Supramolecular cucurbit[8]uril induced protein dimerization
Dang, T.D.

DOI:
10.6100/IR733409

Published: 01/01/2012

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 author's 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

Citation for published version (APA):

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?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Supramolecular cucurbit[8]uril induced protein dimerization

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 dinsdag 19 juni 2012 om 16.00 uur

doort

Dũng Thanh Đặng

geboren te Đồng Tháp, Việtnam
Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. L. Brunsveld
Table of Contents

Chapter 1
Molecular control over protein dimerization; opportunities for supramolecular chemistry
Abstract
1.1 Introduction ..............................................................................................................................2
1.2 Small molecule induced protein dimerization .................................................................6
  1.2.1 Gene expression ................................................................................................................7
  1.2.2 Controlling post-translational processing of proteins ..................................................10
  1.2.3 Signal transduction/ kinase activation ...........................................................................11
1.3 Metal ion-mediated protein dimerization ........................................................................13
1.4 Supramolecular inducers of protein dimerization .............................................................14
  1.4.1 Cyclodextrin-induced protein dimerization .............................................................16
  1.4.2 Cucurbit[8]uril-induced protein dimerization ...........................................................17
1.5 Conclusions .............................................................................................................................18
1.6 Aim and outline of this thesis ..............................................................................................19
1.7 References .............................................................................................................................22

Chapter 2
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril
Abstract
2.1 Introduction .............................................................................................................................26
2.2 Results and discussion .........................................................................................................27
  2.2.1 Construction of proteins ..............................................................................................27
  2.2.2 Cucurbit[8]uril induced protein homodimerization .................................................28
  2.2.3 Size exclusion chromatography and ESI-MS measurement ......................................28
  2.2.4 Cucurbit[8]uril-induced protein heterodimerization ...............................................29
  2.2.5 Reversible inducer of protein dimerization ...............................................................30
2.3 Conclusions .............................................................................................................................31
2.4 Experimental section ...........................................................................................................32
2.5 References .............................................................................................................................38
Chapter 1

Molecular control over protein dimerization; opportunities for supramolecular chemistry

Abstract

Protein-protein interactions and protein dimerization play a key role in many biological processes. Most of cellular events such as enzyme activation, transcriptional cofactor recruitment, signal transduction, and even pathogenic pathways are significantly regulated via protein dimerization or protein-protein interactions. Understanding and controlling the molecular mechanisms that regulate protein dimerization is crucial for biomedical research. The limitations of engineered protein dimerization provide an opportunity for external molecules to induce protein dimerization in biological events. In this chapter, molecular control over protein dimerization and opportunities for supramolecular chemistry in this respect are discussed. Small molecule-induced protein dimerization-based approaches provide powerful tools to modulate the functionality of dimerized proteins in gene expression, post translational processing and signal transduction. Supramolecular chemistry as an orthogonal approach to induce protein dimerization represents a novel promising system. This chapter highlights some of the well known approaches and subsequently brings supramolecular chemistry forward as a novel means to control protein dimerization. Specific focus is on cucurbit[8]uril-based host-guest chemistry as a novel, robust and versatile supramolecular approach to explore and modulate the molecular mechanisms of protein dimerization such as fluorescent protein assembly, caspase-8 and caspase-9 dimerization and activation and surface membrane protein dimerization.
1.1 Introduction

Protein-protein interaction is a crucial biological process in which proteins interact, as for example homo- or heterodimers, to form a functional assembly (Fig. 1). In fact, proteins rarely show function and activity in their isolated form in a biological environment. The self assembly of proteins to form dimers or higher oligomeric aggregates is a common biophysical phenomenon, which occurs in every cellular compartment such as cell membranes, the nucleus, and the cytosol. All cellular pathways such as enzymatic activation\[1;2\], signal transduction\[3\], and even pathogenic pathways\[4\] are significantly regulated via protein dimerization.

![Diagram of protein dimerization](image)

Figure 1. Schematic representation of protein homo- and hetero-dimerization

Regulation of protein dimerization is an essential process to develop the physical structure of organisms under the stimuli of intrinsic or extrinsic factors in the natural environment\[5;6\]. Therefore, understanding the molecular mechanisms of protein dimerization and their function represents the cutting edge of research and provides multiple entries for biomedical applications. For example, while the dimerization of caspase-9 triggers enzymatic activity\[7\], procaspase-9 exists in the inactive monomeric form under physiological conditions in the cellular cytosol. The presence of ATP and cytochrome c (from mitochondria under stress stimuli) leads to the formation of a complex with the cofactor Apaf-1 (apoptotic protease activating factor 1) to generate an apoptosome. The apoptosome is responsible for recruitment of procaspase-9 proteins to induce caspase-9 dimerization, leading to rearrangement of the catalytic active site. Dimerized, activated caspase-9 cleaves and activates caspase-3/7, triggering apoptosis\[8;9\] (Fig. 2). Different from the mechanism of caspase-9 activation, but analogous in the need for dimer formation, the dimerization of caspase-8 is stimulated by formation of a death-inducing signaling complex (DISC) with the Fas ligand-binding Fas death receptor\[1\]. Procaspase-8, consisting of a death effector domain and a protease domain, also exists in the monomeric inactive form under physiological...
conditions. The stimulus of a death signal enables a Fas ligand to bind to the Fas receptor, which recruits procaspase-8 to generate the DISC complex. The formation of DISC allows the processing of procaspase-8 through self-cleavage, thereby generating the dimeric form. The dimer caspase-8 shows activity by cleaving caspase-3/7, which again dimerizes itself and ultimately leads to apoptosis (Fig. 2).

Figure 2. Activation mechanism of caspase-8 and caspase-9 via controlled dimerization. a) Extrinsic apoptotic pathway: the Fas ligand binds to the Fas death receptor which allows procaspase-8 to generate DISC, thereby promoting the formation of dimerized, activated caspase-8 (1a). The active caspase-8 cleaves (2a) and activates caspase-3/7 to dimerize (3a), resulting in apoptosis (4a). b) Intrinsic apoptotic pathway: An apoptosome is generated by the binding of cytochrome c to Apaf-1. The apoptosome is responsible for the recruitment of procaspase-9 leading to an increased local concentration of procaspase-9, resulting in generation of active caspase-9 dimers (1b). Once activated, the caspase-9 dimer cleaves (2b) and activates procaspase-3/7 to form a caspase-3/7 dimer (3b), causing apoptosis (4b).

Protein dimerization at the cellular membrane is a crucial factor for the activation of membrane proteins in signal transduction pathways. Dimerization of membrane protein receptors occurs upon binding of ligands such as growth factors or polypeptide hormones. Indeed, many protein-tyrosine kinase receptors are activated by dimerization. For example, epidermal growth factor receptor (EGFR) is a member of four structurally related receptor
tyrosine kinases which locate on the cell plasma membrane and is activated in the dimeric form by its specific binding to ligands such as EGF (epidermal growth factor) and TGFα (transforming growth factor α)\cite{11,12}. The monomeric EGFR typically resides in the inactive form, but gets activated to the dimeric form in the presence of EGF\cite{13}. The crystal structure of the EGFR:EGF complex demonstrates that the binding of EGF to domains I and III of EGFR causes a conformational change in domain II that leads to EGFR dimerization\cite{13-15}. EGFR dimerization stimulates the intrinsic intracellular activity of its tyrosine kinase through autophosphorylation of several tyrosine residues of EGFR at its C-terminus (Fig. 3). Protein-tyrosine kinase activation initiates several signal transduction cascades, resulting in DNA synthesis and cell proliferation\cite{16}.

![Figure 3. Phosphorylation of intracellular tyrosine residues of EGFR occurs via ligand (E = EGF)-induced EGFR dimerization.](image)

Protein engineering represents one approach to induce or control protein dimerization, thereby facilitating an increase in protein stability and/or function\cite{17} (Fig. 4). Engineering a dimeric interface of initiator caspase-9 resulted in caspase activation through induced proximity. Shi and coworkers have generated a dimeric caspase-9 by replacing five residues in the β6 strand of caspase-9 (Gly\textsuperscript{402}-Cys-Phe-Asn-Phe\textsuperscript{406}) with those normally present in caspase-3 (Cys\textsuperscript{264}-Ile-Val-Ser-Met\textsuperscript{268}), resulting in a dimeric interface of an engineered caspase-9. The engineered caspase-9 functioned as homodimer in buffered solution, leading to an increase in enzymatic activity \textit{in vitro} and in cell-based studies\cite{18}. Additionally, numerous examples of engineered protein dimerization generated by a so-called “coiled-coil zipper” approach have also been demonstrated\cite{19-21}. The coiled-coil zipper structure functions through the hydrophobic interaction between leucine rich motifs which form homo- or heterodimeric states. An example in this respect is the dimer formation between c-Jun and c-Fos, to form functional DNA transcriptional factors\cite{22}. The dimer formation of leucine-rich zippers provides a bioengineering approach that enables induced dimerization of
proteins bearing leucine-rich repeats. For example, the introduction of a leucine zipper motif to quiescent cell proline dipeptidase (QPP), enabled QPP homodimerization, which is essential for QPP activation\(^\text{[23]}\). The leucine zipper motifs were also applied for induced dimerization of other proteins such as protein kinase (MLK-3)\(^\text{[24]}\) and tyrosine hydroxylase\(^\text{[25]}\) in which leucine zipper-induced protein dimerization showed a significant increase in enzymatic activity.

Notwithstanding the great success achieved, the current approaches to engineered protein dimerization interfaces have their limitations, especially in terms of the control over the dimerization event. For example, an engineered dimeric caspase-9 could not sufficiently rearrange to allow for optimal proximity/activity of the caspase-9 active-sites, resulting in significantly lower enzymatic activity compared with the Apaf1-activated caspase-9 (the natural mechanism of activation for caspase-9 in the apoptotic pathway)\(^\text{[18]}\). Mutations made in the active domains of target proteins may change their biological structure and function. Furthermore, approaches based on the addition of dimeric interfaces, such as via addition of leucine zipper-induced domains, are irreversible and do not provide a switching mechanism to control or regulate dimerization. The function of resulting proteins in their biological processes is therefore analogously only difficult to control. Therefore, protein dimerization approaches which are based on external molecular signals, capable of inducing or blocking the dimerization, provide strong entry points to explore and control the molecular mechanisms of protein dimerization.

Figure 4. Schematic representation of protein engineering approaches towards protein dimerization and molecule controlled protein dimerization approaches.
Currently, a powerful chemical approach used to control protein dimerization is via the use of small molecules\[^{28}\]. More recently also the use of metal ions and synthetic host-guest systems has been explored (Fig. 4). These chemical elements act by bringing two proteins together and induce their dimerization, chemical inducers of dimerization, possibly resulting in the activation or inhibition of biological events. The reversibility of chemically-induced protein dimerization is attractive for biochemical research, as it enables an added degree of control over protein dimerization and activation. In this chapter, a schematic overview is provided of protein dimerization mediated by different chemical inducers of dimerization. Examples are highlighted based on the regulation of enzymatic activity, DNA transcription, post-translational processing, and signal transduction.

1.2 Small molecule induced protein dimerization

The concept of small molecule induced protein dimerization is based on the use of low molecular weight organic compounds bearing bifunctional moieties which interact simultaneously with two proteins or protein domains. A chemical inducer of protein dimerization acts as dimerizer to bring protein molecules together to form either a homo- or a heterodimer (Fig. 5).

![Figure 5. Schematic representation of small molecule induced protein dimerization: A) protein homodimerization induced by a symmetric bifunctional molecule, B) protein heterodimerization induced by an asymmetric bifunctional molecule.](image)

Small molecule approaches to induced protein dimerization have been demonstrated based on natural products and their derivatives such as rapamycin, cyclosporine, FK506, and coumermycin (Fig. 6). For example, the natural product rapamycin has emerged as the first biofunctional dimerizer to induce heterodimerization of proteins\[^{27-30}\]. The most prominent molecular feature of rapamycin is its two chemically distinct protein binding domains: one part of the molecule binds with high nanomolar affinity to the FK506-binding protein (FKBP12), the other molecular part to the FRB domain of mTOR, FRAP (FKBP-rapamycin}
associated protein), overall resulting in dimerization of the proteins involved\[31\]. Rapamycin is thus capable of inducing heterodimerization of fusion proteins featuring FKBP and FRB domains. In contrast to rapamycin, coumermycin has two of the same protein-binding moieties and can be used to induce homodimerization of Gyrb (bacterial DNA gyrase B)\[32-34\]. The concept of small molecule-induced protein dimerization can be extended to novel synthetic compounds as well. For example, a synthetic dimer of FK506, named FK1012, promotes FKBP12 homodimerization\[35\], or a synthetic dimer of cyclosporine named (CsA)\(_2\): can induce dimerization of cyclophilin\[36\]. These small molecules form orthogonal dimerizing systems which can be exploited to regulate for example transcription factors, post-translation processing and signal transduction pathways.

\[\text{Figure 6. Chemical structures of small molecules inducers of protein dimerization. Natural products and their derivatives: rapamycin and coumermycin, (cyclosporin A)\(_2\) and FK1012.}\]

1.2.1 Gene expression

Gene expression is the fundamental event at which the genetic information contained within a gene’s DNA sequence is translated into a functional protein. Regulated gene expression by transcriptional activators is a crucial mechanism for the development and homeostasis of cells and organisms. Indeed, the human genome encodes thousands of transcriptional activators.\[37\] Regulating the activities of transcriptional activators is possible by controlling the interaction of proteins with DNA. Therefore, a small molecular inducer can be used to gain control over gene expression by using the small molecule induced dimerization of fusion proteins targeting DNA. For example, rapamycin-controlled gene activation via protein dimerization has been demonstrated using rapamycin-induced dimerization of the FRB domain and FKBP. Fusion of an FRB domain to an activation
domain (VP16) and a DNA-binding domain (Gal4) to an FKBP domain led to rapamycin-induced dimerization generating transcriptional activator functionality, and the promotion of gene expression (Fig. 7)\textsuperscript{[38,39]}. Since toxicity of the natural rapamycin inhibits cell proliferation, Crabtree and co-workers have developed non-toxic rapamycin analogs which were successfully used to control gene expression\textsuperscript{[27]}.

Alternatively, Verdine and coworkers showed that a FK506-activator peptide conjugate behaves as a transcriptional coactivator. This coactivator is capable of bringing an FKBP-GAL4 fusion protein to the basal transcriptional machinery, which then activates gene transcription \textit{in vitro} and \textit{in vivo} (Fig. 8)\textsuperscript{[40]}.

![Figure 7. Schematic representation showing rapamycin-induced dimerization of FRB-fused VP16 and FKBP-fused Gal4, resulting in activation of the reporter gene.](image)

![Figure 8. A) Synthesis of co-activator conjugation of activator peptide to FK506, B) Schematic representation of the synthetic co-activator functioning as a bridge between the DNA binding protein domain GAL4-FKBP and the basal transcriptional apparatus during gene expression.](image)

A small molecule induced protein dimerization approach was also used to study the function of activators and repressors which have been used to control gene transcription \textit{in}
Indeed, the recruitment or displacement of an activator or a repressor can be a crucial mechanism for the regulation of gene expression. In order to prove this concept, Crabtree and coworkers used the co-occupancy of FK506 and rapamycin to regulate the function of transcription activators and repressors. To activate transcription, FK506 firstly recruited an activator to the promoter, resulting in an activation of transcription. Subsequently, rapamycin competes with FK506, leading to a displacement of the activation domain from the promoter, resulting in inhibition of transcription. Alternatively, the rapamycin binding domain (FRB) was fused to a transcriptional repressor allowing rapamycin to recruit the repressor to the promoter, thereby inhibiting transcription.

