M+H]$^+$= 271.17, [2M+H]$^+$= 541.00. TLC in 10% MeOH / CH2Cl2, Rf= 0.

**Lithocholic acid methylester-benzylguanine 15**

10 mg (0.02 mmol) of succinimide activated lithocholic acid and 6.5 mg (0.02 mmol) of aminomethylbenzylguanine 5 were suspended in 1 mL of DMSO and 3.2 µL (0.04 mmol) of triethylamine were added. The mixture was stirred overnight and the solvent was removed under high vacuum. The remaining residue was dissolved in dichloromethane with 1% of methanol and the solution was washed twice with water. The remaining substance was purified via a small column using 10 % methanol in dichloromethane as eluent. 5 mg (0.007 mmol, 40 %) of pure 15 were isolated. $^1$H NMR (400 MHz, CDCl3) δ = 7.48 – 7.25 (m, 5H, Ar, H8'), 5.45 (s, 2H, CH2-O-), 4.94 (s, 2H, NH2), 4.66 (s, 1H, H3), 4.36 (s, 2H, CH2-NH), 3.66 (s, 3H, -COOCH3), 2.42 – 0.82 (m, 34H), 0.63 (s, 3H, H18). ESI-MS: $\text{C}_{39}\text{H}_{54}\text{N}_{6}\text{O}_{5}$; calculated: M= 686.42; found: [M+H]$^+$= 687.6, [2M+H]$^+$= 1373.9. TLC in 10% MeOH / CH2Cl2, Rf= 0.6.

**Lithocholic acid-benzylguanine 16**

5 mg (0.007 mmol) of 15 were dissolved in 500 µL of methanol. 40 µL of a solution of 10% LiOH in water were added to the solution and the reaction was followed by TLC. After completion, the pH was adjusted to 7 by adding 2 M HCl in water. A small
amount of water was added to the solution and the organic phase was separated. The solvent was removed under reduced pressure. The deprotection was quantitative. $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ = 7.86 (s, 1H, H8’), 7.54 – 7.23 (m, 4H, Ar), 5.53 (s, 2H, -CH$_2$O-), 4.60 – 4.43 (m, 1H, H3), 4.27 (s, 2H, -CH$_2$-NH), 2.30 – 0.86 (m, 3H), 0.68 (s, 3H, H18). LC/MS: rt=7.4 min, C$_{38}$H$_{52}$N$_6$O$_5$, calculated: M= 672.40; [M+H]$^+$= 673.58, [2M+H]$^+$= 1345.83.

**Lithochoic acid methyl ester with amine linker 17**

To a solution of 2.3 mL (10.5 mmol) 4,7,10-Trioxa-1,13-tridecanediamine in 140 mL dry dichloromethane 700 mg (1.3 mmol) of 14, dissolved in 70 mL dry dichloromethane, were added slowly over 4 h. The mixture was stirred at room temperature for another 2 h. The solvent was removed at 40 °C under reduced pressure. The crude product was purified via column chromatography using 5% MeOH/CH$_2$Cl$_2$ and 10% MeOH/CH$_2$Cl$_2$ with 1% NH$_3$ (28% in H$_2$O) as eluents to give 420 mg (0.66 mmol, 50%) of 17. $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ = 5.23 (s, 1H, NH), 4.58 (s, 1H, H3), 3.70 – 3.50 (m, 15H, -CH$_2$O-, CH$_3$COO-), 3.26 (m, 2H, -CH$_2$-NH), 2.80 (t, J=6.7, 2H, -CH$_2$NH$_2$), 2.40-2.15 (m, 2H, H23), 2.04 – 0.82 (m, 36H), 0.64 (s, 3H, H18). $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 174.80, 156.44, 74.46, 70.61, 70.57, 70.21, 69.52, 56.54, 56.00, 51.50, 42.73, 41.88, 40.45, 40.17, 39.63, 38.81, 35.78, 35.37, 35.08, 34.56, 33.20, 32.73, 31.07, 31.01, 29.54, 28.99, 28.19, 27.08, 27.05, 26.36, 24.19, 23.35, 20.82, 18.27, 12.03. MALDI-TOF: C$_{36}$H$_{64}$N$_6$O$_7$, calculated: M= 636.47; found [M+H]$^+$= 637.41, [M+Na]$^+$= 659.40. TLC in 10% MeOH / CH$_2$Cl$_2$ + 1% NH$_3$ (28% solution in H$_2$O), R$_f$= 0.3.

**Lithochoic acid methyl ester benzylguanine with linker 18**

a) To a suspension of 24 mg (0.09 mmol) disuccinimidylcarbonate in 2 mL CH$_2$Cl$_2$, 5.7 $\mu$L (0.09 mmol) triethylamine were added and the mixture was stirred at room temperature for 30 min. A solution of 58 mg (0.09 mmol) 17 in 2 mL CH$_2$Cl$_2$ was added dropwise to the mixture and stirred at room temperature for 2h. The solvent was removed under reduced pressure and the crude product was used in the next reaction without purification.

b) To a solution of 29 mg (0.04 mmol) of the crude in 1 mL DMF, 10 mg (0.04 mmol) O$_5^-$-(4-Amino-methyl-benzyl)guanidine were added. Triethylamine (8 $\mu$L, 0.06 mmol) was added to the suspension and the mixture stirred overnight to yield a clear solution. The solvent was removed in vacuo. The residue could be purified via column
chromatography using a gradient from 2.5% MeOH/ CH2Cl2 to 10% MeOH/ CH2Cl2 yielding 24 mg pure 18 (0.03 mmol, 70%). 1H NMR (400 MHz, CD3OD) δ = 7.83 (s, 1H, H8'), 7.51-7.29 (m, 4H, Ar), 5.53 (s, 2H, -ArCH2O-), 4.48 (m, 1H, H3), 4.32 (s, 2H, -ArCH2-NH), 3.67-3.43 (m, 13H, CH3COO, -CH2O-), 3.31 (m, 2H, -CH2NHCOO-), 3.23 (t, J=6.6, 2H), 3.15 (t, J=6.7, 2H), 2.40-2.15 (m, 2H, H23), 2.05 – 0.84 (m, 36H), 0.66 (s, 3H, H18). 13C NMR (100 MHz, CD3OD) δ 176.46, 161.70, 161.11, 158.77, 141.48, 136.72, 129.70, 128.40, 79.50, 75.87, 71.54, 71.18, 70.16, 69.92, 68.72, 57.87, 57.40, 52.04, 44.56, 43.90, 43.38, 41.86, 41.49, 39.13, 38.78, 37.20, 36.71, 36.18, 35.70, 33.96, 32.24, 31.90, 31.14, 30.86, 29.24, 28.23, 28.19, 27.60, 25.26, 23.86, 21.94, 18.77, 12.48. MALDI-TOF: C50H76N8O9 Calculated: M=932.57, [M+H]+ = 933.67, [M+Na]+ = 955.65. TLC in 10% MeOH / CH2Cl2, Rf = 0.4.

**Lithocholic acid benzylguanine with linker 19**

20 mg of 18 were dissolved in 1 mL MeOH and 50 μL of 10% LiOH/H2O were added and the solution stirred at room temperature for 5 days. The pH of the solution was then adjusted to pH 6 by adding 2 M HCl. CHCl3 was added and the organic phase was concentrated and dried in high vacuum. The deprotection was quantitative. 1H NMR (400 MHz, CD3OD) δ = 8.15 (s, 1H, H8'), 7.41 (m, 4H, Ar), 5.60 (s, 2H, -ArCH2O-), 4.50 (m, 1H, H3), 4.33 (s, 2H, -ArCH2-NH), 3.77 – 3.42 (m, 10H, -CH2O-), 3.31 (m, 2H, -CH2NHCOO-), 3.28 – 3.12 (m, 4H), 2.40 – 2.13 (m, 2H), H23, 2.06 – 0.85 (m, 36H), 0.68 (s, 3H, H18). 13C NMR (101 MHz, CD3OD) δ 176.95, 176.93, 159.89, 140.85, 134.32, 128.80, 127.25, 78.28, 74.65, 70.30, 69.97, 68.93, 68.71, 56.68, 56.26, 45.53, 43.28, 42.70, 42.18, 40.66, 40.31, 37.92, 37.54, 35.99, 35.50, 34.98, 34.49, 32.74, 31.12, 30.83, 29.96, 29.65, 28.03, 27.02, 26.40, 24.06, 22.66, 20.75, 17.60, 11.30. MALDI-TOF: C49H74N8O9 calculated M=918.56, found: [M+Li]+ 925.65, [M+Na]+ 941.63.

**Mono-6-azido-6-deoxy-β-cyclodextrin 21**

For the synthesis, a literature protocol from Tang et al. was followed. A mixture of 2 g (1.6 mmol) monotosylated β-cyclodextrin and 2 g (30.6 mmol) NaN3 was suspended in 200 mL of H2O. The mixture was refluxed for 5 h and stirred at room temperature for additional 6 h. The mixture was poured into acetone and the resultant precipitate was isolated by centrifugation. The precipitate contained unmodified β-cyclodextrin 2 (about 40%) and mono-6-azido-deoxy-6-cyclodextrin 21. It was used without further purification in the next step. LC/MS: rt=3.09 min, C42H69N3O34 calculated M=1159.38, found: [M+H]+ =1159.8, [M+Na]+ = 1182.3.
Cyclodextrin-benzylguanine 22

10 mg of $4^{16}$ (0.03 mmol) were dissolved in 500 µL DMF. 75 mg of the cyclodextrin-azide were added together with 4 mg of CuSO$_4$ (0.02 mmol) and a solution of sodiumascorbate in a 1:1 mixture of H$_2$O and DMF. The mixture was stirred for two weeks and the crude product precipitated with acetone. The precipitate was isolated by filtration, suspended in water, centrifuged and the supernatant lyophilized. The crude product was purified using reversed phase HPLC (C18 column, gradient from 2 to 15% CH$_3$CN/H$_2$O) to yield 9 mg (0.006 mmol) (20%) of pure 22. LC/MS: rt=3.11 min, C$_{58}$H$_{84}$N$_8$O$_{36}$ calculated: M=1468.50, Found: [M+H]$^+$=1470.2, [M+2H]$^{2+}$= 735.7.

Figure 6: LC-MS spectra of cyclodextrin-benzylguanine 22; TIC (top) and UV (bottom).

$^1$H-NMR was measured in D$_2$O at different temperatures. A double set of signals in the aromatic region suggested that the product contained a mixture of conformations where the benzylguanine moiety was complexed partially in the cyclodextrin cavity. It was therefore impossible to assign any signals. More material is needed to fully characterize this compound.

Benzyguanine-discotic 24

A solution of 5 mg (13 µmol) benzylguanine-acid conjugate $6^{15}$ together with 4.7 mg HBTU (13 µmol) and 2.2 µL of DIPEA in 250 µL of dry DMA was prepared and incubated for 15 min. 100 µL of this solution (6 µmol) were added to a solution of 10 mg amine-discotic $23^{28}$ (3 µmol) in dry DMA. The solution was stirred overnight at room temperature. After all the amine discotic had reacted, as followed by MALDI-TOF, the solvent was evaporated and the crude mixture purified via size exclusion
chromatography (SX-1 beads in DMF) to yield 10 mg (2.7 μmol, 90%) of functionalized discotic. \(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta = 15.52\ (s, 3H), 14.49\ (s, 3H), 9.60\ (d, J=8.4, 3H), 9.33\ (m, 3H), 9.06\ (s, 3H), 8.51\ (s, 3H), 8.18\ (s, 1H, CONH'), 8.07\ (s, 1H, H8'), 7.70 – 7.47\ (m, 6H), 7.35\ (s, 6H), 7.20 – 6.93\ (m, 4H, Ar'), 5.51\ (s, 2H, BzO), 4.82\ (d, 2H, BzNH), 4.35-4.20\ (m, 18H), 4.01 – 3.20\ (m, 190H), 2.29\ (m, 4H, COCH\(_2\)CH\(_2\)CH\(_2\)CO), 1.98\ (m, 2H, CH\(_2\)CH\(_2\)CO). MALDI-TOF: \(C_{176}H_{261}N_{19}O_{63}\) calculated \(\text{M}=3675.13\), found [M+Na]^+ = 3697.36.

5.1.1 Plasmids
Plasmids were constructed by Dr. Hoang Nguyen.

To construct the plasmids, pHT403 (Strep-EYFP-SNAP-His) and pHT404 (Strep-ECFP-SNAP-His), the fragments encoding the EYFP and ECFP was amplified by PCR using a pair of primers, ON007, 5’-ggcggatccgctgagcaagggcagggagct-3’ (underlined, BamHI site) and ON008, 5’-ggcgaacctgctgtcagctgccagc-3’ (underlined, ScaI site) with pR2GY\(^36\) and pHT559\(^37\) as templates. The BamHI/ScaI-digested products were ligated into pHT402, respectively. The plasmid pHT402 (Strep-SNAP-His) was constructed by introducing DNA fragment encoding for SNAP-tag using a pair of primers, ON005, 5’-ggcggatccgctgagctgccagc-3’ (underlined, BamHI site) and ON006, 5’-ggcggatccgctgagctgccagc-3’ (underlined, EcoRI site) with pSEMS1-26m (Covalys) as a template into pHT401\(^38\) at BamHI and EcoRI.

5.1.2 Protein expression and purification
Plasmids pHT403 (Strep-EYFP-SNAP-His) and pHT404 (Strep-ECFP-SNAP-His) were transformed into \(E. coli\) BL21 cells using heat shock for 60 sec at 42°C. Transformants were selected on ampicillin (125 mg/L) agar plates. A 10 mL of LB medium containing 100 mg/L ampicillin was inoculated with a single colony and the culture was grown overnight at 37 °C. This preculture was used to seed 2 L of fresh LB medium containing antibiotics, and the culture was incubated at 37 °C until the absorbance at 600 nm (OD600) reached 0.7-0.9. After cooling down to 15 °C, IPTG was added to a final concentration of 0.5 mM, and overnight (or 6 h) induction was performed at 15°C. Cells were harvested by centrifugation (4500 rpm, rotor SLA-3000, 20 min, 4 °C). The pellet was frozen at -80°C. After thawing it was resuspended in BugBuster Reagent (Novagen) according to the suppliers’ instructions. Benzonase Nuclease was added to the suspension which was then incubated on a shaker for 20 min at room temperature. The lysate was cleared by centrifugation (20000 rpm, rotor SA300, 30 min, 4 °C). Columns (EconoPac, BioRad, 20 mL) containing 10 mL of HisBind resin
(Novagen) were equilibrated according to the suppliers instructions. The supernatant was applied to the columns at room temperature. The beads were washed with 10-15 column volumes of washing buffer (0.5 M NaCl, 20 mM Tris HCl, 30 mM imidazole, pH 8). The protein was eluted using buffer containing 0.5 M imidazole. The pooled fractions were concentrated and the buffer exchanged three times to storage buffer (25 mM Na-Phosphate, 50 mM NaCl, pH 7.5) and concentrated to 10 mg/mL. The protein was shock frozen in multiple aliquots using liquid nitrogen and stored at -80 °C. Yields were typically about 20 mg protein per liter culture medium. Proteins were analysed by LC/MS: Strep-ECFP-SNAP-His, $C_{2217}H_{3412}N_{596}O_{649}S_{11}$ calculated Mol. Wt. 49151, found 49156; Strep-EYFP-SNAP-His, $C_{2223}H_{3410}N_{596}O_{648}S_{12}$ calculated Mol. Wt. 49237, found 49235 and SDS-PAGE.

5.1.3 Protein modification

Ligations were performed at protein concentrations of 10-20 μM with 2 to 4 equivalents of benzylguanaine modified ligand in phosphate buffer (25 mM Na-phosphate, 50 mM NaCl) with 1 mM DTT added. Stock solutions with ligand concentrations between 0.1-10 mM were prepared in different solvents depending on the solubility of the ligand. LABG 16 was dissolved in MeOH, LA(n=3)BG 19 was dissolved in DMSO or MeOH, CDBG 22 was dissolved in water and BGDisc 24 was dissolved in MeOH. After addition of the ligand, the mixture was incubated at 37°C for 1 to 5 hours. To remove the small molecule, the buffer was exchanged several times using Amicon filters with a MWCO of 30000 Da. The ligation products were analysed by LC/MS and/or SDS-PAGE. LC/MS: ECFP-SNAP-LA: $C_{2250}H_{3459}N_{597}O_{653}S_{11}$ calculated Mol. Wt. 49673, found 49702 (MagTran); EYFP-SNAP-CD: $C_{2276}H_{3489}N_{599}O_{683}S_{12}$ calculated Mol. Wt. 50555, found 50585 (MassLynx); ECFP-SNAP-LA(n=3): $C_{2261}H_{3481}N_{599}O_{657}S_{11}$ calculated Mol. Wt. 49919, found 49943 (ProMass)

5.1.4 Fluorescence spectroscopy

Fluorescence spectroscopy was performed with a Varian Cary Eclipse photoluminescence spectrometer using quartz cuvettes. Samples were excited at 340 nm for the discotic. Protein concentrations were measured with nanodrop using $A_{515}$ and a molar extinction coefficient of 84 000 M⁻¹cm⁻¹ for yellow fluorescent proteins (YFP). In the case of mixtures of benzylguanaine discotic and inert discotic, the mixtures of the discs were prepared at least 5 h prior to the measurement to allow for random mixing of the stacks. After that the protein solution was added.
FRET-measurements with lithocholic acid modified protein

Fluorescence spectroscopy was performed as described in Chapter 3. Mixtures of mYCD (Chapter 2) with CFP-SNAP and CFP-SNAP-LA (ligated to 16) were prepared and emission spectra recorded exciting CFP at 410 nm. FRET-ratios \( \frac{I(527 \text{ nm})}{I(475 \text{ nm})} \) were calculated (Table 1) and a little higher ratio in case of the modified CFP-SNAP-LA was observed compared to the mixtures with the unmodified control protein CFP-SNAP. However, these ratios are significantly lower than 0.62 which had been observed for mixtures of 0.5 µM mYCD and 0.5 µM mCLA (Chapter 2).

Table 1: FRET ratios for mixtures of mYCD with CFP-SNAP and CFP-SNAP-LA in phosphate buffer, pH7.5, 4 mM DTT.