Figure 9. Schematic representation of coumermycin/novobiocin-regulated gene expression. Coumermycin-induced λR-GyrB transactivator dimerization and subsequent binding to the λ operator, thereby switching on gene expression. The presence of novobiocin inhibits coumermycin-induced λR-GyrB transactivator dimerization, resulting in a down-regulation in gene expression.

A pair of well studied molecules – coumermycin/novobiocin – have been used to regulate gene expression in mammals via coumermycin-induced homodimerization of protein GyrB (Fig. 9)\textsuperscript{[32]}. The transactivator for coumermycin-induced dimerization was constructed by fusion of GyrB to the λ repressor-binding domain (λR). Therein, the site directed mutagenesis of λR was also generated to reduce the basal activity. The mutant λR fused to GyrB could effectively and reversibly increase the regulation of gene expression by coumermycin-induced transactivator dimerization. In this case, the efficiency of dimerization for the purpose of gene expression was determined using the reporter genes, genetically encoded green fluorescent protein and luciferase. The monomer of the λR-GyrB transactivator was incapable of bind to λ operator, resulting in downregulation of target gene expression. The cell permeable coumermycin-induced dimer formation of the λR-GyrB
transactivator bound to λ operator, thereby activating the target gene expression. Thus, the luciferase activity was approximately 50-fold greater in the presence of coumermycin, while the luciferase activity was effectively inhibited by the presence of dose-dependent novobiocin. This could be explained by novobiocin behaving as a competitor for binding to GyrB, resulting in inhibition of coumermycin-induced λR-GyrB transactivator dimerization. Indeed, the ability of coumermycin to induce dimerization of λR-GyrB at nanomolar concentrations renders this molecule a potential dimerizer for the regulation of gene expression in eukaryotic cells.

The use of small molecule induced protein dimerization provides a powerful means to regulate gene expression. Cell-permeable and protein-specific binding of small molecules shows considerable potential for controlling transcriptional factors not only in vitro but also in vivo.

1.2.2 Controlling post-translational processing of proteins

Post-translational processes are crucial for the regulation of protein structure and function. Understanding and controlling the molecular mechanism of protein post-translational modifications are required to modulate protein degradation or function. Small molecules can be used to control these processes. For example, the use of small molecule in control over intein splicing system has been described. The I\textsubscript{N} fragment was fused between the FKBP domain and a target protein, and the I\textsubscript{C} fragment was fused between FRB domain and a second target protein\cite{42,43}. The ability of rapamycin to selectively induce hetero-dimerization of FKBP and FRB led to a regeneration of intein function. Subsequently, the two target proteins were released from the intein and able to be ligated each other via a native peptide bond (Fig. 10).

![Figure 10. Schematic representation of chemically induced intein reassembly using rapamycin-induced dimerization of FKBP/FRB fused with two intein halves, thereby releasing the target proteins.](image)

The use of small molecule induced protein dimerization to control the stability or rescue of proteins in living cells has also been demonstrated. Crabtree and coworkers described that FRB\textsuperscript{'}-not only bound to FKBPI2 in the presence of a rapamycin analog (C20-MaRap) but also bore functionality which conferred reversible instability on the fusion proteins. In the
absence of rapamycin analog, the glycogen synthase kinase-3β (GSK-3β) fused to FRB’ (GSK-3βFRB’) was rapidly degraded. Interestingly, C20-MaRap induced dimerization of FKBP12 and GSK-3βFRB’, which might lock FRB’ in a folded state, resulting in the stabilization of the GSK-3β protein. This system may provide a means to control the stability or degradation of target proteins. Another robust approach to rescue proteins from the proteasome is by using both small molecule induced dimerization and splicing of ubiquitin hydrolysis. For example, split ubiquitin for the rescue of function (SURF) was based on the complementation of genetically split ubiquitin under the control of rapamycin-induced dimerization of FRB and FKBP. The strategy was as follows: (1) the FKBP was fused to the N-terminal fragment of ubiquitin to form FKBP-UbN, (2) the C-terminal fragment of ubiquitin was fused to a protein of interest and FRB to form the FRB-UbC-protein complex, and subsequently fused to a degradation signal (degron) which would cause degradation of the fusion protein by proteasome recognition. In the absence of rapamycin, the proteasome recognizes and “kills” the protein fusion through degron domain recognition, thereby promoting degradation of protein fusion. Addition of rapamycin caused dimerization of FRB and FKBP, which resulted in the reassembly and function of ubiquitin, thereby releasing the protein of interest from the degron and rescuing its function.

1.2.3 Signal transduction/kinase activation

Chemically induced protein dimerization has also been investigated to control the activation of certain kinase family members, in order to study signal transduction. This general approach was also based on the ability of small cell-permeable molecules to recognize and simultaneously bind to specific protein binding domains. To understand the molecular mechanism, Schreiber and coworkers generated a membrane-associated protein fused to the FKBP domain which localized in the plasma membrane. The other rapamycin/rapamycin analog-binding domain (FRB or FRB’) was fused to a kinase of interest. Addition of rapamycin or rapamycin analog led to the recruitment of the protein kinase to the plasma membrane via chemical interaction of the FKBP-rapamycin-FRB or FKBP-rapamycin analog-FKB’, thereby enabling kinase activation resulting in signal transduction. Furthermore, the architecture of coumermycin induced GyrB dimerization has also been used to investigate the molecular mechanism of protein kinase Raf activation. The Raf protein was genetically fused to the GyrB domain in which the activation of the GyrB-Raf fusion protein was determined by assaying a haemagglutinin-tagged version of the Raf target MAP kinase (HA-MEK) for activity (Fig. 11). To provide a membrane-anchored form of GyrB, GyrB was fused to a transmembrane domain (HγR, human interferon-γ receptor). Coumermycin acts as a chemical dimerizer of GyrB fusion protein leading to the formation of the protein heterodimer Raf-GyrB/coumermycin/GyrB-HγR, which is located at the plasma membrane, while the protein homodimer 2Raf-GyrB/coumermycin remained at
the cytoplasm. In both cases, Raf was active, thereby stimulating the activity of HA-MEK.

Figure 11. Coumermycin-induced Raf dimerization and activation. Introduction of Raf protein to GyrB allowed coumermycin to induce Raf-GyrB/GyrB-HyR heterodimer which stayed in the plasma membrane and induce Raf-GyrB homodimer which stayed in the cytoplasm. Raf was active in dimeric form, thereby stimulating the activity of HA-MEK.

The synthetic (CsA)$_2$ molecule with the capacity for bifunctional binding can induce protein cyclophilin dimerization, which was used to control programmed cell death[36]. A membrane protein responsible for signal transduction of programmed cell death was generated by fusing cyclophilin between the cytoplasmic tail of the Fas receptor and the myristoyl group (Fig. 12A). The myristoyl group enabled the cyclophilin-Fas complex to localize on the cell membrane of Jurkat cells via the Fas tail. Upon addition of the cell permeable (CsA)$_2$, dimerization of cyclophilin occurred leading to dimerization of the Fas tail. The formation of the dimeric Fas tail activated the Fas signaling pathway, leading to cell death. Another small molecule – FK1012 – induces homo-dimerization of FKBP and can also be used to gain control over programmed cell death. The Fas cytoplasmic domain was fused between poly FKBP and myristoyl which located on the cell membrane (Fig. 12B). The presence of FK1012 mediated the aggregation of the Fas-poly FKBP receptor leading to activation of Fas signaling transduction, and eventually cell death[47].

Protein dimerization induced by the specific binding of cell permeable high affinity small natural products or synthetic molecules represents a power tool for controlling homo- or hetero-dimerization of proteins in numerous biological processes such as gene expression, post-translation and signaling transduction. Thus, the small molecule approach shows very high promise for drug development and other biomedical applications. However, this approach does require the construction of large fusion proteins, in which the required protein domains, such as FKBP, GyrB and cyclophilin, contribute substantial mass to the final protein construct, potentially affecting the biological activity of the target protein.
Therefore, alternative methods for molecular control over protein dimerization are additionally required. Metal ion and supramolecular host-guest chemistry mediated control over protein dimerization might address this.

**Figure 12.** The pathway of programmed cell death: A) (CsA)$_2$ induced dimerization of Fas-cyclophilin receptor. Cyclophilin was fused between myristoyl and Fas domain, the presence of (CsA)$_2$ induced dimerization of Fas-cyclophilin receptor led to activation of the Fas domain, resulting in cell death$^{[36]}$; B) FK1012 induced dimerization of the Fas-FKBP receptor. The Fas domain was fused between myristoyl (M) and FKBP domain. FK1012 induced dimerization of Fas-FKBP receptor led to activation of the Fas domain, resulting in cell death$^{[47]}$.

### 1.3 Metal ion-mediated protein dimerization

Metal ion-mediated protein dimerization has recently been demonstrated$^{[48-52]}$. Tezcan and coworkers for example generated hybrid coordination motifs based on the simultaneous binding of metal ion to a natural Histidine amino acid (His) and non natural ligand (quinolate) on the α helical surface of protein cytochrome cb$_{562}$$^{[53]}$ (Fig. 13). The 5-amino-8-hydroxyquinoline (Quin), which binds metal ions with high affinity, was covalently ligated to cysteine at position 70 of cytochrome cb$_{562}$. Thus binding of metal ions such as Ni$^{2+}$, Co$^{2+}$, Cu$^{2+}$ and Zn$^{2+}$ induced cytochrome cb$_{562}$ dimerization resulting in an increase of global protein stability. Interestingly, metal ion-induced protein dimerization was used to generate a structural superposition closely resembling bZip-type transcription factors, suggesting potential applications for the recognition of biological targets. In addition, the use of metal-templated ligand synthesis to induce protein dimerization represents a promising approach to controlling cell signaling or gene expression events$^{[48]}$. 
1.4 Supramolecular inducers of protein dimerization

Supramolecular chemistry has initially been inspired by biomolecules and their higher orderd structure. Recently, first example of supramolecular used to modulate and control protein dimerization has been reported. Supramolecular systems incorporating natural or synthetic components have been engineered with desirable properties for use in biochemical research, such as water solubility\(^ {54,55}\), guest-specific binding, and low-toxicity\(^ {56}\). In contrast to small molecule inducers of dimerization which bind to protein domains, supramolecular chemistry-based host systems typically bind much smaller cationic and neutral guests. Supramolecular induced protein dimerization is fundamentally based on the noncovalent interaction of supramolecular hosts with guest molecules. Two supramolecular host molecules, cyclodextrin and cucurbit[8]uril, have been widely explored as tools for the selective and reversible control over protein dimerization in both buffered solution and living cells\(^ {57,58}\). The natural product cyclodextrin and its analogs are sugar-based non-symmetrical host molecules, which can selectively bind a hydrophobic guest to form a complex with a 1:1 stoichiometry in aqueous solution. For example, the cavity of the β-cyclodextrin variant recognizes and binds lithocholic acid with high affinity (\(K_a = 10^9 \text{ M}^{-1}\)) (Fig. 14),\(^ {59}\) which opens up the possibility for β-cyclodextrin to recognize and bind protein-lithocholic acid conjugates.

Figure 13. Crystal structure of Ni\(^{2+}\)-mediated cytochrome cb\(^{562}\) (HQuin1) dimer. Figure reproduced with permission from \(^ {53}\).
**Figure 14.** \(\beta\)-cyclohexextrin recognizes and binds lithocholic acid in its cavity with high affinity \((K_d = 10^{-6} \text{M})\).

\(\beta\)-cyclohexextrin

\[
\begin{align*}
\text{Buffer solution} & \\
\beta\text{-cyclohexextrin/ Lithocholic acid complex (}\ K_d = 10^{-9} \text{M})
\end{align*}
\]

Cucurbit[8]uril is the eight membered homologue of the cucurbit[n]uril family of glycoluril based macrocycles (Fig. 15A). Cucurbit[8]uril has shown highly attractive biochemical applications due to its good water-solubility, low toxicity, and capacity to bind various cationic guest molecules\[^{60-62}\]. The cavity of cucurbit[8]uril is sufficiently large to bind two synthetic guest molecules simultaneously such as two equivalents of \(N\)-phenylpiperazine, aminoacridizinium, naphthyl derivative, coumarin and neutral red under acidic conditions\[^{60}\]. The favorable recognition of two guests by cucurbit[8]uril allows to form highly stable ternary complexes in aqueous solution. For example, Urbach and coworkers demonstrated the recognition of two tripeptide phenylalanine-glycine-glycine (FGG) by cucurbit[8]uril to form a 1:2 stoichiometric complex with high affinity \((\text{ternary } K = 10^{11} \text{ M}^{-2})\[^{63}\] with a key interaction between the N-terminal amine functionality of the peptide and the rim of the cucurbit[8]uril cavity (Fig. 15B). In addition to binding two peptides in a 1:2 complex, cucurbit[8]uril also binds an electron-deficient guest molecule such as methylviologen, and an electron-rich guest molecule, such as dihydroxynaphthalene or naphthalene derivatives, simultaneously to form a stable \((1:1:1)\) ternary complex\[^{64,65}\] (Fig. 15C). The selective noncovalent interaction of cucurbit[8]uril and guest elements provides a means to reversibly control dimerization of proteins incorporating these supramolecular guest elements. The easy design and synthesis of supramolecular host-guest systems opens up the possibility of modulating and controlling protein dimerization.
1.4.1 Cyclodextrin-induced protein dimerization

The applicability of supramolecular induced protein-protein interaction was first demonstrated using supramolecular host-guest elements attached to synthetic peptides\[^{67}\]. The specific recognition of the adamantyl group by $\beta$-cyclodextrin allowed a $\beta$-cyclodextrin-conjugated synthetic peptide to selectively recognize and bind an adamantane-conjugated peptide. This self-assembled peptide dimer enabled strong and selective DNA recognition. DNA recognition by supramolecular peptide dimerization could be reversed by inhibiting the supramolecular dimerization with either free $\beta$-cyclodextrin or adamantane. For supramolecular protein dimerization, a pair of enhanced cyan fluorescent protein (eCFP) and enhanced yellow fluorescent protein (eYFP) was firstly selected as the model proteins due to their facilitated manipulation and well-defined structures. These fluorescent proteins can be conjugated to the $\beta$-cyclodextrin and lithocholic acid host guest system at the C-terminus of the proteins. A high affinity and selective recognition of lithocholic acid by $\beta$-cyclodextrin enabled the association of the two fluorescent proteins. In this case, the degree of protein association was monitored by donor-receptor energy transfer\[^{57}\] (Fig. 16). In principle, an optimized synthetic host-guest complex would enhance the interaction of host-guest conjugated proteins and would thus be attractive from the point of view of studying protein-protein interactions at lower concentration. For example, in order to increase the stability of the supramolecularly induced protein dimerization complex, the $\beta$-cyclodextrin host was modified to heptakis-[6-deoxy-6-(2-aminoethyl-sulfanyl)]-$\beta$-cyclodextrin. This molecular upgrading of the $\beta$-cyclodextrin side-chains brought about a 10 fold increase in binding to lithocholic acid\[^{68}\]. The optimization of synthetic host-guest systems is not a unique approach to increase the affinity of protein-protein interaction: engineering of the dimeric interface of proteins has also been used to increase and stabilize the supramolecular protein dimerization.
complexes. For example, point-mutated fluorescent proteins (S208F and V224L), which normally show weak intrinsic affinity for dimerization, formed strong and stable supramolecular protein complexes on ligation of host-guest elements \[^{[69]}\]. To prove the concept of reversible supramolecular protein dimerization, addition of β-cyclodextrin to the mixture containing β-cyclodextrin-appended eYFP and lithocholic acid-appended eCFP (as a competitor for lithocholic acid) resulted in inhibition of protein dimerization \[^{[57;69;70]}\].