<table>
<thead>
<tr>
<th>c(mYCD) / µM</th>
<th>0.5 µM CFP-SNAP</th>
<th>0.5 µM CFP-SNAP-LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>1.2</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>1.8</td>
<td>0.54</td>
<td>0.56</td>
</tr>
<tr>
<td>2.4</td>
<td>0.57</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Also, mixtures of CFP-SNAP-LA(n=3) (ligated to 19) with mYCDNH₂ (Chapter 3) were prepared and emission spectra of mixtures of 0.5 µM of each component were recorded in buffer containing 1 mM TCEP. The mixture of CFP-SNAP-LA(n=3) with mYCDNH₂ exhibited a ratio of 0.61 which is significantly less than was observed for the mixture of mCLA with mYCDNH₂ (Chapter 3). However, it is an indication that FRET takes place but is less efficient probably due to increased distance between the chromophores.

6 References

3 Commercially available from New England BioLabs
Labeling of SNAP-tag fusion proteins with supramolecular ligands

Chapter 6

Supramolecular immobilization of SNAP-tag fusion proteins

Supramolecular immobilization of proteins on surfaces and vesicles is being investigated in this chapter. Fluorescent proteins were site-selectively labeled with bisadamantane and ferrocene via the SNAP-tag technology. In cooperation with the group of Pascal Jonkheijm at the University of Twente, the immobilization of these proteins on patterned and unpatterned β-cyclodextrin and cucurbituril surfaces was studied. It is described how initial problems with nonspecific binding due to the SNAP-tag modification were overcome in the case of the bisadamantane functionalized proteins leading to conditions for specific binding to β-cyclodextrin surfaces. The binding of ferrocene functionalized fusion proteins to cucurbit[7]uril monolayers on gold was difficult to optimize due to nonspecific binding of the unfunctionalized protein. However, if the binding conditions are optimized, the functionalization of proteins using the SNAP-tag allows for immobilization of modified proteins via a supramolecular host-guest interaction on surfaces. This provides a new entry in fabricating supramolecular protein patterns on surfaces taking advantage of the high labeling efficiency of the SNAP-tag. Additionally, together with the group of Bart Jan Ravoo in Münster, the binding of bisadamantane functionalized proteins to β-cyclodextrin vesicles could be shown.
1 Introduction

The immobilization of proteins on surfaces is important for applications in different fields of bioanalytics and biomedicine\(^1\). Protein modified surfaces are also of interest for the study of cell growth and cell differentiation. The interaction of cells with modified surfaces plays an important role in tissue engineering and the design of medical implanted materials\(^2\). Supramolecular chemistry can be a tool for protein immobilization that allows control over distribution and density of biological ligands on a surface. The dynamics and switchability of certain host-guest systems, for example, could mimic Nature in a different and potentially better way than covalent chemistry does. The immobilization of proteins based on biological recognition motifs, such as the interaction between biotin and streptavidin or between NiNTA and His-tagged proteins, has already been successfully applied\(^3\). Recently, the use of synthetic molecular systems for the controlled assembly of proteins on surfaces has been studied by different groups. Huskens et al. have used so called β-cyclodextrin molecular printboards for the formation of protein patterns on surfaces by applying different techniques such as microcontact printing\(^4\) and nano imprint lithography\(^5\). Monolayers of β-cyclodextrin have also been applied as platforms for antibody recognition which were in the next step used for the immobilization of cells\(^6\). Not only monolayers based on cyclodextrins have been used for protein assembly on surfaces but also monolayers of cucurbituril, taking advantage of the interaction between ferrocene-derivatives and cucurbit[7]uril\(^7,8\). A different field of protein assembly at interfaces is the interaction of proteins with vesicles. Synthetic amphiphiles decorated with supramolecular ligands can self-assemble into bilayers which can be seen as a model of a biomembrane\(^9\). Therefore the study of such vesicles is of great interest for the understanding of cellular processes. The decoration of polymersomes that carried cyclodextrins on the outside with adamantane functionalized enzymes is an example for the use of host-guest chemistry to decorate vesicles with functional proteins\(^10\).

In this chapter the SNAP-tag technology for the functionalization of proteins with supramolecular ligands, as described in the previous chapter, will be applied for the immobilization of proteins on surfaces and the assembly on cyclodextrin decorated vesicles. The use of the SNAP-tag for these applications is especially interesting as it is compatible with cellular applications. It would for example be possible to label SNAP-tag fusion proteins with supramolecular ligands in the cell or in cell lysates and
Supramolecular immobilization of SNAP-tag fusion proteins

subsequently immobilize these proteins on an appropriately supramolecular decorated surface. Because of the supramolecular nature of the interaction, the corresponding surfaces could be patterned with proteins in a homogeneous, site-selective manner, and the proteins could be reversibly released in a controlled manner. The attachment of a large SNAP-tag required studying the influence of this protein domain on the surface immobilization of the proteins under study. Therefore previously studied host-guest interactions and surfaces were chosen as model systems. As surfaces, cyclodextrin molecular printboards\textsuperscript{11,12} and monolayers of cucurbituril on gold\textsuperscript{13} were selected, and additionally vesicles consisting of amphiphilic cyclodextrins were chosen (Figure 1).

Figure 1: Ferrocene- and bisadamantane- functionalized SNAP-tag fusion proteins for immobilization on CB[7] monolayers (1), β-cyclodextrin-surfaces (2) and β-cyclodextrin-vesicles (3).

The guest-molecules selected for the assembly of the proteins were a bisadamantane guest and a ferrocene derivative as the same moieties had been used previously for the immobilization of proteins on the corresponding surfaces\textsuperscript{8,11,12}. The assembly of proteins on vesicles has not been studied before with the same scaffolds, but the decoration of cyclodextrin vesicles with fluorescent small molecules via a bisadamantane interaction motif has been described\textsuperscript{14}. As proteins, the fluorescent
proteins CFP and YFP were chosen to allow for an easy visualization of protein patterns. Additionally, these proteins have been immobilized before, labeled via expressed protein ligation, which allows for a comparison of the two labeling techniques with respect to ease of protein labeling and immobilization.

In the following the synthesis of the two supramolecular benzylguanine (BG) conjugates for labeling of the SNAP-tag fusion proteins and the subsequent ligation to the proteins will be described. The immobilization studies of the proteins on the different surfaces will be discussed as well as the interaction of adamantane labelled proteins with cyclodextrin vesicles.

2 Synthesis of modified proteins

2.1 Synthesis of adamantane and ferrocene BG-conjugates

The SNAP-tag technology was selected for the supramolecular modification of the proteins under study, because of the efficient labeling as described in the previous chapter. This labeling strategy, developed by K. Johnsson and co-workers\textsuperscript{15}, requires the supramolecular ligands to be conjugated to an $O^6$-benzylguanine which later reacts with the enzyme alkylguanine transferase (SNAP-tag) that is fused to the proteins of interest. The first ligand to be attached to a benzylguanine was a bisadamantane unit\textsuperscript{11,14}. The two adamantane moieties were attached via an oligoethylene glycol spacer to a phenyl unit which was then coupled to the benzylguanine. The design of molecule 9 is based on work in the group of Reinhoudt and Huskens et al. where a similar scaffold has been successfully applied for the assembly of fluorescently labeled adamantanes on cyclodextrin surfaces\textsuperscript{16}. The synthesis is depicted in Scheme 1. Monoadamantyl-functionalized tetraethylene glycol bromide 4 was reacted with methyl-3,5 dihydroxy benzoic acid 5 in acetone in the presence of potassium carbonate and 18-crown-6\textsuperscript{17}. The raw product was purified via column chromatography and pure 6 was isolated in 70 % yield. The bivalent wedge protected as a methylester was treated with potassium hydroxide to yield free acid 7 in 92 % yield without further purification. In the last step 7 was preactivated using HBTU and then coupled to benzylguanine-amine 8\textsuperscript{18,19} in DMF. The pure final product was obtained in 60 % yield after purification via silica column chromatography.
Supramolecular immobilization of SNAP-tag fusion proteins

Scheme 1: Synthesis of bisadamantane-benzylguanine 9.

The second guest molecule that was conjugated to benzylguanine was an amino-methyl ferrocene element\textsuperscript{20}. The design of molecule 12 was similar to a ferrocene-cysteine conjugate which had been ligated to fluorescent proteins and immobilized on surfaces by J. Young et al.\textsuperscript{8}. The ferrocene was attached to a short linker carrying an amine which allowed coupling to the carboxylic acid of benzylguanine 10\textsuperscript{21}. Carboxylic acid 10 was preactivated with HBTU and then slowly added to the ferrocene-amine 11 in DMF at 0 °C and after addition stirred at room temperature. The crude product was purified via silica chromatography to isolate 62 % of the pure product 12.