![Schematic representation of cyclodextrin-induced assembly and disassembly of eYFP, eCFP functionalized with cyclodextrin and lithocholic acid, respectively. Image reproduced from reference \[^{[57]}\].](image)

1.4.2 Cucurbit[8]uril-induced protein dimerization

Another attractive supramolecular host molecule for use in reversible protein dimerization studies is cucurbit[8]uril. Unlike cyclodextrin, cucurbit[8]uril selectively binds and dimerizes two guests simultaneously within its hydrophobic cavity, and with high affinity\[^{[63;64]}\]. Control over protein dimerization by the cucurbit[8]uril is therefore a highly desirable approach using proteins that only incorporate guest molecules. For example, cucurbit[8]uril recognizes and binds methylviologen (MV) and naphthalene simultaneously (Np), thereby forming a ternary cucurbit[8]uril:MV:Np complex, which could induce hetero-dimerization of CFP with YFP. In this case, CFP and YFP were chemically provided with Np and MV, resulting in CFP-Np and YFP-MV, respectively. The presence of cucurbit[8]uril mediated hetero-dimerization of CFP-Np with YFP-MV, resulting in an increase of energy transfer from donor CFP to acceptor YFP\[^{[71]}\] (Fig. 17).
The recognition of small guest molecules by supramolecular hosts is an orthogonal approach to reversibly controlled protein dimerization. However, the application of supramolecular induced protein dimerization is currently still limited to the reversible controlled dimerization of fluorescent proteins. Control over other proteins, such as membrane protein dimerization, transcription factors, and enzyme activation by a supramolecular based system needs to be established to enable the study of increasingly complex systems in biomedical applications. Additionally, the selective recognition of natural protein elements by supramolecular host-molecules would prevent the need for protein conjugation reactions and provides a very versatile entry to the application of supramolecular chemistry in biology.

1.5 Conclusions

Protein dimerization plays a key role in almost all biological processes. Control over protein dimerization using small molecules is an important concept for studying the fundamental underlying molecular processes. The use of external molecules to induce protein dimerization in part overcomes the limitations of protein engineering approaches. Indeed, the use of small natural or artificial molecular dimerizers has led to an improved
understanding of the role of protein-protein interactions in biological processes such as transcriptional coactivators, as well post translation processing and signaling transduction events. Recently, control over protein dimerization has also been achieved using metal ions and supramolecular approaches. In particular, the selective recognition of small guests by synthetic host molecules to form 1:1 complexes or 1:2 ternary complexes enables the reversible control of protein dimerization using proteins prefunctionalyzed with small guest elements. The field of supramolecular induced dimerization represents an orthogonal approach for studying functional protein dimerization, for example, in the case of enzyme activation, transcriptional cofactors and membrane proteins in different cellular settings.

1.6 Aim and outline of this thesis

Understanding and controlling the molecular mechanisms that regulate protein dimerization is a fundamental aspect of biomedical research with numerous applications. Unfortunately, current methods for the external control over these interactions are still only achievable for limited cases. A robust and versatile supramolecular approach could, therefore, provide a strong point of entry to explore and modulate the molecular mechanisms of protein dimerization. The aim of this thesis is the use of supramolecular cucurbit[8]uril-induced dimerization as a powerful tool to reversibly control the dimerization and activation of proteins featuring genetically encoded peptide tags. Chapters 2 and 3 reveal that the supramolecular host molecule cucurbit[8]uril can be efficiently used to induce and reversibly control the dimerization of fluorescent proteins incorporating a genetically encoded N-terminal phenylalanine-glycine-glycine (FGG) peptide motif. The proteins with an FGG-tag are easily generated using autocleavage of an intein system under pH and temperature control. Cucurbit[8]uril selectively binds and dimerizes the FGG-tag in its hydrophobic cavity and via a key interaction between the N-terminal amine functionality of the peptide and the carbonyl rim of cucurbit[8]uril (Fig. 18). In search of a proof of principle, we studied the cucurbit[8]uril-induced dimerization of fluorescent proteins (CFP and YFP variants) using Förster resonance energy transfer (FRET), size exclusion chromatography, dynamic light scattering (DLS) and small angle X-ray scattering (SAXS). Cucurbit[8]uril-induced dimerization of proteins with an FGG-tag via a supramolecular host-guest noncovalent interaction could be reversed through addition of a small synthetic competitor (methyl viologen). The novel concept of cucurbit[8]uril-induced protein tetramerization is described in chapter 4. The combination of dimeric protein interface and a genetically encoded FGG motif at the N-terminus of a protein allowed for the use of cucurbit[8]uril to induce protein tetramerization, which was analyzed by size exclusion chromatography and DLS.
Figure 18. Schematic model of protein homodimer is formed by noncovalent supramolecular interaction between Q8 and a genetically encoded N-terminal FGG motif.

Chapter 5 describes control over enzyme dimerization and activation using a synthetic supramolecular molecule as a novel and attractive strategy for biological investigations. There is a pressing need to understand the molecular mechanisms of enzyme activation for many potential applications. Therefore, supramolecular chemistry represents a power tool to understand the role of enzyme dimerization, which can be the stimulus for enzyme activation. In this chapter, we focus on a cucurbit[8]uril-induced protein dimerization approach which is utilized to reversibly control dimerization and activation of caspase-9 in buffer solution. Cucurbit[8]uril can selectively recognize and dimerize engineered caspase-9 bearing a genetically encoded N-terminal FGG motif (Fig. 19), which leads to a significant increase in enzymatic activity. The significant increase in activity of cucurbit[8]uril-induced caspase-9 dimerization also supports the concept of an activation mechanism for enzymes, which relies mainly on dimerization. Furthermore, the high activity of cucurbit[8]uril-induced caspase-9 can be fully reversed using a competitor ligand, illustrating the concept of reversible control over enzyme dimerization and activation.

Chapter 6 shows the regulation of caspase activity using synthetic supramolecular chemistry to be an excellent tool for biomedical research. The use of cucurbit[8]uril as a molecular dimerizer not only increased the enzymatic activity of caspase-9 (in chapter 5) but also fully regenerated enzymatic activity of inactive mutants of caspase-8. In this chapter, we focus on deactivity of caspase-8 by engineering of a mutant monomeric caspase-8, and full regeneration of the mutant caspase-8 activity by cucurbit[8]uril-induced dimerization. A synthetic substrate (Ac-IETD-AFC) and a natural substrate (caspase-3) were used to determine the enzymatic activity of the mutant caspase and its cucurbit[8]uril-induced dimeric form. The mutant showed no affinity with caspase-3 resulting in incompetent cleavage of caspase-3. Upon addition of cucurbit[8]uril full enhancement of catalytic activity was observed. These results again reiterate the importance of correct protein dimerization for active site-rearrangement and activation of the enzymes and the power of a supramolecular approach which allosterically stabilizes the protein dimer.
Figure 19. Cucurbit[8]uril induced caspase-9 dimerization. A) Schematic representation of N-terminal FGG-bearing monomeric caspase-9 (large and small subunits) and its dimerization into an enzymatically active homodimer via supramolecular-induced host-guest complexation with cucurbit[8]uril. b) Crystal structure representation of a caspase-9 dimer (2AR9) [18], portraying the close localization of the two N-termini of the two large subunit.

The final part (chapter 7) of the investigation focuses on supramolecular induced dimerization of membrane proteins in living cells. The results from these explorative studies offer promising indications that the previously discussed concept might indeed be applicable to proteins in a cellular environment, making cucurbit[8]uril-induced protein dimerization a strong molecular concept in biomedical research. In this chapter, we first generated membrane proteins, EGFR (epidermal growth factor receptor), consisting of an N-terminal FGG motif located at the extracellular membrane. Cucurbit[8]uril-induced EGFR dimerization, by the putative supramolecular interaction between cucurbit[8]uril and the genetically coded FGG motif, might result in changes in localization of EGFR at the plasma membrane. Ongoing studies on cucurbit[8]uril-induced dimerization of fluorescent model proteins will provide further insights in the potential of membrane protein assembly using supramolecular chemistry.
1.7 References

Molecular control over protein dimerization; opportunities for supramolecular chemistry


Chapter 2

Protein dimerization induced by supramolecular interactions with cucurbit[8]uril

Abstract

The supramolecular host molecule cucurbit[8]uril can be used to reversibly control the dimerization of proteins bearing a genetically incorporated N-terminal phenylalanine-glycine-glycine (FGG) peptide motif. Proteins having this motif are genetically generated by using an N-terminal intein system. The specific recognition of proteins bearing the FGG peptide motif allows cucurbit[8]uril to induce protein homodimerization and heterodimerization. The cucurbit[8]uril induced protein dimer is stable and can be separated on a size exclusion chromatography column. The supramolecular interaction between cucurbit[8]uril and two FGG peptide motifs can be reversed by addition of a small competitor ligand (methyl viologen), which leads to inhibition of protein dimerization. The combination of a very short peptide motif with the easy accessibility of both the genetically encoded proteins and the synthetic supramolecular molecules holds great promise for applications as supramolecular inducers of protein dimerization in biological processes.

This work has been published: Protein dimerization induced by supramolecular interactions with cucurbit[8]uril. Hoang D. Nguyen, Dung T. Dang, Joost L.J van Dongen, and Luc Brunsveld. Angew. Chem. Int. Ed. 2010, 122, 907-910
2.1 Introduction

The combination of supramolecular chemistry with proteins provides opportunities to reversibly control protein properties. Proteins have for example been decorated with synthetic supramolecular guest-ligands for immobilization on host-molecule modified surfaces\(^1\)\(^-\)\(^2\), and vesicles decorated with β-cyclodextrin were used as platform to immobilize adamantane-modified enzymes\(^3\). Supramolecular dimerization of two peptides enabled sequence-specific recognition of DNA,\(^4\)\(^-\)\(^5\) and a β-cyclodextrin based host-guest system allowed controlling the functional reassembly of two split-Green Fluorescent Protein (GFP) fragments\(^6\). The supramolecular interaction of β-cyclodextrin with lithocholic acid, both attached to protein C-termini, enabled selective induction of protein hetero-dimerization in solution and in cells\(^7\)\(^-\)\(^8\).

![Figure 1. Schematic representation of supramolecular-induced protein homodimerization of two monomeric yellow fluorescent proteins with an N-terminal FGG peptide motif by cucurbit[8]uril.](image)

Control over protein dimerization with synthetic molecules is highly desirable, and the approaches with small molecules binding to specific, additionally incorporated, protein domains have been very successful\(^9\)\(^-\)\(^12\). Control over protein dimerization via a very small, genetically encoded, peptide motif is an attractive target in this respect as it would require only very minor protein modifications. Synthetic supramolecular molecules such as cyclodextrins\(^13\)\(^-\)\(^15\) and cucurbituril\(^15\)\(^-\)\(^18\) have been shown to selectively recognize amino acids, and synthetic receptors have been developed to recognize protein elements\(^19\)\(^-\)\(^22\), some of which with a high degree of affinity and selectivity\(^23\). However, synthetic supramolecular host molecules have not been applied thus far as inducers of protein dimerization. Here we show that the supramolecular host molecule cucurbit[8]uril allows to induction and reversible control of dimerization of proteins with a genetically incorporated N-terminal phenylalanine-glycine-glycine (FGG) peptide motif (Fig. 1). The FGG peptide tag is easily incorporated via molecular biology techniques, and the induction of protein dimerization is stable and can be reversed with a synthetic supramolecular guest molecule.
2.2 Results and discussion

2.2.1 Construction of proteins

The FGG tripeptide has been shown to bind as a dimer into one cucurbit[8]uril molecule, with an important interaction of the N-terminal amine functionality of the peptide with the cucurbit[8]uril carbonyl rim\textsuperscript{[23]}. We therefore incorporated this motif at the N-terminus of proteins as recognition element to induce protein dimerization with cucurbit[8]uril. Proteins typically are expressed with an N-terminal methionine. Therefore an intein-based protein expression was used to generate proteins with an N-terminal phenylalanine. Two monomeric fluorescent proteins\textsuperscript{[24]} incorporating the FGG motif (Cyan Fluorescent Protein FGG-mCFP and Yellow Fluorescent Protein FGG-mYFP) were designed and generated. The monomeric character of the proteins ensures absence of unspecific protein aggregation.\textsuperscript{[7,8,24]} Apart from an N-terminal intein domain for post-expression auto-cleavage, the proteins featured an N-terminal Chitin Binding Domain (CBD) for purification. The FGG-proteins were expressed at 15 °C, and subsequently the N-terminal intein-tag was released from the target proteins via overnight incubation at room temperature and pH 7.0. Two fluorescent reference proteins featuring an N-terminal methionine were also generated (Met-CFP and Met-YFP). All proteins were obtained pure as evidenced by SDS-PAGE and LC-ESI-MS (Fig. 2 and Table 1).

Figure 2. A) Generation of proteins bearing an N-terminal FGG motif by intein splicing. B) SDS-PAGE analysis of proteins Met-CFP, Met-YFP, FGG-mCFP and FGG-mYFP.
2.2.2 Cucurbit[8]uril induced protein homodimerization

We first studied the cucurbit[8]uril induced homodimerization of FGG-mYFP (Fig. 1) with homo Förster Resonance Energy Transfer\[^{25}\] (homo-FRET) studies (Fig. 3). The homodimerization of two fluorescent proteins typically results in a decrease of the fluorescence anisotropy, due to intermolecular energy transfer. Indeed, upon the addition of cucurbit[8]uril to a solution of FGG-mYFP, the anisotropy of the YFP fluorescence decreased. The anisotropy decrease is cucurbit[8]uril concentration dependent, as expected for a supramolecular recognition process. The cucurbit[8]uril thus binds to the proteins and induces their dimerization. Addition of a large excess of cucurbit[8]uril results in a small increase of the fluorescence anisotropy, possibly resulting from a less optimal protein orientation for FRET due to low affinity cucurbit[8]uril binding to amino acid side-chains.

![Graph showing fluorescence anisotropy titration of cucurbit[8]uril to a 2 μM FGG-mYFP solution in 10 mM phosphate buffer at pH7, observed by using a 500 nm excitation wavelength and scan spectra over a range of emission wavelengths (524-532 nm)](image)

**Figure 3.** Fluorescence anisotropy titration of cucurbit[8]uril to a 2 μM FGG-mYFP solution in 10 mM phosphate buffer at pH7, were observed by using a 500 nm excitation wavelength and scan spectra over a range of emission wavelengths (524-532 nm)

2.2.3 Size exclusion chromatography and ESI-MS measurement

The supramolecular induced protein dimerization could also be proven by size exclusion chromatography (Fig. 4A) and mass spectrometry (Fig. 4B). Addition of increasing amounts of cucurbit[8]uril to a solution of FGG-mYFP resulted in the appearance of a second population at higher molecular weight, which increased with increasing cucurbit[8]uril concentration. This second species features a molecular weight (58,624 Da) which corresponds to exact the molecular weight of 2 FGG-mYFP and 1 cucurbit[8]uril complex (58,623 Da). The ternary complex coexists next to the monomer form of the protein at intermediate cucurbit[8]uril concentrations. These results not only show that the cucurbit[8]uril binds two FGG motifs and induces protein dimerization, but also that the induced protein dimer is stable and can be separated on a purification column.
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril

Figure 4. A) Size exclusion chromatograms on a Superdex 200 column of a 20 μM FGG-mYFP solution with increasing concentrations (0, 2, 5, 10, and 20 μM) of cucurbit[8]uril in 10 mM phosphate buffer pH 7. B) Mass spectrometry measurements of (FGG-mYFP):cucurbit[8]uril ternary complex (mass calculated 58623 Da, mass determined 58624 Da) in 10 mM phosphate buffer pH 7.

2.2.4 Cucurbit[8]uril-induced protein heterodimerization

To further evaluate and investigate the scope of the cucurbit[8]uril induced FGG-protein dimerization, protein heterodimerization studies with CFP and YFP variants were performed (Fig. 5a). Addition of cucurbit[8]uril to an equimolar mixture of FGG-mCFP and FGG-mYFP resulted in the occurrence of a strong hetero-FRET (Fig. 5b). The strong energy transfer is even more notable, considering that the supramolecular induced protein heterodimerization in this case is most probably accompanied by 50 percent homodimers, which do not contribute to the hetero-FRET. The hetero-FRET system was used to evaluate the selectivity of the cucurbit[8]uril induced FGG-protein dimerization over proteins with an N-terminal methionine. The addition of cucurbit[8]uril to a solution of FGG-mCFP and FGG-mYFP resulted in an increase of the 527 nm/475 nm ratio from 0.46 to 2.73 (Fig. 5c). Addition of the same amount of cucurbit[8]uril to a mixture of two N-terminal methionine reference proteins did not result in an increase of the 527 nm/475 nm ratio (from 0.48 vs. 0.46). Neither was an increase in the FRET ratio observed in a mixture of a CFP protein with a methionine and a YFP protein with an FGG motif (from 0.45 vs. 0.50). The cucurbit[8]uril induced protein dimerization is thus selective for proteins featuring the FGG motif over proteins with an N-terminal methionine. These results additionally show that the protein dimerization does not result from unselective interactions of the cucurbit[8]uril with amino acid side chains at the periphery of the protein.
Figure 5. FRET studies with CFP-YFP protein pairs under the control of cucurbit[8]uril. a) schematic model of the supramolecular induced protein dimerization; b) representative spectra of a FGG-mCFP and FGG-mYFP mixture (both at 1 μM) in the absence (grey) and presence (dark) of cucurbit[8]uril (1.5 μM) by using a 410 nm excitation wavelength; c) comparison of the 527 nm/475 nm FRET ratios observed without (grey bars) and with (dark bars) cucurbit[8]uril (1.5 μM) for different protein mixtures (both at 1 μM) in 10 mM phosphate buffer pH 7.

2.2.5 Reversible inducer of protein dimerization

Supramolecular induction of protein dimerization provides the opportunity to non-covalently reverse the induced protein dimerization, with a small synthetic molecule (Fig. 6a). Methyl viologen, or paraquat[26], was therefore titrated to a solution of cucurbit[8]uril induced fluorescent protein heterodimers. Upon addition of the competitive methyl viologen ligand, a rapid decrease of the hetero-FRET effect was observed (Fig. 6b, c), resulting in the restoration of the original CFP fluorescence (Fig. 6b, dark line), devoid of energy transfer to YFP. These results show that the FGG motif can be displaced by methyl viologen, thus accomplishing a complete reversal of the supramolecular induced protein dimerization with a bio-orthogonal supramolecular ligand.
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril

2.3 Conclusions

The results reported in this chapter show that a very short, genetically-encoded, N-terminal FGG peptide motif allows control over protein dimerization with the supramolecular host molecule cucurbit[8]uril as inducer of dimerization in buffered solution. The supramolecular protein dimerization is significantly stable to allow the protein dimers to be separated by size-exclusion chromatography, and the recognition is selective for the FGG-motif over a classical N-terminal methionine. Finally, the supramolecular induced protein dimerization can efficiently be reversed with a small synthetic ligand that competes with the peptide motif for cucurbit[8]uril binding, allowing switching off the protein dimerization. The combination of the attractively, very short peptide motif with the easy accessibility of both the genetically encoded proteins and the synthetic supramolecular molecules holds great promises for the application as supramolecular inducers of protein dimerization, both for protein homodimerization, such as for example observed in dimerizing enzymes and membrane proteins, and for the stabilization of weakly associating protein heterodimers.
2.4 Experimental section

2.4.1 Construction of plasmids

The fluorescence proteins with N-terminal FGG motif were generated by using the N-terminal intein from plasmid pTWIN-1 (New England Biolabs). DNA encoding for FGG-mCFP was amplified by PCR using pHT476\(^7\) as a templates and primers ON179, 5’-gcc tgc tct tcc aac ttt ggt ggc ggc agc cac cat cac cat cac c-3’ and ON180, 5’- ggc cat gga tcc tta ctt gta cag ctc gtc cat gcc gga -3’. DNA encoding for FGG-mYFP was amplified using pHT486\(^7\) as a template and primers ON181, 5’- ggc tgc tct tcc aac ttt ggt ggc gca agc tgg agc cac ccg cag-3’ and ON180. These two SapI- and BamHI-treated PCR fragments were ligated into pTWIN1 at SapI and BamHI using T4 DNA ligase resulting in plasmids pHT582 and pHT584, respectively (Fig. 7).

*Figure 7. Construction of protein expression plasmid pHT584 (FGG-mYFP).*
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril

The plasmids carrying DNA encoding for fluorescence proteins with methionine at the N-terminus were constructed as follows. DNAs encoding for CFP and YFP were amplified by PCR using ON175, 5’-aaa gga ggg gtc tcc cat ggt gag caa ggg cga gga gc-3’ and ON180, 5’-ggc cat gga tcc tta ctt gta cag etc gtc cat gcc gag-3’ with template pHT477 and pHT482, respectively. The Nco/BamHI-treated CFP fragment was cloned in pHT449 at NcoI and BamHI to fuse with His-tag at N-terminus resulting in pHT559. The Nco/BamHI-treated YFP fragment was cloned in pHT531 at NcoI and BamHI to fuse with a Strep-tag at the N-terminal resulting in pHT560. Plasmid pHT449 was constructed by removing the N-terminal CBD-intein part and introducing hybridized oligos, ON071, 5’-tat gga agc gag cca cca tca cca tca cca tca tgc-3’ and ON072, 5’-agc atg atg atg atg atg atg gct cgc ttc ca-3’ encoding for MEASHHHHHHHSHKQGEELFTGVVPILVELGDVNGHKFSVSGEGDATYGGKTLKFICTTGGTPWPWTLLTTLWTVQCFSRYPDHMKQHDFKFSAMPEGYQERTIFFKDPGNKYKTRAEVKFDGTLVNRKIGIDFKEDGINLGHKLEYNYISHVYITADQKNGIKANFKirHniedgsvqladhyqqntpigaqvpvllpdpnhy1sqtqsklkdpekrmdtv11efvtaagitlmgde1yk

Amino acid sequences:

**FGG-mCFP**

\[
\text{FGGASHHHHHHHMSVSKGEELFTGVVPILVELGDVNGHKFSVSGEGDATYGGKTLKFICTTGGTPWPWTLLTTLWTVQCFSRYPDHMKQHDFKFSAMPEGYQERTIFFKDPGNKYKTRAEVKFDGTLVNRKIGIDFKEDGINLGHKLEYNYISHVYITADQKNGIKANFKirHniedgsvqladhyqqntpigaqvpvllpdpnhy1sqtqsklkdpekrmdtv11efvtaagitlmgde1yk}
\]

**FGG-mYFP**

\[
\text{FGGASWSHPQFEKSAMVSKGEELFTGVVPILVELGDVNGHKFSVSGEGDATYGGKTLKFICTTGGTPWPWTLLTTLWTVQCFSRYPDHMKQHDFKFSAMPEGYQERTIFFKDPGNKYKTRAEVKFDGTLVNRKIGIDFKEDGINLGHKLEYNYISHVYITADQKNGIKANFKirHniedgsvqladhyqqntpigaqvpvllpdpnhy1sqtqsklkdpekrmdtv11efvtaagitlmgde1yk}
\]

**Met-CFP**

\[
\text{MEASHHHHHHHHAMSVSKGEELFTGVVPILVELGDVNGHKFSVSGEGDATYGGKTLKFICTTGGTPWPWTLLTTLWTVQCFSRYPDHMKQHDFKFSAMPEGYQERTIFFKDPGNKYKTRAEVKFDGTLVNRKIGIDFKEDGINLGHKLEYNYISHVYITADQKNGIKANFKirHniedgsvqladhyqqntpigaqvpvllpdpnhy1sqtqsklkdpekrmdtv11efvtaagitlmgde1yk}
\]
**Met-YFP**

MESAWSHPQFEKAMVSGEEELFTGVVIPWELDGDVNGHKFSVSGEGGDATYGKLTLKFICTTGKLVPWPTLVTTFYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLVNRIELKGIDFKEDGNILGHKLEYNYNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIDGPVLPLDNHLYLQSAKDPNEKRDMVLLFVTAAAGITLGMDELYK

### 2.4.2 Protein expression and purification

The plasmids pHT559 (Met-CFP), pHT560 (Met-YFP), pHT582 (FGG-mCFP) or pHT584 (FGG-mYFP) were transformed into *E. coli* strain BL21(DE3). The bacteria were cultured in LB medium containing 100 μg/mL of ampicillin, and the cells were grown at 37 °C, shaking at 250 rpm to an A₆₀₀ of 0.5, then IPTG was added to a final concentration of 0.5 mM. The cells were continuously incubated overnight at 15 °C, shaking 250 rpm before being harvested. The pellet was resuspended into the BugBuster protein (Novagen) extraction reagent plus bezonase nuclease (to degrade DNA and RNA), and the insoluble material was removed by centrifugation at 20,000 rpm for 30 minutes at 4 °C. The soluble fraction was applied to a 15 mL column filled with chitin beads (New England Biolabs) for FGG-mCFP and FGG-mYFP through gravity flow, and the column was washed with 20 volumes of sodium phosphate buffer 10 mM, pH 7 and subsequently overnight incubated at room temperature. The column was then eluted with sodium phosphate 10 mM, pH 7 buffer, and the target protein was collected. In case of Met-CFP and Met-YFP, the soluble fractions were loaded on the columns containing Ni-NTA Superflow or Strep-Tactin Superflow, respectively and purification steps were performed following the manuals (Qiagen). All the proteins were dialysed in the phosphate buffer, concentrated using Amicon Ultra 10,000 NMWL (Milipore) and kept at -80 °C until use. The target proteins were analyzed by SDS-PAGE and ESI-MS (Fig. 2B, 8).

### 2.4.3 LC-ESI-MS analysis of proteins

The LC-ESI-MS experiments were performed using a LCQ Fleet (Thermo Scientific). Proteins went through the C₄ column under elution with a water/acetonitrile gradient (5% water: 95% acetonitril for 1 minute to 70% water: 30% acetonitril in 9 minutes and stay at 70% water: 30% acetonitril for 2 minutes) with 0,1 % formic acid before they were introduced to the ion source and mass spectrometry. The mass of the proteins was deconvoluted by Magtran software (Fig. 8, Table 1)
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril

![Figure 8. LC-ESI-MS of analysis of FGG-mCFP, mass spectrum and deconvoluted mass of FGG-mCFP.](image)

### Table 1. Calculated and measured mass of proteins

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Calculated mass (Da)</th>
<th>Measured mass (Da)</th>
<th>Error (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGG-mCFP</td>
<td>28548</td>
<td>28546</td>
<td>-2</td>
</tr>
<tr>
<td>FGG-mYFP</td>
<td>28646</td>
<td>28652</td>
<td>+6</td>
</tr>
<tr>
<td>Met-CFP</td>
<td>28472</td>
<td>28480</td>
<td>+8</td>
</tr>
<tr>
<td>Met-YFP</td>
<td>28501</td>
<td>28509</td>
<td>+8</td>
</tr>
</tbody>
</table>

#### 2.4.4 Fluorescence spectroscopy measurements

All samples for fluorescence spectroscopy measurements were measured in phosphate buffer (10 mM, pH 7.0) in quartz cuvettes of 10 mm light path (Hellma). The concentration of the proteins was measured with a NanoDrop ND1000 (Nanodrop technologies) and adjusted using the absorbance at 435 nm (A435) and a molar extinction coefficient of 32,500 M⁻¹cm⁻¹ for CFP and by using A515 and a molar extinction coefficient of 84,000 M⁻¹cm⁻¹ for YFP. All of the fluorescence measurements were carried out with a Cary Eclipse fluorescence spectrophotometer (Varian), and all fluorescence data were recorded at 20 °C with an excitation wavelength of 410 nm, slit width of 5 nm. Measurement parameters and sample concentrations were kept constant over all measurements to enable data comparison.
2.4.5 Size exclusion chromatography

The monomeric and cucurbit[8]uril dimerized proteins (FGG-mYFP) were analyzed on a superdex 200 column, 1x30 cm (GE Healthcare) (LC-20AD, Shimadzu). The samples were analyzed and eluted in a sodium phosphate buffer (10 mM, pH 7) at the constant flow rate of 0.1 mL/min, and absorption detection was performed at 514 nm. 40 μl of sample was analyzed in each run, consisting of 20 μM protein and variant concentrations of cucurbit[8]uril.

A SEC calibration curve was generated using the following protein standards (MW-GF-200, Sigma) as markers (Fig. 9):
1. Amylase: 200000 Da
2. Bovine serum Albumin (BSA): 66000 Da
3. Yellow fluorescent protein (YFP): 28647 Da
4. Myoglobin: 17600 Da

Detection of these proteins was accomplished using a PDA detector, scanning from 200 nm- 600 nm. The proteins showed the following retention times:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Retention time (min)</th>
<th>Log (M.W)</th>
<th>M.W (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amylase</td>
<td>12.034</td>
<td>5.30</td>
<td>200000</td>
</tr>
<tr>
<td>BSA</td>
<td>14.081</td>
<td>4.82</td>
<td>66000</td>
</tr>
<tr>
<td>YFP</td>
<td>15.71</td>
<td>4.46</td>
<td>28647</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>18.009</td>
<td>4.25</td>
<td>17600</td>
</tr>
</tbody>
</table>

The data were fitted with a quadratic and the exclusion time of the complex (2 FGG-mYFP + 1 cucurbit[8]uril) complex (x = 14.29 min.) resulted in a napparent molecular weight of 55 Da, which compares very well to the calculated molecular weight (58 Da)

![Figure 9. Calibration curve of size exclusion chromatography.](image)

2.4.6 Size exclusion chromatography-ESI-MS measurements

The FGG-mYFP protein alone (20 μM) (left) and a mixture of cucurbit[8]uril (5 μM) and FGG-mYFP (20 μM) (right) were first run on a superdex 200 column (GE Healthcare) (LC-20AD, Shimadzu) in an ammonium acetate buffer (30 mM, pH 6.5), optimal for MS response, at the constant flow rate of 0.1 mL/min. Detection of the protein and protein complex with
Protein dimerization induced by supramolecular interactions with cucurbit[8]uril were performed using the total ion count of the MS signal (Q-Tof Ultima™ Global, Micromass-UK).