Scheme 2: Synthesis of ferrocene-benzylguanine 12.
2.2 Ligation to proteins

The two fluorescent proteins used for the immobilization studies were CFP and YFP which were both expressed as SNAP-tag fusion proteins in *E.coli* (chapter 5). The SNAP-fusion proteins can be labeled by mixing with a small excess of the supramolecular ligand and incubation at 37 °C for 0.5-2 hours. Compared to expressed protein ligation it is a very fast and efficient labeling reaction. The ligation of bisadamantane 9 was performed with a 3-fold excess of ligand compared to the protein. Due to the low solubility of the ligand in buffer, a small amount of precipitate formed during the ligation. After two hours the reaction was complete and the precipitate was removed by centrifugation. The excess of ligand was removed by exchanging the buffer, resulting in pure adamantane functionalized proteins, Ad₂CFP and Ad₂YFP. The ligation of CFP to ferrocene 12 (5 eq) was complete as well after incubation for two hours at 37 °C. The buffer was exchanged to remove excess of the ligand. Purity and integrity of all ligated proteins was confirmed by LC/MS (see experimental section of this chapter Figure 9, Figure 10) and SDS-PAGE. These examples demonstrate that labeling via the SNAP-tag is highly efficient even for supramolecular ligands with low solubility yielding pure proteins without any further purification step.

Scheme 3: Ligation of fluorescent SNAP-fusion proteins to benzylguanine conjugates 9 and 12. Ligation was carried out in phosphate buffer, pH 7.5 at 37°C using 5 eq of 12 and 3 eq of 9.
3 Immobilization of adamantane proteins on CD-surfaces

The interaction between the adamantane modified protein, Ad₂YFP, and monolayers of β-cyclodextrin was studied at the University of Twente.

3.1 Surface plasmon resonance measurements

Studies of protein immobilization were performed using surface plasmon resonance. Self-assembled monolayers of cyclodextrins on gold were prepared according to a published procedure using heptathioether-β-cyclodextrin 2\textsuperscript{22}. The samples were equilibrated in a microfluidic cell with running buffer before the protein solution was flown over the surface. As control the unligated SNAP-tag fusion protein YFPSNAP was used to screen for nonspecific binding. First experiments were performed with phosphate buffered saline (PBS) as running buffer. Under these conditions, a significant amount of nonspecific binding was observed. It is known that proteins can bind nonspecifically to cyclodextrin monolayers on surfaces\textsuperscript{11}. For different proteins, different solutions have been found such as the addition of cyclodextrin\textsuperscript{12} to the solution or the use of a monovalent adamantane-hexa(ethylene glycol)\textsuperscript{11} which forms a supramolecularly controlled layer of hexa(ethylene glycol) on the surface that prevents protein binding. The monovalent adamantane can be replaced by the bivalent scaffold which is attached to the protein of interest and therefore ensures specific binding\textsuperscript{11}. Here different concentrations of the detergent Tween20 were added to the running buffer to suppress the binding of the unmodified YFPSNAP. At the optimized concentration of 0.005 % Tween20, binding of 2 µM YFPSNAP could not be detected anymore and at the same time a significant binding of 2 µM Ad₂YFP was observed in the SPR sensorgram (Figure 2). At high protein concentrations of about 10 µM, nonspecific binding of YFPSNAP was observed even in the presence of Tween20. However, the binding of the adamantane protein was in all cases significantly higher. This shows that in the presence of optimized concentrations of a surfactant, bisadamantane functionalized SNAP-tag fusion proteins bind specifically to the cyclodextrin monolayers. The addition of Tween20 can prevent nonspecific binding at low protein concentrations.
To study the concentration dependence of Ad2YFP binding to the cyclodextrin monolayers, a series of SPR-measurements at different concentrations has been carried out. The saturation levels of the different measurements were plotted against the corresponding concentrations of Ad2YFP (Figure 3a). The obtained data points were fitted to a 1:1 Langmuir model, which gave $K_{LM} = 1.6 \times 10^6$ M$^{-1}$. This binding constant is significantly weaker than expected for bisadamantane scaffolds on cyclodextrin monolayers. For the binding of a bis(adamantyl)-functionalized calix[4]arene$^{23}$, for example, the Langmuir fit resulted in binding constants $K_{LM}$ between $10^9$ and $10^{10}$ M$^{-1}$. Here, a comparison of these two binding constants is considered to be valid as the surfaces and binding entities are identical. The only difference between the bisadamantane 9 and the bis(adamantyl)calixarene is the attachment of the ethylene glycole linkers to a phenyl unit (9) and to a calixarene$^{23}$. The significantly weaker binding of the bisadamantane-conjugated protein compared to the binding of bisadamantane-conjugated small molecules can most probably be attributed to the addition of Tween20 to the running buffer. An interaction of the surfactant with the cavities of the cyclodextrin or the adamantanes does not only suppress the nonspecific binding of the YFPSNAP, but also significantly weakens the interaction of the Ad2YFP with the CD-monolayer. To test this hypothesis, the binding of an unfunctionalized bisadamantane (Figure 3b) to the CD-monolayers in PBS was compared to binding in PBS with 0.005 % Tween20. The bisadamantane was flown over the surface at 17 µM concentration in both buffers. In the PBS without surfactant, a significantly higher binding was observed (Figure 3b, grey). The addition

Figure 2: SPR sensorgram of the immobilization of 2 µM YFP-SNAP and 2 µM Ad2YFP on CD-monolayers in PBS-buffer with 0.005% Tween (= running buffer = RB).
of Tween20 lead to almost completely suppressed binding of the small molecule
(Figure 3b, black). This finding supports the explanation of the weaker binding of
Ad$_2$YFP to the CD-monolayers in the presence of detergent. Tween20 might interact
with the cyclodextrin cavities or the adamantane and therefore weakens the binding
of the adamantane at the applied concentrations of the surfactant.

Figure 3: a) Langmuir plot of saturation levels from SPR binding experiments with Ad$_2$YFP on CD-
monolayers performed at different concentrations; b) SPR sensorgrams of small molecule Ad$_2$NH$_2$
(17 µM) binding to CD-surfaces in PBS containing 0.005 % Tween20 and 0.2 % MeOH (black) and PBS
with 0.2 % MeOH (grey). Binding was first done in PBS buffer, then the surface was washed and then
binding was repeated in Tween buffer. The experiment has been repeated in reversed order.

Overall the binding of Ad$_2$YFP to cyclodextrin monolayers is specific in the presence
of Tween20 and exhibits a binding with an apparent binding constant $K_{LM}$ of
1.6 x 10$^6$ M$^{-1}$.

3.2 Protein patterns

Microcontact printing$^{24}$ was the first technique that was applied to fabricate patterns
of Ad$_2$YFP on cyclodextrin molecular printboards. However, Ad$_2$YFP appeared to
aggregate under various printing and inking conditions. With microcontact printing
SNAP-tag fusion proteins can therefore only be printed inhomogeneously. The
formation of homogeneous protein patterns is important, because then the
probability that the protein is still in its native state and therefore also functional is
higher. Here the goal is to generate protein patterns with functional proteins.
Therefore a technique was chosen that avoids the printing step with the protein and
instead is based on the incubation of preformed patterns of the host-molecules.
Cyclodextrin patterns on glass, backfilled with polyethylene glycole, prepared by UV-lithography were incubated with the functionalized protein. As before, the unfunctionalized protein YFP-SNAP was used as a control. First tests using the same buffer as in the SPR-experiments did result in significant amount of unspecific binding. To block any free spots on the surface and to suppress nonspecific binding, the cyclodextrin patterns were incubated with a solution of bovine serum albumin (BSA) first. After washing of these surfaces in buffer, a 2 µM protein solution was applied to the surface which was incubated in a humidity chamber before being washed with buffer. The incubation with YFPSNAP did not result in a protein pattern (Figure 4 left) showing that the blocking step with BSA efficiently repressed nonspecific binding of the SNAP-tag fusion protein to the cyclodextrin surfaces. When the functionalized Ad₂YFP was applied under the same conditions, well defined, homogeneous protein patterns were observed by fluorescence microscopy (Figure 4 middle and right).

**Figure 4:** UV-lithography patterns incubated with YFP-SNAP (left), Ad₂YFP (middle), Ad₂YFP zoom (right).

It has been shown that the incubation of β-cyclodextrin patterns with functionalized SNAP-tag fusion proteins at µM concentrations is a well suited technique to fabricated protein patterns via the applied host-guest interaction. The described optimized protocol resulted in nice homogeneous protein patterns.

## 4 Interaction of adamantane proteins with CD-vesicles

As a mimic of a receptor-protein interaction the binding of adamantane functionalized fluorescent proteins to β-cyclodextrin vesicles was studied together with the group of Bart Jan Ravoo in Münster. Amphiphilic β-cyclodextrins have been synthesized which form bilayer vesicles in buffered solution²⁵. The host-guest
interaction of these vesicles with a bivalent adamantane scaffold carrying a fluorescent dye is already known\textsuperscript{14}. Ad\textsubscript{2}YFP and Ad\textsubscript{2}CFP are carrying the same bisadamantane scaffold and should therefore allow for binding of the fluorescent proteins to the surface of the vesicles.