Figure 10. SEC-ESI-MS data. 1) FGG-mYFP alone; 2) mixture of cucurbit[8]uril (5 microM) and FGG-mYFP (20 microM) signal of ternary complex. a) SEC trace, visualized via total ion count; b) ESI-MS spectra of indicated peak area; c) deconvoluted spectra.
Chapter 2

2.5 References

Chapter 3

Molecular characteristics of cucurbit[8]uril induced protein dimerization

Abstract

Cucurbit[8]uril selectively binds and dimerizes proteins bearing a genetically incorporated, N-terminal phenylalanine-glycine-glycine (FGG) peptide motif. Dynamic light scattering shows the particle size of a protein with an FGG motif in the presence of cucurbit[8]uril to double, whereas the size of a protein particle with an N-terminal methionine does change in the presence of cucurbit[8]uril. Cucurbit[8]uril induced protein dimerization is also characterized using small angle X-ray scattering (SAXS). The results show that two subunits of proteins have a z-shaped structure and are in very close proximity forming a very compact protein dimer, despite being connected by a flexible peptide linker. The combination of a short genetic peptide motif and the synthetic supramolecular host molecule holds a promising approach for protein dimerization in biological processes.
3.1 Introduction

Protein dimerization is a crucial mechanism for protein functioning in almost all cellular events\(^1\)-\(^{10}\). Chemical control over protein dimerization allows for understanding and modulating biological processes\(^{11-17}\) and for the construction of artificial protein architectures\(^{18-23}\). Recently, synthetic host-guest chemistry has emerged as an attractive approach to gain control over protein dimerization\(^{24}\). The non-covalent selective interaction of an orthogonal host-guest pair provides a molecular mechanism to induce the dimerization of appended proteins. Two complementary synthetic supramolecular elements can be attached to the biomolecules of choice and mediate their assembly\(^{25-29}\). Alternatively a supramolecular host molecule can be applied as an inducer of dimerization, by binding to two proteins simultaneously. For example, the noncovalent binding of cucurbit[8]uril to the two guest molecules methylviologen and naphthalene results in the formation of a ternary complex which has found applications to modify polymers\(^30-33\), and to induce protein heterodimerization\(^{27}\). Interestingly, cucurbit[8]uril also binds twofold to the tripeptide phenylalanine-glycine-glycine (FGG) with high association constant\(^{34,35}\). The two-fold recognition of a peptide motif by cucurbit[8]uril has enabled the application of cucurbit[8]uril as a supramolecular inducer of protein dimerization for engineered protein structures. Proteins featuring a small, genetically encoded N-terminal FGG peptide motif can be brought to dimerize upon the simple addition of cucurbit[8]uril\(^{36}\) (Fig. 1).

![Figure 1. Protein dimerization is mediated by cucurbit[8]uril.](image)

The highly promising approach of supramolecular controlled protein dimerization lays out the need for a better molecular and structural understanding of the resulting protein assembly process. This understanding is required for subsequent design based approaches for cucurbit[8]uril induced protein dimerization. Here the molecular aspects of the cucurbit[8]uril induced protein dimerization are explored as well as a structural study on the
resulting protein dimer. Fluorescent proteins featuring diverse genetically encoded N-terminal tags were evaluated concerning recognition by cucurbit[8]uril. The structural aspects of size and shape of the protein dimerization events were evaluated by a combination of dynamic light scattering (DLS) and small angle X-ray scattering (SAXS). The results provide clear design parameters for supramolecular induced protein dimerization, concerning motif recognition and protein assembly process.

3.2 Results and discussion

3.2.1 Cucurbit[8]uril induced dimerization of proteins with an N-terminal aromatic amino acid

The molecular characteristics of the cucurbit[8]uril induced protein dimerization were evaluated on a set of fluorescent proteins with diverse, mainly aromatic N-termini. The intein system was applied as a crucial technique to generate proteins with aromatic amino acids at their N-terminus[37]. A set of fluorescent proteins featuring the N-terminal aromatic amino acids F, W, or Y was prepared. Förster Resonance Energy Transfer (FRET) was monitored for a protein donor-acceptor pair of monomeric CFP (mCFP) and mYFP, leading to so-called hetero-FRET[38]. The monomeric variants of the fluorescent proteins[39] were selected to avoid background protein dimerization by an intrinsic affinity between proteins. Previously, we reported that addition of cucurbit[8]uril to a solution of FGG-mCFP/FGG-mYFP resulted in an increase of the FRET ratio (527 nm / 475 nm) from 0.46 to 2.72. Moreover, addition of cucurbit[8]uril to a mixture of Met-CFP/Met-YFP (reference proteins) did not result in any increase of FRET ratio (Fig. 2). Cucurbit[8]uril thus selectively binds and dimerizes proteins displaying the FGG motif at the N-terminus in preference to proteins with an N-terminal methionine, most frequently found in nature[36].

To further explore the substrate selectivity, proteins featuring WGG or YGG motifs were evaluated. Cucurbit[8]uril selectively binds and dimerizes peptides with these aromatic amino acids has been reported[34]. Addition of cucurbit[8]uril to a solution of WGG-mCFP/WGG-mYFP or YGG-mCFP/YGG-mYFP resulted in an increase in the FRET ratio from 0.47 to 1.15 or from 0.46 to 0.57, respectively (Fig. 2). The significantly stronger FRET ratios observed for the phenylalanine terminated proteins over the tryptophan and tyrosine terminated proteins highlight the preferred recognition of the FGG-motif in proteins by cucurbit[8]uril, in line with results observed for isolated peptide motifs[34].
Figure 2. a) Normalized fluorescence emission spectra of CFP/YFP and cucurbit[8]uril: 1) FGG-mCFP/FGG-mYFP (A) + cucurbit[8]uril, 2) WGG-mCFP/WGG-mYFP (B) + cucurbit[8]uril, 3) YGG-mCFP/YGG-mYFP (C) + cucurbit[8]uril, 4) Met-CFP/Met-YFP (D) + cucurbit[8]uril. b) Comparison of the 527 nm/475 nm FRET ratios observed without (grey bars) and with (dark bars) cucurbit[8]uril (1.5 μM) for different protein mixtures (both at 1 μM).

3.2.2 Protein dimerization evaluation via dynamic light scattering

The molecular characteristics and size of the cucurbit[8]uril induced protein assembly were further evaluated using dynamic light scattering studies. Specifically, the determination of the hydrodynamic radius of the resulting protein assemblies in aqueous solution was established. Measurements on the isolated FGG-mYFP protein revealed scattering particles with an average size of 2.5 nm. This would theoretically correspond to a protein of around 29 kDa (based on a spherical protein particle) and indeed fits nicely to the calculated mass of FGG-mYFP, of 28 kDa. Upon addition of cucurbit[8]uril to the FGG-mYFP, the size of the particles in solution increases to 3.1 nm, corresponding to a theoretical protein of around 48 kDa (Fig. 3a). This size nicely corresponds to a dimerized FGG-mYFP protein (56 kDa). Reference experiments show no effect of cucurbit[8]uril addition to control protein MGG-mYFP (Fig. 3b). These light scattering results clearly show that addition of cucurbit[8]uril to
monomeric FGG-mYFP protein leads to the formation of protein dimer with a well-defined, compact size.

![Graph](image)

**Figure 3.** Dynamic light scattering data for isolated FGG-mYFP (a, grey line) and MGG-mYFP (b, grey line) at 40 μM, and after the addition of cucurbit[8]uril (40 μM) to FGG-mYFP (a, dark line) and MGG-mYFP (b, dark line).

### 3.2.3 Structural determinations via small angle X-ray scattering

In order to get more structural insight in the cucurbit[8]uril induced protein dimer, small angle X-ray scattering (SAXS) studies were performed. SAXS has emerged as a useful technique to study the structure and interactions of biological macromolecules in solution\(^{42-44}\). The monomeric FGG-mYFP (910 μM) and the complex of FGG-mYFP:cucurbit[8]uril (857 μM:428 μM) in 10mM sodium phosphate buffer at pH 7 were studied by SAXS. The buffer solution and the buffer solution containing cucurbit[8]uril without protein were also measured. These background scattering files were subtracted from the sample scattering after transmission correction in order to determine the scattering of the protein particles only.

The complex of the FGG-mYFP:cucurbit[8]uril in solution was also measured at a lower concentration (214 μM of FGG-mYFP:107 μM of cucurbit[8]uril). The results showed no systematic deviations between the scattering patterns of the two concentrations (857 μM of
FGG-mYFP:428 μM of cucurbit[8]uril and 214 μM of FGG-mYFP:107 μM of cucurbit[8]uril) after concentration normalization (Fig. 4). This confirmed that there was no significant effect of particle interactions at the high concentration. The scattering intensity of the monomeric FGG-mYFP at small scattering angles was significantly lower than of the FGG-mYFP:cucurbit[8]uril complex, which means that bigger particles of the FGG-mYFP protein are formed in the presence of cucurbit[8]uril.

Figure 4. Background subtracted and concentration normalized small angle X-ray scattering curves of monomeric FGG-mYFP (910 μM FGG-mYFP, no cucurbit[8]uril - full squares) and cucurbit[8]uril-induced FGG-mYFP dimer at two different concentrations (857 μM FGG-mYFP: 428 μM cucurbit[8]uril - open triangles and 214 μM of FGG-mYFP:107 μM of cucurbit[8]uril-open circles) in 10 mM sodium phosphate buffer at pH 7.

Guinier analysis of the scattering curve leads to the radius of gyration $R_g$ and the forward scattering intensity $I(0)$ of the scattering particles. The so-called Guinier plot (natural logarithm of scattering intensity versus square of scattering vector) of the two samples is shown in Figure 5. In such a plot a linearly decaying scattering intensity can be seen at the limit of small scattering vectors ($qR_g < 1$). $R_g$ can be determined from the slope of the decay.

The radius of gyration ($R_g$) of both the FGG-mYFP monomer and the cucurbit[8]uril-induced FGG-mYFP dimer were summarized in table 1, since $R_g$ values correspond to the conformation and distribution of two-domain proteins\cite{45}. The results showed that the $R_g$ of the FGG-mYFP monomer was observed at 2.3 nm, larger than the expected $R_g$ of 1.8 nm of a YFP parent (protein calculated from the high resolution crystal structure 1YFP). The larger size nicely corresponds to the presence of a flexible linker (strep-tag and FGG motif). Addition of cucurbit[8]uril to the protein solution led to a change in the scattering profile. The $R_g$ of the cucurbit[8]uril-induced FGG-mYFP dimer (2.7 nm) was significantly larger than
that of the FGG-mYFP monomer. The selective recognition of the FGG motif allows cucurbit[8]uril to induce the FGG-mYFP dimer\textsuperscript{[56]}, resulting in an increase of the $R_g$.

![Graph](image)

*Figure 5. Guinier plots of SAXS intensities of monomeric FGG-mYFP (full squares) and cucurbit[8]uril-induced FGG-mYFP dimer (open triangles). The fitting region was between 0.1 and 0.65 nm$^{-1}$. This region contains 82 data points.*

<table>
<thead>
<tr>
<th>Sample</th>
<th>$R_g$ [nm]</th>
<th>$I(O)$ [a.u.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical 1YFP (pdb)</td>
<td>1.8</td>
<td>-</td>
</tr>
<tr>
<td>FGG-mYFP</td>
<td>2.3</td>
<td>6.59$\times$10$^3$</td>
</tr>
<tr>
<td>FGG-mYFP + cucurbit[8]uril</td>
<td>2.7</td>
<td>9.16$\times$10$^3$</td>
</tr>
</tbody>
</table>

For the determination of the 3 dimensional structure of the cucurbit[8]uril-induced FGG-mYFP dimer, it is beneficial to make use of prior knowledge of the system. Most importantly this is the structure of YFP which is known from X-ray crystallography (1YFP). It can be assumed that the structure of the beta-barrel subunit does not change dramatically in the complex and thus this unit can be used as a rigid body for the modelling. The two FGG motifs are linked within one cucurbit[8]uril molecule. Thus, the system can be described as 2 rigid bodies (1YFP) linked together at the end of their N-terminal tails which are both 21 amino acids long. The relative orientations of the rigid bodies and the folding of the tails have to be optimized to give best agreement with the SAXS data.

The rigid body of the proteins and their complexes was calculated using the BUNCH program\textsuperscript{[46]}. Modeling of a two-domain protein, based on the fluorescence and light scattering results, against the SAXS data was performed for the cucurbit[8]uril induced dimers. A simulated annealing algorithm was employed to find the optimal positions and
orientations of the available high resolution coordinates of the two fluorescent domains. The portion with unknown structure (the flexible N-terminus) was represented as a chain of 21 dummy residues. Furthermore, the 3 N-terminal amino acids, the FGG motif, of the two proteins were imposed to be in close proximity. The calculation was performed in 18 successive runs of BUNCH with the same data set (Fig. 6).

![Figure 6. 18 fits with the program BUNCH (lines) to the experimental scattering curve of dimeric FGG-mYFP (symbols).](image)

The results showed the variation among the resulting structural models is small, which indicates a certain degree of flexibility of the system. All of the resulting models have z-shaped structure with a small contact area of the YFP rigid bodies. The most probable structure of the dimeric FGG-mYFP is shown in figure 7. The two YFP domains are arranged in a z-shaped structure with a significant connection area. In addition, the axis of the two YFP barrels is almost parallel to each other which indicate the z-shaped structure on average is aligned in one plane. This resulting model completely agrees with studies on the conformation of chimeric fluorescent proteins connected via a covalent flexible linker. These proteins featured structures in which the two subunits also show an elongated conformation instead of a compact side-by-side conformation\[42\]. The structures of the flexible protein ends are significantly less well defined since the flexible linker may disorder the conformation of two domains\[42\]. In all models, the two subunits are in very close proximity. The results show that the flexible linker dimerized via the cucurbit[8]uril between the two subunits does not contribute to the inter-domain distance. As a results the two subunits are well ordered, which is fully in line with the strong energy transfer observed for the cucurbit[8]uril-induced fluorescent protein dimers (Fig. 2).
Figure 7. A typical result of one run of the program BUNCH for dimeric FGG-mYFP. The same structure is shown 2 times with different viewing modes. A) The YFP building blocks can be seen in a ribbon illustration. B) The space filling illustration is shown: The YFP rigid blocks are shown in light grey and heavy grey. C) A view on the most probable structure of the dimeric FGG-mYFP complex.

3.3 Conclusions

The fluorescent proteins featuring a genetically encoded tripeptide motif at the N-terminus are versatile model proteins for the evaluation of supramolecular host cucurbit[8]uril-induced protein dimerization. Cucurbit[8]uril preferentially recognizes and dimerizes proteins featuring N-terminal FGG motifs and to a lesser extend WGG motifs. The induced dimerization leads to compact dimeric protein structures, resulting in strong energy transfer. The cucurbit[8]uril/FGG-mYFP (1:2) complex forms a compact particle and reveals a z-shape structure of the two YFP subunits. The two beta-barrels are in very close proximity, forming a compact protein dimer in solution. The long flexible linker clearly does not negatively disturb the protein alignment process. The results provide molecular insight in the optimal supramolecular motif for protein dimerization, as well as structural design rules for applying the concept of cucurbit[8]uril induced protein dimerization to other proteins. Clearly, a close
proximity of the N-termini of the proteins under study facilitates an optimal interaction between the protein domains in the supramolecular complex. Application of this supramolecular technology in other protein dimerization events, such as activation of dimeric enzymes and membrane proteins has become now feasible, putting cucurbit[8]uril forward as a novel building block concept in synthetic biology.

3.4 Experimental section
3.4.1 Construction of plasmids

The monomeric fluorescent cyan and yellow proteins (mCFP, mYFP) with N-terminal FGG, WGG, YGG, motifs were generated by using the N-terminal intein from plasmid pTWIN-1 (New England). DNAs encoding for FGG-mCFP and FGG-mYFP and two references Met-CFP and Met-YFP were described in literature[^36]. The plasmid WGG-mCFP/WGG-mYFP, YGG-mCFP/YGG-mYFP and MGG-mYFP were generated from FGG-mCFP/FGG-mYFP templates by using QuickChange Site-Directed Mutagenesis Kit (Qiagen) to replace F with W, Y and M, respectively.

Amino acid sequences:

**FGG-mCFP**

<table>
<thead>
<tr>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGGASHHHHHHHSMVSKGEELFTGVVPILVLEDGVNGHKFSVSGEGEDATYGBKTLKFICITGKLPVPWPTLVTTLTWGVQCFRSYPDHMKQHDFFKSAAMPEGYVQERTIFFKDGNYKTRAEVKFEKDTILVNRILEKGDIFKEDGNILGKLEYNISHNVIYITAKQKNGIKANFKIRHNNEDGSVQLADHYQQNTPIDGPVLPDNHYLSIQSKLSKDPNEKRDHMVLLEFVTAAIGTLGMDELYK</td>
</tr>
</tbody>
</table>

**FGG-mYFP**

<table>
<thead>
<tr>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGGASWSHPQFEKSAVMVSKGEELFTGVVPILVLEDGVNGHKFSVSGEGEDATYGBKTLKFICITGKLPVPWPTLVTTLTWGVQCFRSYPDHMKQHDFFKSAAMPEGYVQERTIFFKDGNYKTRAEVKFEKDTILVNRILEKGDIFKEDGNILGKLEYNISHNVIYITAKQKNGIKANFKIRHNNEDGSVQLADHYQQNTPIDGPVLPDNHYLSIQSKLSKDPNEKRDHMVLLEFVTAAIGTLGMDELYK</td>
</tr>
</tbody>
</table>

**WGG-mCFP**

<table>
<thead>
<tr>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGGASHHHHHHHSMVSKGEELFTGVVPILVLEDGVNGHKFSVSGEGEDATYGBKTLKFICITGKLPVPWPTLVTTLTWGVQCFRSYPDHMKQHDFFKSAAMPEGYVQERTIFFKDGNYKTRAEVKFEKDTILVNRILEKGDIFKEDGNILGKLEYNISHNVIYITAKQKNGIKANFKIRHNNEDGSVQLADHYQQNTPIDGPVLPDNHYLSIQSKLSKDPNEKRDHMVLLEFVTAAIGTLGMDELYK</td>
</tr>
</tbody>
</table>

[^36]: Reference citation
WGG-mYFP

\textbf{WGGASWSHPQFEK}A\textbf{SMVSKGEELFTGV}PIL\textbf{V2LDGDVNGH}KFSVSGE\textbf{GE}G\textbf{G}D\textbf{ATYGKL}
\textbf{TLKFIC}TTGKL\textbf{PVPWPTLVTT}FGYGL\textbf{QCFARYPDHMKQHDF}FS\textbf{A}MPE\textbf{GYVQ}ERTIFFKDD
\textbf{GNYKTRA}E\textbf{FKEGDTLV}NRI\textbf{EKGIDFKEDGNIL}GHKLEYNYN\textbf{SHVYIMADKQ}NGIKV
\textbf{NF}KIRHN\textbf{EDGSV}QLAD\textbf{HYQ}QNTPIGDPVLLPD\textbf{NH}YLSY\textbf{Q}SLKD\textbf{PNEKRDHMV}LLEF
\textbf{VTAAGITLGMD}E\textbf{LYK}

YGG-mCFP

\textbf{YGGASWSHPQFEK}A\textbf{SMVSKGEELFTGV}PIL\textbf{V2LDGDVNGH}KFSVSGE\textbf{GE}G\textbf{G}D\textbf{ATYGKL}
\textbf{TLKFIC}TTGKL\textbf{PVPWPTLVTT}FGYGL\textbf{QCFARYPDHMKQHDF}FS\textbf{A}MPE\textbf{GYVQ}ERTIFFKDD
\textbf{GNYKTRA}E\textbf{FKEGDTLV}NRI\textbf{EKGIDFKEDGNIL}GHKLEYNYN\textbf{SHVYIMADKQ}NGIKV
\textbf{NF}KIRHN\textbf{EDGSV}QLAD\textbf{HYQ}QNTPIGDPVLLPD\textbf{NH}YLSY\textbf{Q}SLKD\textbf{PNEKRDHMV}LLEF
\textbf{VTAAGITLGMD}E\textbf{LYK}

YGG-mYFP

\textbf{YGGASWSHPQFEK}A\textbf{SMVSKGEELFTGV}PIL\textbf{V2LDGDVNGH}KFSVSGE\textbf{GE}G\textbf{G}D\textbf{ATYGKL}
\textbf{TLKFIC}TTGKL\textbf{PVPWPTLVTT}FGYGL\textbf{QCFARYPDHMKQHDF}FS\textbf{A}MPE\textbf{GYVQ}ERTIFFKDD
\textbf{GNYKTRA}E\textbf{FKEGDTLV}NRI\textbf{EKGIDFKEDGNIL}GHKLEYNYN\textbf{SHVYIMADKQ}NGIKV
\textbf{NF}KIRHN\textbf{EDGSV}QLAD\textbf{HYQ}QNTPIGDPVLLPD\textbf{NH}YLSY\textbf{Q}SLKD\textbf{PNEKRDHMV}LLEF
\textbf{VTAAGITLGMD}E\textbf{LYK}

MGG-mYFP

\textbf{MG}G\textbf{ASWSHPQFEK}A\textbf{SMVSKGEELFTGV}PIL\textbf{V2LDGDVNGH}KFSVSGE\textbf{GE}G\textbf{G}D\textbf{ATYGKL}
\textbf{TLKFIC}TTGKL\textbf{PVPWPTLVTT}FGYGL\textbf{QCFARYPDHMKQHDF}FS\textbf{A}MPE\textbf{GYVQ}ERTIFFKDD
\textbf{GNYKTRA}E\textbf{FKEGDTLV}NRI\textbf{EKGIDFKEDGNIL}GHKLEYNYN\textbf{SHVYIMADKQ}NGIKV
\textbf{NF}KIRHN\textbf{EDGSV}QLAD\textbf{HYQ}QNTPIGDPVLLPD\textbf{NH}YLSY\textbf{Q}SLKD\textbf{PNEKRDHMV}LLEF
\textbf{VTAAGITLGMD}E\textbf{LYK}

Met-CFP

\textbf{MEASH}H\textbf{HHH}HAM\textbf{SVSKGEELFTGV}PIL\textbf{V2LDGDVNGH}KFSVSGE\textbf{GE}G\textbf{G}D\textbf{ATYGKLT}
\textbf{LKFI}CTT\textbf{GKL}\textbf{PVPWPTLVTT}FGYGL\textbf{QCF}RYPD\textbf{HMKQHDF}FS\textbf{A}MPE\textbf{GYVQ}ERTIFFKDD
\textbf{GNYKTRA}E\textbf{FKEGDTLV}NRI\textbf{EKGIDFKEDGNIL}GHKLEYNYN\textbf{SHVYIMADKQ}NGIKAN
\textbf{FK}IRHN\textbf{EDGSV}QLAD\textbf{HYQQN}TPIGDPVLLPD\textbf{NH}YLSY\textbf{Q}SLKD\textbf{PNEKRDHMV}LLEF
\textbf{VTAAGITLGMD}E\textbf{LYK}
**Chapter 3**

**Met-YFP**

MESAWSHPQFEKAMVSKGEELFTGVVIPVELGDVNGHKSVEGEGDATYGLTL
KFICTTGKLVPWPTLVTTFVGLQCFARYPDHMKQHDFKFSAEPGYTVGERTIFKDDG
NYKTRAEVKFEGDITVNIELKGIDFKEGDNILGKLEYMNSHNYIMADKQKNGIKVM
FKIRHNIEDGSVQADHYQQNTIGDPVLLPDNHQSYQSLSKDPNEKRDMVLLDEFV
TAAGITLGMDELYK

**3.4.2 Protein expression and purification**

The plasmids were transformed into *E. coli* strain BL21 (DE3, New England Biolabs). The bacteria were cultured in LB medium containing 100 μg/ml of ampicillin and the cells were grown at 37 °C, shaking 250 rpm to an OD₆₀₀ of 0.5 - 0.7, then IPTG was added to a final concentration of 0.4 mM. The cells were continuously incubated overnight at 15 °C, shaking 180 rpm before being harvested. The pellet was resuspended into the bugBuster protein extraction reagent plus benzonase nuclease, and the insoluble material was removed by centrifugation at 20,000 rpm for 40 minutes at 4 °C. The soluble fraction was applied to a column filled with chitin beads (New England Biolabs) through gravity flow, and the column was washed with 40 volumes of sodium phosphate buffer (10 mM, 100 mM of sodium chloride, pH 7) and subsequently incubated overnight at room temperature. The intein cleaved proteins were then collected in the flow-through using the phosphate buffer. The target proteins were analyzed by SDS-PAGE (Fig. 8) and LC-ESI-MS (Table 2), confirming purity and identity.

![Figure 8. Schematic representation of generation of FGG-protein by intein system. SDS-PAGE of proteins: 1) FGG-mCFP, 2) WGG-mCFP, 3) YGG-mCFP, 4) Met-CFP, 5) FGG-mYFP, 6) WGG-mYFP, 7) YGG-mYFP, 8) Met-YFP.](image-url)
3.4.3 LC-ESI-MS analysis

The LC-ESI-MS experiments were performed using a LCQ Fleet (Thermo Scientific). Proteins went through the C4 column under elution with a water/acetonitrile gradient (5% water: 95% acetonitril for 1 minute to 70% water: 30% acetonitril in 9 minutes and stay at 70% water: 30% acetonitril for 2 minutes) with 0.1% formic acid before they were introduced to the ion source and mass spectrometry. The mass of the proteins was deconvoluted by Magtran software (Table 2).

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Calculated mass (Da)</th>
<th>Measured mass (Da)</th>
<th>Error (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGG-mCFP</td>
<td>28587</td>
<td>28590</td>
<td>+3</td>
</tr>
<tr>
<td>WGG-mYFP</td>
<td>28685</td>
<td>28690</td>
<td>+5</td>
</tr>
<tr>
<td>YGG-mCFP</td>
<td>28564</td>
<td>28566</td>
<td>+2</td>
</tr>
<tr>
<td>YGG-mYFP</td>
<td>28662</td>
<td>28667</td>
<td>+5</td>
</tr>
</tbody>
</table>

3.4.4 Fluorescence spectroscopy

All fluorescence spectroscopy measurements were performed at 20 °C in quartz cuvettes of 0.5 cm path length and 300 μL minimal volume. Samples for experiments were prepared in 10 mM sodium phosphate buffer, pH 7. Protein concentration was measured using NanoDrop (Nanodrop technologies) and adjusted using the absorbance at 435 nm and a molar extinction coefficient of 32,500 M⁻¹cm⁻¹ for CFP and at 515 nm and a molar extinction coefficient of 84,000 M⁻¹cm⁻¹ for YFP. All fluorescence measurements were carried out on a Cary Eclipse fluorescence spectrophotometer (Varian). All steady-state fluorescence data were recorded using an excitation wavelength of 410 nm and at 20 °C. Measurement parameters and sample concentrations were kept constant to make the data comparable. The YFP/CFP ratio is defined as the fluorescent intensity at 527 nm divided by the fluorescent intensity at 475 nm of YFP/CFP mixtures, slit width of 5 nm.

3.4.5 Dynamic light scattering

Dynamic light scattering measurements were carried out with a Zetasizer μV (Malvern Instruments Limited, UK). All samples were diluted in 10 mM sodium phosphate, pH 7. Samples were filtered via 0.1 μm syringe filter into a clean cuvette. The cuvette was then inserted into the unit for 2 minutes for equilibration at 20 °C before a measurement was performed. All experiments were performed at least in duplicate.
3.4.6 Small angle X-ray scattering (SAXS)

SAXS measurements were done with a SAXSess camera\textsuperscript{[49]} (Anton-Paar, Graz, Austria), which was connected to an X-ray generator (Panalytical PW3830) operating at 40 kV and 50 mA with a sealed-tube Cu anode. A Göbel mirror was used to convert the divergent polychromatic X-ray beam into an intense focused line-shaped beam of Cu Kα radiation ($\alpha = 0.154$ nm). The scattering pattern was recorded by a one dimensional silicon strip detector operating in single-photon counting mode (Mythen 1K by Dectris, Switzerland). This detector features 1280 pixels with a height of 50 μm each (active area: 64 x 8 mm$^2$). Cosmic ray removal was done on the raw data before absorption correction and background subtraction. The data were converted into the one-dimensional scattering function I(q), where q is the norm of the scattering vector defined by $q = (4\pi/\lambda)\sin (\theta/2)$. Here $\theta$ is the scattering angle.

The samples were filled into the temperature controlled (20 ± 0.1°C) sample holder (quartz capillary, diameter 1 mm, wall thickness 0.01 mm) and placed in the evacuated SAXS instrument. The samples were exposed for 1 hour (6 frames of 10 min). The individual frames were averaged and the standard deviation was calculated.

For data analysis the program BUNCH\textsuperscript{[46]}, which is part of the ATSAS software package\textsuperscript{[50]}, was used. BUNCH performs modelling of multidomain proteins against SAXS data using a combined rigid body and ab initio modelling approach. The program allows determination of three-dimensional domain structure of proteins when the high resolution structure of individual domains are available, for instance from protein crystallography. A simulated annealing algorithm was employed to find the optimal positions and orientations of available high resolution coordinates of the domains with known structure and from the portion with unknown structure represented as a chain of dummy residues.

In the present work the YFP monomers were represented by the structure published as 1YFP in the protein data base\textsuperscript{[51]} linked to an N-terminal flexible chain of 21 dummy residues. As a condition, the 3 dummy residues at the N-termini (FGG) in the dimeric complex have to be in close proximity. The calculation was performed several times with the same data set, and the resulting structures were compared in order to get information about different possible structures and possible flexibilities in the complex.
3.5 References

Chapter 4

Cucurbit[8]uril-mediated protein homotetramerization

Abstract

Many proteins form homotetrameric complexes which play a crucial role in cellular protein functions. A supramolecular protein homotetramerization approach has been devised which enables the controlled formation of discrete protein tetramers. The supramolecular element cucurbit[8]uril has been used as an inducer of the protein tetramerization in combination with intrinsic affinities between the proteins. The combination of a dimerizing interface on the fluorescent proteins under study (dYFP, dCFP), with a genetically encoded N-terminal phenylalanine-glycine-glycine (FGG) peptide motif allows cucurbit[8]uril to selectively recognize and induce FGG-dYFP or FGG-dCFP homotetramerization. The concept of cucurbit[8]uril-induced protein homotetramerization was elucidated by using a combination of fluorescence anisotropy, dynamic light scattering and size exclusion chromatography experiments. Simple addition of cucurbit[8]uril to a solution of for example FGG-dYFP or FGG-dCFP generates the tetrameric protein assemblies. The cucurbit[8]uril-induced tetrameric protein complex is highly stable and can be separated by size exclusion chromatography. This supramolecular induced protein homotetramerization approach opens up a strong novel concept in generating well-defined synthetic protein assemblies for the stabilization of weakly associating homotetrameric proteins and for regulating the functionality of homotetrameric proteins in biological events.
4.1 Introduction
Many natural proteins form homooligomeric complexes in order to be functional in their cellular context. In the Brenda Enzyme database has shown that around 70% of the enzymes are multimeric enzymes, with most of them being homooligomers\[1;2\]. In addition, a search of the Protein Data Bank (PDB) has also shown that of the homooligomeric proteins 38% are present as homodimers and 10% as homotetramers in which the structures are well defined\[3\]. In fact, most of the homooligomeric protein complexes are homodimeric proteins which are responsible for almost all of its protein function\[4\]. However, a significant amount of homotetrameric proteins exist as functional assemblies and play important roles in many cellular events \[5-9\]. The study and control over protein homotetramerization provides molecular insights in their formation as well as an entry to design and mimic protein function for biomedical research and applications.