A possible method to follow the assembly of the proteins on the vesicle is by observing FRET between CFP and YFP, when the two proteins are brought in close proximity. For the fluorescence measurements the two proteins, CFP and YFP, were premixed in a 1 to 1 ratio and then different amounts of freshly extruded vesicles were added. If the proteins assembled on the vesicles, energy transfer from CFP to YFP should be observed. The cyclodextrin concentration ranged between 5 and 20 µM while the protein concentration was kept between 0.5 and 2 µM. Different conditions were tried and the samples observed over 24 h but no changes in the fluorescence spectrum could be observed. As it was not possible to show the binding of proteins to the vesicles using FRET between CFP and YFP, a different technique was applied. Vesicles were incubated with functionalized or unfunctionalized CFP protein and observed under the fluorescence microscope (Figure 5). The bare vesicles were visible in transmission mode but not in the fluorescent mode. In case of the mixture of vesicles and Ad\textsubscript{2}CFP, fluorescence was detected at the vesicles, showing that the protein bound to the vesicles. The control experiment with SNAP-CFP confirmed that the binding was specific and due to the functionalization with the bisadamantane.

These experiments prove that Ad\textsubscript{2}CFP can bind to the vesicles. However, no FRET was observed when Ad\textsubscript{2}YFP and Ad\textsubscript{2}CFP were mixed with the vesicles as described above. A possible explanation is that not enough proteins were bound to the vesicles at the applied concentrations for efficient energy transfer. A second explanation might be that the proteins which did bind to the vesicles did not come close enough to each other due to steric hindrance caused by the large fusion protein (50 kDa). For efficient energy transfer, the two fluorophores have to be as close as their Förster radius which is about 5 nm\textsuperscript{26}. The average distance of two cyclodextrins on the surface of the vesicles is 2.2 nm\textsuperscript{27}, which in principle should allow for binding of the proteins in close proximity. However incomplete surface coverage probably resulted in the observed absence of FRET. For a better readout, so called split proteins might be applied in the future, which become functional upon reassembly on the vesicles. Furthermore, it would be interesting to apply switchable interactions for the assembly of proteins on vesicles such as the interaction between ferrocene and
β-cyclodextrin which is addressable by electrochemistry or the interaction between azobenzene and α-cyclodextrin\textsuperscript{28} which can be switched with light.

Figure 5: Fluorescence microscope images of cyclodextrin vesicles (50 µM) mixed with CFP (5 µM) and Ad₂CFP (5 µM).

5 Immobilization of ferrocene proteins on CB[7] surfaces
The scope of the SNAP-tag modification for the supramolecular immobilization of proteins was further studied applying the interaction between ferrocene and cucurbit[7]uril (CB[7]). This strategy had been applied before in our group\textsuperscript{8}. Yellow fluorescent proteins had been labeled by expressed protein ligation with a monoferrocene ligand and immobilized specifically on CB[7]uril monolayers on gold. Here, the same supramolecular host-guest system was studied, however now using the SNAP-tag for labeling the fluorescent protein with ferrocene. The formation of CB[7] monolayers on gold is based on adsorption of CB[7] molecules on gold as described by An et al.\textsuperscript{13}. CB[7] forms imperfect monolayers with a surface coverage of about 48 % which has been studied by electrochemistry\textsuperscript{13} and dynamic force spectroscopy\textsuperscript{29}. To study binding of the protein to the surface, SPR measurements were performed, using the same buffer as for the cyclodextrin surfaces (\textit{vide infra}), i.e. PBS with 0.005 % Tween. In the binding experiment the ferrocene functionalized
protein (FeCFP) exhibited a stronger binding than the control CFP-SNAP (Figure 6). However, both proteins, CFP-SNAP and FeCFP, showed high binding to the surface, indicating that the addition of 0.005 % surfactant does not suppress nonspecific binding of the protein to the CB[7] surface sufficiently. In the case of the protein that had been modified via expressed protein ligation, SPR measurements indicated only a small amount of unspecific binding of the nonfunctionalized protein to the surface. Apparently the modification with the SNAP-tag increases the probability of unspecific binding to CB[7] surfaces, as seen also for the CD-surfaces.

![Figure 6: SPR sensorgram for the binding of 2 µM FeCFP (grey) and 2 µM CFP-SNAP (black) to CB[7]-monolayers on gold in PBS buffer containing 0.005 % Tween20.](image)

To reduce the amount of unspecific binding to the surface, the concentration of Tween20 was increased to 0.1 % and before binding of the protein of interest a solution of BSA was flown over the surface (Figure 7). The binding of BSA to the surface is significant as can be seen in both experiments. After the surface has been blocked with BSA, there is hardly nonspecific binding of CFP-SNAP detectable. Under the same conditions, FeCFP still binds strongly to the surface. Apparently, BSA passivates the surface and allows for specific binding of CFP-SNAP. However, the saturation levels of the BSA binding of the two samples are differing significantly, indicating that the monolayers are not perfectly packed with CB[7], which is in agreement with the literature. It seems that the formation of the monolayers is not as reproducible as the cyclodextrin monolayers. Therefore, these results have to be considered as being the starting point for further investigations. Possibly a different way of preparing well defined CB[7] monolayers might lead to more reliable results.
A backfilling of empty spaces on the gold surface with polyethylene glycol, for example, could reduce the amount of nonspecific binding.

**Figure 7:** SPR sensorgrams for the binding of 1 µM FeCFP (a) and 1 µM CFP-SNAP (b) to CB[7] monolayers on gold in PBS buffer containing 0.1 % Tween20 (= RB) after blocking with 1 mg/ mL BSA.

### 6 Conclusions and outlook

The use of SNAP-tag labeled fluorescent proteins for immobilization on β-cyclodextrin and CB[7] monolayers has been described. Fluorescent proteins were successfully modified with two supramolecular guest molecules, bisadamantane and ferrocene. In the second step these constructs were applied for assembly on different surfaces. It was found that the bisadamantane modified YFP binds specifically to β-cyclodextrin monolayers on gold in the presence of a surfactant. The binding affinity was lower than expected which could be explained by binding of the applied surfactant to the applied host-guest pair. The formation of YFP-patterns was achieved by incubation of cyclodextrin patterns on glass with the adamantane functionalized protein. Problems due to nonspecific adsorption of unfunctionalized YFP-SNAP were overcome by adding a blocking step with BSA to the protocol which lead to well defined, homogeneous patterns of functionalized YFP. The bisadamantane functionalized protein was further assembled on β-cyclodextrin vesicles. Binding to the vesicles was visualized by fluorescence microscopy. Overall the binding via the bisadamantane works efficiently as expected from examples from literature, however the functionalization of proteins via the SNAP-tag does lead to a significant amount of unspecific binding. This was experienced as well in the binding of ferrocene functionalized CFP-SNAP to cucurbituril[7] surfaces. In that case a blocking step with BSA was necessary to suppress unspecific binding. However, it will be interesting to further study ferrocene functionalized proteins in combination with
Supramolecular immobilization of SNAP-tag fusion proteins

β-cyclodextrin surfaces, where optimized binding conditions to repress nonspecific binding of the SNAP-tag were established. As the binding of one ferrocene moiety to β-cyclodextrin is significantly weaker ($K_a = 10^3$-$10^4$ M$^{-1}$), than the binding to CB[7]$^{30}$, the synthesis of a bivalent ferrocene would allow for application for surface immobilization. The advantage over the binding to CB[7] is the option to electrochemically switch the interaction$^{31}$. Such a bivalent ferrocene would also be applicable for the assembly on vesicles. First studies in that direction are ongoing.

The different examples for supramolecular protein immobilization on surfaces show that, for every protein-surface combination selective immobilization can be achieved, but that binding conditions have to be optimized individually. The scope of this method to immobilize proteins labeled via the SNAP-tag directly from cell lysates is as such doubtful and will significantly depend on the type of surface and host-guest interaction. When using the SNAP-tag it has to be considered that due to fusion of the protein of interest to the SNAP-tag the potential for nonspecific binding of the protein construct is increased compared to labeling techniques that do not require fusion to an additional tag. The protein labeling reaction, however, does proceed very efficiently, making the SNAP-tag approach still a considerable approach for surface immobilization of purified proteins.

7 Experimental part

7.1.1 General
The adamantane derivatives 6, 7 and 9 were synthesized by Dr. Christian Haase. The surface modifications were carried out by and together with Dr. Dorothee Wasserberg at the University of Twente. CD-heptathioether and CD-heptaamine were provided by the MnF group, UTwente. Adamantane-NH$_2$ (Figure 3b) was kindly provided by Deniz Yilmaz, UTwente. The expression and purification of the proteins SNAP-CFP and SNAP-YFP was described in the previous chapter. The synthesis of ferrocene-amine$^8$ and adamantane-bromine$^{16}$ was carried out as described before. CD-vesicles were prepared and incubated with adamantane-functionalized proteins by Jan-Hendrik Schenkel at the Westfälische Wilhelms-Universität Münster.
Adamantane-TEG-Methylester 6

A solution of 1.33 g (7.92 mmol) Methyl-3,5-dihydroxy-benzoic acid and 6.51 g (16.64 mmol) of Adamantane-TEG-Br\textsuperscript{16} and 0.42 g (1.58 mmol) of 18-crown-6 together with 2.74 g (19.8 mmol) of K\textsubscript{2}CO\textsubscript{3} in 200 mL of acetone was prepared and refluxed for 72 h. The solvent was removed under reduced pressure. The residue was separated between H\textsubscript{2}O and ethyl acetate. The aqueous phase was extracted three more times and the combined organic layers were washed with brine and dried with MgSO\textsubscript{4}. The raw product was purified by normal phase column chromatography using a stepwise gradient from 1 % to 10 % MeOH in CH\textsubscript{2}Cl\textsubscript{2} to yield 4.4 g (70%) of pure product. TLC, cyclohexane/ethyl acetate 1/1, Rf= 0.1 ; LC/MS rt 9.0 min C\textsubscript{44}H\textsubscript{68}O\textsubscript{12} calculated M= 788.47, found [M+H]+= 789.75, \textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ = 7.19 (d, J=2.4, 2H, H6), 6.69 (t, J=2.3, 1H, H7), 4.18 – 4.10 (m, 4H, H4), 3.89 (s, 3H, CH\textsubscript{3}O-), 3.88 – 3.82 (m, 4H, H5), 3.78 – 3.63 (m, 16H, -CH\textsubscript{2}CH\textsubscript{2}O-), 3.63 – 3.52 (m, 8H, -CH\textsubscript{2}CH\textsubscript{2}O-), 2.13 (s, 6H, H3), 1.74 (m, 12H, H2), 1.61 (m, 12H, H1). \textsuperscript{13}C NMR (100 MHz, CDCl\textsubscript{3}) δ 166.16, 159.38, 131.46, 107.59, 106.36, 71.68, 70.87, 70.46, 70.26, 70.22, 69.18, 67.37, 58.89, 51.77, 41.09, 36.09, 30.10.