Figure 1. A) Crystal structure of p53 tetramer-bound DNA (image reproduced from reference \[10\]). B) Active p53 tetramer plays a key role in different cellular functions such as cell control, differentiation and apoptosis.

Protein tetramerization is a process to instate function in many proteins such as transcription factors\[9;11;12\], transport proteins\[13;14\], potassium channels\[15\], water channels\[16;17\] and many enzymes\[18;19\]. For example, a homotetrameric lactose repressor (LacI), assembled as a dimer of dimers, can bind simultaneously to two operators of DNA, thereby forming a loop that blocks the transcription of genes encoding lactose-metabolizing enzymes\[20-22\]. p53 is another transcription factor, a tumour suppressor, which is active as a tetramer and plays a key role in different cellular functions such as apoptosis, cycle control and differentiation (Fig. 1)\[9;11;23-25\]. In the pathway of oxygen transport in the organism, haemoglobin must be tetrameric to be functional\[13\].

Studies of protein tetramerization have demonstrated that almost all of the proteins which exist in tetrameric form have been assembled by a “dimers of dimers” pathway\[26\] (Fig. 2).
Understanding the molecular mechanism of the protein tetramerization assembly provides strategies for controlling its function. For example, the deletion of the carboxy-terminal domain or mutation of the lactose repressor leads to reduced tetramer formation. This decreases the lactose repressor mediated formation of DNA loops\cite{27,28}. In contrast, engineering of novel coiled-coil interfaces for the lactose repressor enables a higher stability of the tetrameric complex\cite{20,21}. As another example, engineering of a monomeric avidin by mutations at its protein interaction interface allows for molecular control via ligand binding; biotin binding to the monomeric avidin induces avidin tetramerization\cite{29}. However, the current tools that allow control over protein tetramerization are limited to a selected number of examples and feature drawbacks such as irreversibility\cite{27,28}, partial reversibility\cite{29}, or reversibility only upon temperature changes\cite{21}, thus preventing molecular control or strongly effecting cellular processes. There is a need for new approaches to induce and control protein tetramerization. Such novel systems would provide molecular diversity to the protein assembly field and can possibly overcome current limitations, for example regarding reversibility. Herein, we provide a supramolecular chemistry system which is able to reversibly induce protein homotetramerization via a combination of intrinsic protein affinities and a genetically encoded peptide tag capable of two-fold binding to cucurbit[8]uril (Fig. 3).

![Diagram](image)

*Figure 2. Schematic representation of a protein homotetramerization assembly pathway: “monomer-dimer-tetramer”.*

Supramolecular host-guest chemistry has recently emerged as a versatile tool to reversibly control protein dimerization\cite{30-33}. The specific recognition of lithocholic acid by β-cyclodextrin allows for example a β-cyclodextrin-conjugated yellow fluorescent protein (YFP) to selectively bind a lithocholic acid-conjugated cyan fluorescent protein (CFP). This self-assembled CFP-YFP heterodimer features significant Förster resonance energy transfer (FRET), which can be reversibly inhibited by the addition of an excess of β-cyclodextrin\cite{33}. Alternatively, the fusion of proteins with a very small genetically encoded phenylalanine-glycine-glycine (FGG) peptide motif at their N-terminus allows cucurbit[8]uril to selectively recognize and induce their homodimerization by the supramolecular interaction between two FGG motifs and the cavity of cucurbit[8]uril\cite{30}. Protein dimerization, both of the homo-
and hetero-type, has thus been successfully achieved via the application of host-guest chemistry. This bodes well for applying supramolecular chemistry to reversibly modulate protein homotetramerization. Supramolecular protein homotetramerization has not been addressed as such thus far, while being an important biological mechanism for controlling protein activity and function as discussed above. Additionally, supramolecular control over protein tetramerization provides a challenging opportunity to push the limits of self-assembly with proteins and synthetic elements even further.

Figure 3. Schematic representation of cucurbit[8]uril-induced protein homodimerization and homotetramerization. (1) Engineering of a monomeric interface of YFP by one mutation (A206K) (grey circle)\(^{34}\); (4) Cucurbit[8]uril-induced protein homodimerization via supramolecular interaction between FGG motif and cavity of cucurbit[8]uril\(^{30}\); (2,3) Engineering of a dimeric interface of YFP by two mutations (S208F and V224L) (red and blue circles) allows dimerizing variant dYFP to exist as dimerized form in two possible orientations: parallel (2) or anti-parallel (3) barrels\(^{35}\); (5,6) Cucurbit[8]uril selectively recognizes and induces YFP homotetramerization in parallel (5) or anti-parallel (6) complexes.
In this study, cucurbit[8]uril-induced homotetramerization of fluorescent proteins (CFP and YFP) is achieved via a 2 step assembly process (Fig. 3): 1) two-fold mutation of the protein interface (S208F and V224L) provides these proteins with an intrinsic affinity for dimerization, yielding so-called dimerizing variants (dYFP and dCFP)\(^{35}\); 2) addition of the small genetically encoded FGG motif at the N-terminus of these proteins allows for a selective recognition by cucurbit[8]uril to induce an orthogonal manner for dimerization. Overall these two orthogonal protein dimerization events are envisioned to lead to the formation of protein tetramers. The formation of cucurbit[8]uril-induced protein homotetramers was elucidated using a combination of fluorescence spectroscopy, dynamic light scattering and size exclusion chromatography.

4.2 Results and discussion

4.2.1 Protein design and expression

Fluorescent proteins are versatile model proteins due to their structural stability, optical detection and easy expression in E. coli\(^{36,37}\). Most fluorescent proteins feature a weak affinity for dimerization in the micromolar range of concentrations\(^{33}\). Previous studies have shown that specific point mutations at the protein surface allow tuning the dimerization affinity. Replacement of alanine (A) with lysine (K) at site 206 of for example CFP and YFP prevents their intrinsic dimerization, leading to the so-called monomeric cyan and yellow fluorescent proteins (mCFP, mYFP)\(^{34}\). In contrast, double replacement of serine with phenylalanine at position 208 and valine with leucine at site 224 (S208F and V224L) enhances the affinity of these proteins to dimerize, leading to the so-called dimerizing CFP and YFP (dCFP, dYFP)\(^{35}\). These dimerizing variants are attractive proteins for the generation of supramolecular controlled protein tetramers, by providing them additionally with an FGG motif. Therefore, dimerizing YFP and CFP proteins featuring either an FGG motif or an MGG motif, as negative control, were designed and generated. In these constructs, the flexible linkers between the FGG motif and YFP or CFP protein was removed to prevent the formation of cucurbit[8]uril-induced protein homodimers.

These proteins were genetically generated by using an intein system as N-terminal autocleavage domain in a pTWIN1 plasmid. All protein constructs were transformed and expressed in BL21 (DE3) cells. The intein-fusion proteins were captured on a chitin bead column after expression for affinity purification. Subsequently, the intein cleaved proteins, featuring their N-terminal peptide motifs, were eluted from the columns after incubation at room temperature and pH 7.0 overnight. All proteins were isolated in pure form and with the correct molecular weight as evidenced by SDS-PAGE and LC-ESI-MS (Fig. 4 and Table 1).

59
4.2.2 Fluorescence anisotropy measurements

Fluorescent anisotropy was applied to study cucurbit[8]uril-induced protein assembly of mYFP and dYFP. The assembly of fluorescent proteins in larger aggregates, be it of the dimer or tetramer type, typically results in a decrease in fluorescence anisotropy, a phenomenon called homo-FRET[30]. The reference monomeric protein MGG-mYFP features a fluorescence anisotropy value of 0.32 (Fig. 5). Addition of cucurbit[8]uril to MGG-mYFP solution did not result in a change of the fluorescence anisotropy, in line with the absence of an interaction of the two components. The FGG-mYFP features the same fluorescence anisotropy value as MGG-mYFP, providing proof that the N-terminal phenylalanine alone does not have an effect on the protein assembly. Addition of cucurbit[8]uril to the FGG-mYFP solution resulted in a decrease of the fluorescence anisotropy value from 0.32 to 0.29. The cucurbit[8]uril induces dimerization of FGG-mYFP via the supramolecular interaction between the FGG motif and the cavity of cucurbit[8]uril[30].

The fluorescence anisotropy value of both FGG-dYFP and MGG-dYFP were observed at around 0.28, similar to that of the FGG-mYFP dimer in the presence of cucurbit[8]uril. These results indicate that the two mutations (S208F and V224L) induce FGG-dYFP to homodimerize. Addition of cucurbit[8]uril to the reference MGG-dYFP protein did not affect the anisotropy, in line with the lack of a host-guest interaction for this protein with the cucurbit[8]uril. Interestingly, addition of the same amount of cucurbit[8]uril to FGG-dYFP resulted in significant decrease of the fluorescence anisotropy value from 0.28 to 0.25. This decrease of the anisotropy value implies that a higher-order oligomerization might be induced in the presence of cucurbit[8]uril. The intrinsically dimerized dYFP assembles into higher-order oligomeric structures by virtue of the subsequent binding to cucurbit[8]uril.
Cucurbit[8]uril-mediated protein homotetramerization

Figure 5. Fluorescence anisotropy values of YFP variants (2 μM) in the absence (grey bars) and presence (dark bars) of cucurbit[8]uril (2 μM) in 10 mM phosphate buffer at pH 7, were observed by using a 500 nm excitation wavelength and scan spectra over a range of emission wavelengths (524-532 nm).

4.2.3 Dynamic light scattering

The cucurbit[8]uril mediated protein assembly process was further investigated using dynamic light scattering (DLS) for the determination of the radius of the protein particles. Measurements on FGG-mYFP revealed scattering particles with an average radius of 2.4 nm (Fig. 6A). This theoretically corresponds to a spherical particle / protein of around 25 kDa and fits nicely to the actual 27 kDa mass of FGG-mYFP. Addition of cucurbit[8]uril to FGG-mYFP increases the size of particles to an average radius of 3.1 nm, which corresponds to a protein mass of around 48 kDa. (It should be noted that the mass calculations are based on a spherical model, possibly not exactly matching the actual shape of the protein assemblies.) The observed size and mass match the 54 kDa mass of a dimerized FGG-mYFP protein and are in line with the fluorescence anisotropy results. Cucurbit[8]uril thus selectively dimerizes FGG-mYFP in aqueous solution[30]. The dimerizing variants of YFP and CFP, FGG-dYFP and FGG-dCFP, scattered with an average radius of 2.9 nm and 3.0 nm, respectively, which are significantly bigger than those of FGG-mYFP. This demonstrates that the dYFP and dCFP proteins are, for at least a major part, dimerized in solution. Addition of cucurbit[8]uril to FGG-dYFP and FGG-dCFP further increase the size of the protein assemblies to 4.1 nm and 4.0 nm, respectively (Fig 6A, 6C). These sizes correspond to a theoretical spherical protein of around 85-92 kDa. This nicely fits to the calculated 108 kDa mass of a protein tetramer. Addition of the same amount of cucurbit[8]uril to reference proteins MGG-mYFP and MGG-dYFP did not result in changes of the particle size of the proteins (Fig. 6B). The results show that the assembly of FGG-mYFP into a dimer and of FGG-dYFP into a tetramer upon addition of cucurbit[8]uril is exclusively due to the recognition of the FGG motif by the cucurbit[8]uril.
4.2.4 Size exclusion chromatography

The cucurbit[8]uril induced homodimer of FGG-mYFP was previously shown to be stable enough to be separated on a size exclusion column. Addition of increasing amounts of cucurbit[8]uril to a solution of FGG-mYFP resulted in the appearance of a second peak at approximately double molecular weight. Similarly we studied the cucurbit[8]uril induced tetramerization of FGG-dCFP by size exclusion chromatography (Fig. 7). The exclusion time of FGG-dCFP in the absence of cucurbit[8]uril were 15.0 minute. The molecular weight of protein eluting at 15.0 minutes corresponds to the mass of 49 kDa (based on the calibration curve) and closely matches the mass of a FGG-dCFP dimer. The size exclusion data thus...
Cucurbit[8]uril-mediated protein homotetramerization

again support that FGG-dCFP exists as a dimer in buffered solution. Interestingly, addition of cucurbit[8]uril to FGG-dCFP resulted in the appearance of a second peak at higher molecular weight. This second peak increased in a cucurbit[8]uril dose-dependent fashion. At the condition of a cucurbit[8]uril/FGG-dCFP ratio of 1:1, the signal for the dimer had disappeared and only the signal for the larger FGG-dCFP assembly was observed. The protein assembly eluted at 13.4 minutes which corresponds to a protein mass of 111 kDa. The observed mass of the protein assembly matches the calculated mass of a tetrameric FGG-dCFP protein, 108 kDa, very closely.

The molecular mechanism of the supramolecular-induced protein homotetramerization results from a combination of intrinsic affinity for protein dimerization and the selective twofold recognition of the N-terminal FGG-motif by cucurbit[8]uril. This molecular recognition process could either be obtained following pathway 2 or pathway 3 (Fig. 3). The presence of well-defined tetrameric species bodes for pathway 3, leading to a well-defined protein tetramer in which the proteins are first dimerized via an anti-parallel arrangement[38;39], followed by the cross-linking of two of the dimers into a tetramer, in line with the frequently encountered dimers of dimers mechanism[26]. Assembly via pathway 2, would initially lead to parallel oriented protein dimers with the N-terminal FGG-motifs located in close proximity. This preorientation would potentially lead to a further stabilization of the protein dimer via N-terminal cross-linking with cucurbit[8]uril.

![Figure 7](image)

**Figure 7.** A) Exclusion times of different molecular weight of proteins for calibration. B) Size-exclusion chromatograms (superdex 200 column) of FGG-dCFP with increasing concentrations of cucurbit[8]uril (0, 5, 10, 20 and 40 μM) (from grey to dark lines).
4.3 Conclusions

Supramolecular protein tetramerization is a highly attractive novel approach to generate well-defined tetrameric assemblies of proteins. The FGG-tagged fluorescent proteins under study assemble into a tetramer upon addition of cucurbit[8]uril. The assembly process is the result of a combination of intrinsic protein affinities and a genetically encoded peptide tag capable of two-fold binding to cucurbit[8]uril. As a result, the assembly process follows a “monomer-dimer-tetramer” pathway. The specific recognition of dimerized proteins bearing an N-terminal FGG motif allows cucurbit[8]uril to induce protein homotetramerization. These resulting complexes are very stable as exemplified by the capacity to be separated by size exclusion chromatography. As such, these protein complexes are easily accessible. The concept of supramolecular induced protein homotetramerization employing a combination of intrinsic protein affinity and a very small genetically encoded protein tags provides an attractive novel approach for the generation of well-defined synthetic protein assemblies and for potential application in regulating the functionality of homotetrameric proteins in biological events.