Adamantane-TEG-OH 7

In a mixture of 5 mL H\textsubscript{2}O, 5 mL EtOH and 10 mL MeOH 1.24 g (1.57 mmol) of the adamantane-TEG-methylester and 0.26 g (4.71 mmol) of KOH were dissolved and refluxed for 5 h. The solvents were removed under reduced pressure and the residue dissolved in CH\textsubscript{2}Cl\textsubscript{2} and H\textsubscript{2}O. The aqueous phase was neutralized with 0.5 M HCl and extracted with ethyl acetate. The combined organic layers were dried with MgSO\textsubscript{4} and after evaporation of the solvent 1.12 g (90 %) product were isolated. TLC in 10 %
Supramolecular immobilization of SNAP-tag fusion proteins

MeOH in CH₂Cl₂, R₉ = 0.5, LC/MS rt = 8.30, C₅₃H₆₆O₁₂ calculated M= 774.46, found [M+H]+= 775.7, ¹H NMR (400 MHz, CDCl₃) δ = 7.25 – 7.23 (m, 2H, H6), 6.74 – 6.68 (m, 1H, H7), 4.19 – 4.11 (m, 4H, H4), 3.88 – 3.81 (m, 4H, H5), 3.75 – 3.63 (m, 16H, -CH₂CH₂O-), 3.59 (t, J=3.2, 8H, -CH₂CH₂O-), 2.13 (s, 6H, H3), 1.77 – 1.70 (m, 12H, H2), 1.67 – 1.52 (m, 12H, H1). ¹³C NMR (100 MHz, CDCl₃) δ 168.82, 159.86, 131.83, 108.64, 107.60, 72.56, 71.38, 70.97, 70.78, 70.72, 69.81, 67.88, 59.42, 41.58, 36.59, 30.65.

Adamantane-benzylguanine 9

First a solution of 134.5 mg (0.173 mmol) of Adamantane-TEG-OH, 65.8 mg (0.173 mmol) HBTU and 57.2 µL (0.346 mmol) DIPEA in 1 mL DMF was prepared and stirred for 10 min at room temperature. Then a suspension of 51.6 mg (0.191 mmol) benzylguanine-amine in 1.2 mL DMF was added to the mixture. The solution was stirred overnight at room temperature. The DMF was removed under reduced pressure and to the residue water and ethyl acetate were added. The phases were separated and the aqueous layer was extracted three times with ethyl acetate and the combined organic phases were dried with MgSO₄. The solvent was evaporated and the residue purified by silica column chromatography using a gradient from 1 % - 10 % MeOH in CH₂Cl₂ which contained 0.5 % formic acid yielding 106 mg (60 %) of pure 9. TLC in 5 % MeOH / CH₂Cl₂ with 0.5 % formic acid, R₉ = 0.3, LC/MS rt 7.9 min, C₅₆H₇₈N₆O₁₂ calculated M= 1026.57, found [M+H]+= 1027.58, ¹H NMR (400 MHz, CDCl₃) δ = 8.21 (s, 1H, H11), 8.05 (s, 1H, CONH), 7.38 (m, 4H, H9,H9'), 7.18 (d, J=2.0, 2H, H6), 6.84 (s, 1H, NH), 6.60 (t, J=2.1, 1H, H7), 5.76 (s, broad, 2H, NH₂), 5.33 (s, 2H, H10), 4.64 (d, J=5.2, 2H, H8), 4.21 – 4.10 (m, 4H, H4), 3.85 – 3.74 (m, 4H, H5), 3.73 – 3.50 (m, 24H, CH₂CH₂O-), 2.10 (m, 6H, H3), 1.71 (m, 12H, H2), 1.57 (m, 12H, H1). ¹³C NMR (100 MHz, CDCl₃) δ 168.03, 159.91, 159.49, 159.22, 159.21, 139.91, 135.52, 134.18, 129.85, 127.90, 106.14, 105.68, 100.29, 72.39, 71.20, 70.68, 70.49, 70.48, 69.62, 69.05, 67.66, 59.18, 44.20, 41.41, 36.38, 30.45.

Ferrocene-benzylguanine 12

Benzylguanine 8²¹ (50 mg, 0.13 mmol) was dissolved in 2 mL dry DMF and incubated with 49 mg (0.13 mmol) HBTU and 22.5 µL (0.13 mmol) DIPEA for 15 min at room temperature. The mixture was slowly added to a solution of 57 mg (0.13 mmol) Ferrocene-amine in 4 mL dry DMF at 0°C. The mixture was stirred for additional 15 min. Then 10 µL TFA were added to the mixture and the solvent was removed at 40°C under reduced pressure. The crude product was absorbed on silica and purified via column chromatography using 10% MeOH/CH₂Cl₂ and 10% MeOH/CH₂Cl₂ with 1%
NH₃ (28% in H₂O) to obtain 63 mg (0.08 mmol, 60%) pure product. MALDI-TOF: C₃₀H₅₀Fe₅N₂O₆ Calculated: M=784.34, Found: [M+H]+ 784.41, [M+Na]+ 807.41.¹H NMR (400 MHz, CD₃OD) δ = 7.82 (s, 1H, H8'), 7.37 (m, 4H, Ar), 5.52 (s, 2H, BzO), 4.35 (s, 2H, BzNH), 4.21 (m, 2H, Cₚ), 4.11 (s, 7H, Cₚ), 3.64 – 3.41 (m, 14H, -CH₂O-, FeCH₂), 3.23 (t, J=6.8, 2H), 2.68 (t, J=7.0, 2H), 2.22 (m, 4H, CH₂O(BG)), 1.97 – 1.83 (m, 2H, COCH₂-CH₂-CH₂CO), 1.73 (m, 4H, CH₂CH₂CH₂); ¹³C NMR (100 MHz, CD₃OD) δ 175.19, 175.11, 161.58, 161.32, 157.20, 140.32, 136.92, 129.66, 128.67, 113.46, 85.45, 71.49, 71.47, 71.20, 70.76, 70.16, 69.87, 69.53, 69.16, 68.59, 49.54, 47.51, 43.83, 37.82, 36.30, 36.20, 30.39, 29.85, 23.26.

### 7.1.2 Protein ligation

Ligation of CFP-SNAP to ferrocene was performed at a protein concentration of 40 µM with 5 equivalents of benzylguanine ligand 12 in phosphate buffer (25 mM Na-phosphate, 50 mM NaCl). A 1.5 mM stock solution of Ferrocene-BG in MeOH was prepared, added to the protein and incubated for 2 hours at 37 °C. To remove the small molecule, the buffer was exchanged several times using Amicon filters with a MWCO of 10000 Da. ECFP-SNAP-Fer: LC/MS rt 8.4 min, C₂₂₅₅H₄₃₅⁹Fe₅₉₉O₆₅₄S₁₁ calculated M= 49785, found 49809.

![Figure 9: LC/MS spectrum of FeCFP.](image-url)
Ligation of CFP-SNAP and YFP-SNAP to the bisadamantane molecule was performed at a protein concentration of 50 μM. Adamantane 9 was dissolved at a concentration of 1 mM in MeOH. 3 equivalents of 9 were added to the protein solution and incubated for 2 hours at 37°C. Then precipitate was removed by centrifugation before exchanging the buffer several times using Amicon filters with a MWCO of 10000 Da. ECFP-SNAP-Ad₂ LC/MS rt 8.6, C₂₆₆₈H₃₄₈₆N₅₉₇O₆₆₀S₁₁ calculated M= 50028 found 50030; EYFP-SNAP-Ad₂ LC/MS rt 8.6, C₂₂₇₄H₃₄₈₄N₅₉₇O₆₅₉S₁₂ calculated M= 50114 found 50136.

**Figure 10:** LC/MS spectrum of Ad₂CFP.

### 7.1.3 CD-surface on gold for SPR

β-cyclodextrin surfaces were prepared according to published procedures.²²,²³ 50 nm thick gold layers on glass obtained from SSens B.V., Hengelo were cleaned for 30 s in piranha solution (conc. H₂SO₄ / 30% H₂O₂, 3:1) and rinsed with copious amounts of water. The freshly cleaned and dried substrates were incubated in a solution of 1 mg of β-cyclodextrin-heptathioether²² in 70 mL of a 2:1 mixture of chloroform and ethanol overnight at 60 °C under Argon. Samples were stored under these conditions until they were used for the SPR measurements. Immediately prior to use the substrates were rinsed with chloroform, ethanol and water. The washing steps were repeated three times and the samples dried in a stream of nitrogen.
7.1.4 CB[7] surface on gold for SPR
The CB[7] monolayer was formed as described by An et al.\textsuperscript{13}. The gold surfaces were cleaned as described for the CD-surfaces and then immersed in a saturated solution of CB[7] in water overnight. Before use they were washed with water and dried in a stream of nitrogen.

7.1.5 Surface plasmon resonance
Surface plasmon resonance (SPR measurements) for the adamantane small molecules and ferrocene compounds were performed as described\textsuperscript{8} in a SPR setup in Kretschmann configuration. Glass substrates covered with a 50 nm gold layer and a CD or CB[7] monolayer is attached to a 70 µL volume microfluidic cell and placed on top of a LaSFN9 prism which in turn is mounted on a goniometer head with which the angle of incident of the exciting laser on the prism can be controlled. Light from a 2 mW HeNe laser of 633 nm wavelength passes through the prism and hits the substrate under a variable angle of incident. The intensity of the light is measured by a large-area photodiode. The set-up allows determination of changes of angle of plasmon resonance with an accuracy of 0.002 deg. The gold substrate was optically matched to the prism using an index matching oil. SPR experiments were performed at a continuous flow of 20 µL/min. SPR measurements with adamantane proteins were performed with an iSPR instrument from IBIS Technologies which is also setup in Kretschmann configuration but uses a laser of 800 nm wavelength. The resonance angle is determined by continuously scanning through the surface plasmon resonance dip. The measurements were performed at a continuous flow of 50 µL/min.

The small molecule adamantane-NH\textsubscript{2} was dissolved in methanol first and then diluted with PBS or PBS + 0.005% Tween 20 to a final concentration of 17 µM Ad\textsubscript{2}NH\textsubscript{2} in 0.2 % MeOH in the corresponding buffer.

7.1.6 Preparation of patterned CD-surfaces by UV-lithography
A piranha cleaned 4 inch borofloat wafer was spin-coated with hexamethyldisilazan (HMDS) for 30 s at 4000 rpm. On top, the photoresist Olin 907-17 was spin-coated for 30 s at 4000 rpm. To remove all solvents, the wafer was prebaked at 95°C for 90 s. Subsequently the Olin resist coated wafer was illuminated through a photomask using an EVG 620 mask aligner with UV-light for 4 s (12 mW / cm\textsuperscript{2} Hg-lamp). After exposure, the wafer was baked for 10 min at 120 °C and then developed during 60 s in OPD-4262. The wafers were rinsed with copious amounts of water, dried and
stored in a nitrogen box until further use. For the formation of the amine terminated monolayer, the complete wafer was cut in small parts. Silanization was carried out by overnight chemical vapor deposition of (trimethoxysilyl)propyl-ethylenediamine (TPEDA) in a vacuum desiccator. To remove the excess of silane, the samples were rinsed thoroughly with ethanol. The amine terminated monolayer was reacted with phenyl diisothiocyanate (ITC). The samples were immersed in a 0.04 M solution of ITC in ethanol for 3 h at 50 °C under argon. After the reaction, the samples were rinsed with ethanol and dried in a stream of nitrogen. The ITC-terminated monolayer was then reacted with β-cyclodextrin-heptaamine\(^{33}\). The samples were incubated in a 1 mM solution of β-cyclodextrin-heptaamine in water at room temperature 40°C for 4 hours under argon. Samples were washed afterwards with water and dried in a stream of nitrogen. To remove the Olin resist from the remaining lines, the samples were immersed in a beaker of 60 mL of acetone and sonicated continuously for at least 2.5 h. Samples were rinsed with acetone and dried in a stream of nitrogen. To control if the resist was completely removed, optical microscopy images were taken. To form a PEG-terminated monolayer, the samples were incubated in a solution of 100 µL PEG-(trimethoxy)silane (AB111226 from ABCR) in 60 mL dry toluene overnight at room temperature under argon. Samples were washed with toluene and stored in a nitrogen box until use for protein immobilization.

### 7.1.7 Protein immobilization on patterned surfaces

The surfaces were first incubated in a humidity chamber with a solution of 1 mg/mL bovine serum albumine (BSA) in PBS with 0.005 % Tween 20 (PBST) for 30 min. The solution was removed with a pipette from the surface and the sample was rinsed three times with PBST and washed two times for 10 min with PBST by incubation on an orbital shaker at 80 rpm. Then, the surfaces were incubated in a humidity chamber with a 2 µM protein solution (SNAP-YFP or Ad2YFP) for 90 min. The solution was removed with a pipette and the sample was washed on an orbital shaker in PBST for 60 min and 5 h. The sample was rinsed with water and dried in a stream of nitrogen. The samples were then investigated using an Olympus IX70 inverted fluorescence microscope with a Hg-lamp as light source using appropriate filters.

### 7.1.8 Vesicles

CD-vesicles were prepared as described in the literature\(^{25}\). Fluorescence microscopy was done with a CKX41 inverted microscope from Olympus equipped with a Hg-lamp (U-RFL-T from Olympus). Pictures were taken with a DX 20 L-FW camera (Kappa opto-
electronics GmbH). Vesicles were mixed with the protein solutions as indicated above and observed under the microscope.

8 References

Supramolecular immobilization of SNAP-tag fusion proteins

29 Personal communication Pascal Jonkheijm.
Supramolecular control over protein assembly

Supramolecular chemistry is the study of non-covalent interactions in and between molecules and the resulting multimolecular complexes. The interactions between molecules in biological systems, such as proteins, oligonucleotides, lipids and their complexes are supramolecular interactions. Even though the initial inspiration for novel synthetic supramolecular assemblies came from biology, it was not until only recently that supramolecular systems started to find application in the investigation and modulation of biological systems themselves. Highly relevant in this respect is the need for synthetic supramolecular systems to meet certain requirements in order to be applicable to biological systems. The applied supramolecular interaction should be bioorthogonal and selective and occur at low concentrations in aqueous medium compatible with the typically dilute concentrations of biomolecules. Furthermore, it must be possible to synthetically modify the supramolecular elements to enable their selective conjugation to target biomolecules. In the first chapter, different examples of host-guest systems as applied to the controlled assembly of proteins are discussed.

In this thesis, the application of host-guest chemistry to controlled protein assembly was investigated. Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) were used as protein model systems because of the ability to reliably monitor their dimerization in solution by fluorescence spectroscopy, as well as study their assembly on different surfaces by fluorescence microscopy. In chapters two, three and four, three different host-guest systems were explored as synthetic host-guest systems for protein assembly. The different supramolecular elements were conjugated to fluorescent proteins via expressed protein ligation. For this purpose, proteins were expressed as C-terminal thioesters using an intein-based system. The protein thioesters were then ligated to supramolecules modified with an N-terminal cysteine.

In the second chapter, two sets of CFP and YFP – monomeric analogues and analogues with a weak affinity for dimerization – were functionalized with the host-guest molecules lithocholic acid and β-cyclodextrin. The two protein pairs were studied as model systems for proteins with different intrinsic affinities in an effort to investigate the interplay between the host-guest interaction and the intrinsic protein-protein interaction. In this way, the synthetic host-guest system was shown to induce selective formation of a protein heterodimer for the monomeric analogues. In the case of the dimerizing analogues, the host-guest interactions and the intrinsic
protein-protein interactions acted cooperatively, leading to heterodimer formation at lower protein concentrations. Additionally, the introduction of cysteine side chains next to the host- and guest-elements allowed for covalent locking of the supramolecular-induced protein complex through reversible disulfide bridge formation. The disassembly of the covalently-locked heterodimer could then be induced in a stepwise manner. The trajectory of the disassembly process was found to differ for the two protein pairs as a result of the interplay between the supramolecular host-guest system and the intrinsic protein affinity.

The host-guest complex of lithocholic acid and β-cyclodextrin enables induced protein heterodimerization at μM concentrations. To gain control over protein dimers at even lower concentrations, an advanced host-guest complex consisting of lithocholic acid in combination with heptakis-[6-deoxy-6-(2-aminoethylsulfanyl)]-β-cyclodextrin was investigated in chapter three. This host-guest complex had been reported to possess a significantly higher affinity than the previously applied system with β-cyclodextrin. For protein ligation of heptakis-[6-deoxy-6-(2-aminoethylsulfanyl)]-β-cyclodextrin, conditions were developed for the synthesis and purification of a mono-cysteine intermediate. Expressed protein ligation to YFP and subsequent purification via ion-exchange chromatography yielded pure protein modified with the novel β-cyclodextrin derivative. Protein dimerization of the newly modified protein with a lithocholic acid modified CFP was investigated in solution with fluorescence spectroscopy. First results indicated that the new host-guest complex is indeed applicable to supramolecular protein dimerization at submicromolar concentrations, in line with the original design.

In chapter four, the use of cucurbit[8]uril as an inducer of protein heterodimerization between two modified fluorescent proteins was investigated. Cucurbit[8]uril is known to form a ternary host-guest complex with methylviologen and naphthalene derivatives. Therefore, the proteins CFP and YFP were ligated to methoxynaphthalene and methylviologen respectively and the selective formation of a protein heterodimer in the presence of cucurbit[8]uril could be observed. In addition to the normal enhanced fluorescent protein variant, a second pair of fluorescent proteins, carrying hydrophobic mutations, was investigated to study unspecific aggregation due to the methylviologen modification. In this way it could be shown that the presence of cucurbit[8]uril shields the methylviologen thereby enabling the use of the ternary host-guest complex for controlled protein assembly in solution.
Chapters five and six of the thesis investigate the application of the SNAP-tag technology for the labeling of proteins with supramolecular elements. SNAP-tag labeling is compatible with intracellular labeling and labeling in cell lysates, possibly making it superior to expressed protein ligation for the use of such supramolecular systems to answer more biologically-relevant questions. First, different supramolecular ligands were attached to an $O^6$-benzylguanine scaffold, which is required for reaction with SNAP-tag fusion proteins. Fluorescent proteins were expressed as SNAP-tag fusion proteins and then ligated to benzylguanine conjugates of lithocholic acid, $\beta$-cyclodextrin and a supramolecular polymer. Compared to the previously used ligation technique, the SNAP-tag labeling is a fast and high yielding method for the ligation of proteins to different supramolecular elements. The last chapter deals with the application of SNAP-tag-labeled proteins for immobilization on different surfaces in collaboration with groups from the University of Twente and the Westfälische Wilhelms Universität Münster. Proteins were first conjugated to bisadamantane and a ferrocene respectively. Then the immobilization on $\beta$-cyclodextrin surfaces and cucurbit[7]uril surfaces was studied. The binding of the functionalized proteins was influenced by the SNAP-tag fusion, such that individual optimization of binding conditions was required. Under optimized conditions, adamantane functionalized fluorescent proteins were specifically immobilized on $\beta$-cyclodextrin surfaces and $\beta$–cyclodextrin vesicles. Furthermore the attachment of ferrocene to a SNAP-tag fusion protein enabled immobilization on a cucurbit[7]uril surface. Taking into account the high protein labeling efficiency, the SNAP-tag is therefore a valid approach for constructing new protein modified materials.
Dana Uhlenheuer was born 10\textsuperscript{th} of June 1983 in Dortmund (Germany). After finishing secondary education (Gymnasium) in 2002, she started her chemistry studies at the Technische Universität Dortmund. As part of her studies she worked for three months at Schering AG Berlin in the organic synthesis laboratory. In 2007 she obtained her diploma in Chemistry with honors in the group of prof. dr. Waldmann under the supervision of dr.ir. Brunsveld at the Max Planck Institute of Molecular Physiology. For her diploma work titled ‘Supramolecular modulation of protein-protein interactions’ she was awarded the ‘Jahrgangsbestenpreis’ for the best diploma thesis of the year. In 2007 she started her PhD thesis in the same group continuing the work on supramolecular chemistry for protein assembly. In September 2008 she moved from the Max Planck Institute in Dortmund to the Technische Universiteit Eindhoven to continue her project in the group of prof.dr.ir. Brunsveld in the laboratory of chemical biology. The most important results of this work are presented in this dissertation.
During the last years I have been supported by and worked together with many people whom I want to thank on these last pages of my thesis.

First of all, I want to thank my supervisor prof. dr. ir. Luc Brunsveld: Luc, thank you for giving me the chance to work on such an interesting and challenging project. I am grateful for your support and your constant belief in the success of the project. I learned a lot in your group where you manage to keep an inspiring, stimulating and always enjoyable working atmosphere.

I want to express my gratitude to the members of my reading committee, prof. dr. V. Subramaniam, prof. dr. B. J. Ravoo, prof. dr. R. Sijbesma and dr. M. Merkx for critically reading my thesis which helped to improve the quality of this book. Furthermore I thank prof. dr. E. W. Meijer and dr. ir. P. Jonkheijm for taking part in the committee.

Different collaborations gave me the opportunity to learn new techniques and apply my knowledge in different contexts. I am happy that already in the beginning of my work the collaboration with the group of Vinod Subramaniam in Twente was initiated. I would like to thank Christian Blum for the help with the lifetime experiments described in chapter 2. Two years later, I came back to Twente for surface immobilization experiments in the group of Pascal Jonkheijm together with Dorothee Wasserberg. I also want to thank the people in Twente for the very friendly and helpful working atmosphere. Dodo, ich danke dir für viele amüsante und interessante Stunden im Labor, in denen ich viel gelernt habe über Spektroskopie, Oberflächen und vieles mehr. Ich bin froh, dass unsere Projekte beschrieben in Kapitel 2 und 6 so erfolgreich waren. Außerdem möchte ich mich bei Martin Schneiders und Jan-Hendrik Schenkel aus der Gruppe von Bart Jan Ravoo in Münster bedanken, die mir einen Einblick in die Welt der Cyclodextrin-Vesikel ermöglicht haben. Furthermore I would like to thank the group of Frans Mulder in Groningen who tried to measure NMR of a supramolecular protein complex. Even though it turned out to be too complex, I enjoyed the stay in your group and it was impressive to see what NMR can do.

If one wants to work successfully on an interdisciplinary topic, it means working together in a team. I was lucky to be part of such a team in our group consisting of Hoang Nguyen, Dorothee Wasserberg, Jacqui Young, Lech Gustav Milroy, Katja Petkau, Christian Haase and Li Zhang. Hoang is especially acknowledged for everything he did on the biology side of the project. Hoang, I am happy that even
Acknowledgements

though our dream of doing supramolecular chemistry in cells turned out to be too ambitious for the moment, you were always motivated to help me with protein purification, culturing mammalian cells or cloning another construct. I learned a lot from you. Jacqui is thanked for the initial studies on the cucurbit[8]uril induced protein dimerization described in chapter 4. Jacqui, I learned a lot from you regarding synthesis and NMR analysis. I want to thank Christian and Lech for synthesizing the bisadamantane molecule described in chapter 6 and the cysteamine cyclodextrin investigated in chapter 3. Katja synthesized the discotic molecule that was used for labeling studies in chapter 5. Furthermore she is acknowledged for designing the cover of this thesis and help with design related questions. Finally, I would like to thank Marcel Scheepstra who joined me for a project for his studies and successfully synthesized the viologen modified protein described in chapter 4. Matt, Jacqui, Katja, Lech, Dodo and Marcel are thanked for proofreading the thesis at different stages.

After moving from Dortmund to Eindhoven, we got a lot of support and help with getting started again. I want to especially thank Hans, who did all the ordering and was extremely important in getting all the bits and pieces together for the lab. Thanks to Joke for all the support and help with official questions and organizing social group and lab activities. I also appreciated the help of prof. Schubert’s group who shared the lab with us in the beginning. Without you, it would have been quite empty during the initial period in Eindhoven and precious compounds would still be trapped in the desiccators… I want to thank Wencke for all the work connected to expanding the lab and keeping it safe and workable with the steadily growing number of people. I also want to thank the people from Rint’s group who filled the lab with life (and experience) after the Schubert group left. I enjoyed working with Maarten’s group in the biolab and the joined Friday morning meetings which were an excellent platform to create new ideas and get valuable feedback. Tom, Katja, Bob and Marko are acknowledged for the efforts they put in setting up the autosampler on the NMR. For mass spectrometry, the analytical lab offers many possibilities and I am happy that I could make use of some of them. I want to especially thank Ralf and Lou for many nice MALDI spectra and always being helpful.

Over the last years in the group in Dortmund and Eindhoven, I was always fortunate to have exceptionally friendly and helpful colleagues in my office and the lab. Especially the good atmosphere in the Brunsveld group made work enjoyable – I hope that you can keep the helpful and friendly spirit. You made up for the less nice experiences in and around the lab. The joined activities in the evenings and
weekends definitely made life in Eindhoven enjoyable – for example Sascha’s parties, Burns Night at Tessa & Matt, barbecues, the evening out after the student(s) left, . . .

I will miss you a lot and hopefully stay in contact with many of you.

I would like to thank Melissa and Katja for being the paranimfen at my defense. Melissa, you were very important by helping me with research related questions and it was great that you always made time to discuss (also about unrelated things). I am happy that also after you moved we stay in contact and you still always have time to chat. Katja, die letzten drei Jahre haben wir viel Zeit zusammen verbracht (in der Wohnung, im Büro, im Labor und in der Freizeit). Ich danke dir für deine wichtige Unterstützung in diesen (nicht immer ganz einfachen) drei Jahren. Ich bin froh, dass wir uns immer noch nicht auf die Nerven gehen und ich eine neue Freundin gewonnen habe. Ich werde dich vermissen.


Lieber Marcel, die letzten Worte sind für dich. Um es kurz zu machen – in den letzten 1.5 Jahren bist du zum wichtigsten Menschen in meinem Leben geworden und ich wünsche mir, dass das noch lange so bleibt.