4.4 Experimental section

4.4.1 Construction of plasmids

The fluorescent proteins with an N-terminal FGG motif were generated by using the N-terminal intein from plasmid pTWIN-1 (New England Biolabs). DNA encoding for FGG-mYFP, FGG-dCFP, and FGG-dYFP were amplified by PCR using pHT486, pHT483 and pHT484 as templates, respectively, with primers ON12 5'-gcg tgc tct tcc aac ttt ggt ggc gtg agc aag ggc gag gag ctc ttc-3', ON180 5'-ggc cat gga tcc tta ctt gta cag ctc gtt gcc gag -3'. DNA encoding for MGG-dYFP was amplified by PCR using pHT484 as a template with primers ON136 5'-gcg tgc tct tcc aac atg ggt ggc afg ggt ggc afg ggt ggc afg ggt aag ggc gag cag ctc ttc -3', ON180 5'-gcc cat gga tcc taa ctt gta cag ctc gtt gcc gag -3'. The SapI- and BamHI-treated PCR fragments were ligated into pTWIN1 at SapI and BamHI using T4 DNA ligase resulting in plasmids of FGG-mYFP, FGG-dCFP, FGG-dYFP and MGG-dYFP, respectively. The plasmid MGG-mYFP was generated from pHT584 template by a single point mutation (F→M) using forward primer ON30 5'-gac atc att gta cac aac atg ggt ggc gca agc tgg agc-3', and reverse primer ON31 5'-gct cca gct tgc gcc acc cat gtt gtc tag aat gat gtc-3'.

Amino acid sequences of proteins:

FGG-mYFP

FGGVSKGEEELFTGVVPVLVLEDGDVNGHKSVESGEGDATYKLTKFICTTGKLPVPWPTLVTTFGYGLQCFARYPDHMKQHDFFKASMEPGYVQERTIFFKDDGNKTRAEVKFEGDTR VNRIELKIDFKEDGNILGHKLEYNHNYIMADKQKNGIKVNFKIRHNIEDGSLADHYQQNTPIGDGPVLLPDHNYLSYQSKLSKDPNEKHDMVLLFVTAAAITLGMDELYK
Cucurbit[8]uril-mediated protein homotetramerization

FGG-dCFP
FGGVSKGEELFTGVPIVEGDGDVGKFSVSGEGEDATYGLTLKLFICTTGKLPVWPLTVTTLTWGVQCSRPHYDMMKQHDFKKSAMPEGYQYERTIFFKDDGNYKTRAEEKFEDGVTNRILEKGDIFKEDGNLGHKLQYNHSHWYIMADKQNGIQKVENFKIRHNIEDGSVQLADHYYQNNTPIDGVPVLLPDNHLYLSQSALF KDPMKREDHMVLLEFLTAAGITLMGDELYK

FGG-dYFP
FGGVSKGEELFTGVPIVEGDGDVGKFSVSGEGEDATYGLTLKLFICTTGKLPVWPLTVTTFGYGLQCFARYPDMMKQHDFKKSAMPEGYQYERTIFFKDDGNYKTRAEEKFEDGVTNRILEKGDIFKEDGNLGHKLQYNHSHWYIMADKQNGIQKVENFKIRHNIEDGSVQLADHYYQNNTPIDGVPVLLPDNHLYLSQSALF KDPMKREDHMVLLEFLTAAGITLMGDELYK

MGG-mYFP
MGGASWSHPOFESAVMSKGEELFTGVPIVEGDGDVGKFSVSGEGEDATYGLTLKLFICTTGKLPVWPLTVTTFGYGLQCFARYPDMMKQHDFKKSAMPEGYQYERTIFFKDDGNYKTRAEEKFEDGVTNRILEKGDIFKEDGNLGHKLQYNHSHWYIMADKQNGIQKVENFKIRHNIEDGSVQLADHYYQNNTPIDGVPVLLPDNHLYLSQSALF KDPMKREDHMVLLEFLTAAGITLMGDELYK

MGG-dYFP
MGGSGGSGSGGSGKGEELFTGVPIVEGDGDVGKFSVSGEGEDATYGLTLKLFICTTGKLPVWPLTVTTFGYGLQCFARYPDMMKQHDFKKSAMPEGYQYERTIFFKDDGNYKTRAEEKFEDGVTNRILEKGDIFKEDGNLGHKLQYNHSHWYIMADKQNGIQKVENFKIRHNIEDGSVQLADHYYQNNTPIDGVPVLLPDNHLYLSQSALF KDPMKREDHMVLLEFLTAAGITLMGDELYK

4.4.2 Protein expression and purification

The plasmids FGG-mCFP, FGG-mYFP, FGG-dCFP, FGG-dYFP, MGG-mYFP and MGG-dYFP were transformed into E. coli strain BL21(DE3). The bacteria were cultured in LB medium containing 100 μg/ml of ampicillin, and the cells were grown at 37 °C, shaking 250 rpm to an A600 of 0.5, then IPTG was added to a final concentration of 0.5 mM. The cells were continuously incubated overnight at 15 °C, shaking 250 rpm before being harvested. The pellet was resuspended into the bugBuster protein extraction reagent plus bezonase nuclease (to degrade DNA and RNA) and the insoluble material was removed by centrifugation at 20,000 rpm for 30 minutes at 4 °C. The soluble fraction was applied to a column filled with chitin beads (New England Biolabs) through gravity flow, and the column was washed with 20 volumes of sodium phosphate buffer 10 mM, pH 7 and subsequently overnight incubated at room temperature. The column was then eluted with sodium phosphate 10 mM, pH 7 buffer, and the target protein was collected. All proteins were dialysed in phosphate buffer, concentrated using Amicon Ultra 10,000 NMWL (Milipore) and kept at -80 °C until further use. The target proteins were analyzed by SDS-PAGE and ESI-MS.
4.4.3 LC-ESI-MS analysis of proteins

The LC-ESI-MS experiments were performed using a LCQ Fleet (Thermo Scientific). Proteins went through the C4 column under elution with a water/acetonitrile gradient (5% water: 95% acetonitril for 1 minute to 70% water: 30% acetonitril in 9 minutes and stay at 70% water: 30% acetonitril for 2 minutes) with 0.1 % formic acid before they were introduced to the ion source and mass spectrometry. The mass of the proteins was deconvoluted by Magtran software (Table 1).

Table 1. Calculated and measured mass of proteins

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Calculated mass (Da)</th>
<th>Measured mass (Da)</th>
<th>Error (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGG-mYFP</td>
<td>27158</td>
<td>27162</td>
<td>+4</td>
</tr>
<tr>
<td>FGG-dYFP</td>
<td>27176</td>
<td>27178</td>
<td>+2</td>
</tr>
<tr>
<td>FGG-dCFP</td>
<td>27090</td>
<td>27093</td>
<td>+3</td>
</tr>
<tr>
<td>MGG-dYFP</td>
<td>27763</td>
<td>27766</td>
<td>+3</td>
</tr>
<tr>
<td>MGG-mYFP</td>
<td>28630</td>
<td>28626</td>
<td>-4</td>
</tr>
</tbody>
</table>

4.4.4 Fluorescence anisotropy

The fluorescence anisotropy measurements were carried out at 2 μM of protein in the absence and presence of 2 μM of cucurbit[8]uril in 10 mM sodium phosphate buffer, pH 7 by using the Cary Eclipse fluorescence spectrophotometer (Varian). All the samples were prepared under room temperature in quartz cuvettes of 0.5 cm path length and 300 μl minimal volume. The fluorescence anisotropy data were recorded by using an excitation wavelength of 500 nm and scan spectra over a range of emission wavelengths (524-532 nm) at 20 °C.

4.4.5 Dynamic light scattering

Dynamic light scattering study was performed by using the Zetasizer μV (Malvern Instruments Limited, UK). For FGG-proteins, the isolated proteins were analyzed at concentration of 40 μM, and the complex of protein and cucurbit[8]uril was analyzed at ratio of (40 μM protein: 40 μM cucurbit[8]uril). For MGG-proteins, the isolated proteins were analyzed at a concentration of 30 μM, and the complex of protein and cucurbit[8]uril was analyzed at ratio of (30 μM protein: 30 μM cucurbit[8]uril). All samples for measurement were diluted in 10 mM sodium phosphate, pH 7 and then filtered via 0.1 μm syringe filter into a clean cuvette. The cuvette was inserted into the unit for 2 minutes of equilibration at 20 °C before a measurement was performed.
4.4.6 Size exclusion chromatography

The dimeric protein (FGG-dCFP) and cucurbit[8]uril-induced oligomerized protein (FGG-dCFP) were analyzed on a superdex 200 column (GE Healthcare) (LC-20AD, Shimadzu). The samples were analyzed and eluted in a sodium phosphate buffer (10 mM, pH 7) at the constant flow rate of 0.1 mL/min and absorption detection was performed at 514 nm. 40 μL of sample was analyzed in each run, consisting of 40 μM protein and variant concentrations of cucurbit[8]uril.

A SEC calibration curve was generated using the following protein standards (MW-GF-200, Sigma) as markers (Fig. 8):
1. Amylase: 200,000 Da
2. Bovine serum Albumin (BSA): 66,000 Da
3. Cyan fluorescent protein (CFP): 28,500 Da
4. Myoglobin: 17,600 Da

Detection of these proteins was accomplished using a PDA detector, scanning from 200 nm-600 nm. The proteins showed the following exclusion times:

<table>
<thead>
<tr>
<th>Calibration</th>
<th>Amylase (min)</th>
<th>BSA (min)</th>
<th>CFP (min)</th>
<th>Myoglobin (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
<td>12.4</td>
<td>14.375</td>
<td>16.2</td>
<td>18.1</td>
</tr>
<tr>
<td>Log (M.W)</td>
<td>5.3</td>
<td>4.8</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>M.W (Da)</td>
<td>200,000</td>
<td>67,000</td>
<td>28,500</td>
<td>17,600</td>
</tr>
</tbody>
</table>

![Figure 8. Calibration curve of size exclusion chromatography, and data fitted with a quadratic.](image-url)
4.5 References


Chapter 5

Cucurbit[8]uril mediated supramolecular dimerization and activation of caspase-9

Abstract

Control over protein dimerization and activation using synthetic molecules is at the forefront of biomedical research. The recognition of specific short peptide motifs in proteins by a synthetic supramolecular receptor addresses a need in this respect for novel inducers of protein dimerization. Cucurbit[8]uril is a small synthetic supramolecular host molecule that can selectively recognize and dimerize an engineered caspase-9, via a genetically encoded N-terminal FGG motif. Cucurbit[8]uril induced caspase-9 dimerization leads to strong enzymatic activation of the induced protein assembly. The supramolecular induced caspase-9 protein dimer is 50-fold more catalytically active than the caspase-9 monomer and can be fully reversed using a competitor ligand. Supramolecular induction of caspase-9 dimerization leads to activation, superior to conventional techniques based on protein engineering of the dimerization motif. Cucurbit[8]uril is a supramolecular inducer of dimerization, which facilitates the formation of the caspase-9 dimer at low concentrations and simultaneously sustains the correct active site rearrangement upon dimerization. In our opinion this system holds great promises, not only for studying caspase-9 dimerization and activation, but also for the huge array of other protein homodimerization events, such as for example observed in other dimerizing enzymes and membrane proteins.
5.1 Introduction

Caspase-9 is a critical protein in the apoptosis pathway and belongs to the caspase cysteine protease family responsible for cleaving proteins at specific aspartate residues[1-4]. In the cell, caspase-9 (casp-9) exists primarily in its inactive monomeric form (zymogen), becoming activated only upon induced dimerization by other auxiliary factors[5-8]. For example, apoptosomes, which include the apoptotic protease activating factor 1 (Apaf-1) and cytochrome c (from mitochondrion), are responsible for the recruitment of monomeric casp-9 to increase the local concentration levels of the protein in favour of dimer-driven activation. Once activated, casp-9 initiates cleavage of procaspase-3, generating their active forms, which signal for apoptosis (Fig. 1).

Figure 1. The intrinsic apoptotic pathway is mediated by casp-9. Under stress stimuli, cytochrome c is released from the mitochondria and associates with the apoptotic protease activating factor 1 (Apaf-1) and procasp-9, forming the apoptosome. Casp-9 becomes active by proteolytic processing and is released from the apoptosome. Active casp-9 cleaves and activates caspase-3, resulting in apoptosis[9].

Casp-9 exists primarily in its inactive monomeric state in dilute solution and simple methods to control casp-9 dimerization independent from the apoptosome are absent. The absence of such biochemical methods severely limits the molecular control achievable over casp-9 dimerization. This hampers for example the study of the molecular effects of the interplay between casp-9 dimerization and conformational changes. Only via engineering of
the casp-9 dimerization interface using specific point mutations, a constitutively dimeric casp-9 variant could be generated up to now\textsuperscript{[11]}. This dimeric casp-9 featured an approximately 5-fold higher catalytic activity compared to the wild type casp-9, but was still much less active than the Apaf-1 activated casp-9. Principle limitations of these engineered dimeric casp-9 constructs are their non-reversibility and possible effect of the point mutations on the conformation of the active site upon dimerization. New approaches to reversibly control casp-9 dimerization and activation are therefore highly sought.

![Figure 2. Cucurbit[8]uril induced casp-9 dimerization. a) Schematic representation of N-terminal FGG-bearing monomeric caspase-9 (large and small subunits) and its dimerization into an enzymatically active homodimer by supramolecular-induced host-guest complexation with cucurbit[8]uril. b) Crystal structure representation of a casp-9 dimer (2AR9)\textsuperscript{[11]}, portraying the close localization of the two N-termini of the two large subunits.]

Control over protein dimerization and activation using synthetic molecules is on the forefront of biomedical research\textsuperscript{[12-14]}. To be amenable to small molecules, proteins of interest typically require the introduction of additional protein domains. Control over protein dimerization via a very small, genetically encoded, peptide motif is highly attractive in this respect as a short peptide motif would impose the smallest possible influence on the protein of study. Synthetic supramolecular molecules such as the donut-shaped cyclodextrins and cucurbiturils\textsuperscript{[15]} and other more complex synthetic receptors\textsuperscript{[16]} have been shown to selectively recognize amino acids and protein elements\textsuperscript{[17]}. Strong and specific recognition of two short peptide motifs simultaneously can be best achieved using cucurbit[8]uril as supramolecular host molecule\textsuperscript{[18]}. Introduction of an N-terminal phenylalanine-glycine-glycine (FGG) motif in fluorescent proteins, allows cucurbit[8]uril to selectively induce the homodimerization of these proteins\textsuperscript{[19]}. This supramolecular inducer of protein dimerization acts via a molecular mechanism that involves the selective recognition of the two N-terminal phenylalanines via hydrophobic interactions of the phenyl groups inside the cucurbit[8]uril cavity and interactions of the protonated N-terminal amine functionality with the carbonyl rim of the cucurbit[8]uril\textsuperscript{[18]}. Since it is postulated that the key step for casp-9 activation relies on enhancing the dimer formation, we envisaged to use the cucurbit[8]uril-FGG supramolecular
system as a novel way to reversibly induce caps-9 dimerization and activation. Herein, we describe the generation of engineered caps-9 constructs featuring an N-terminal FGG motif and the applicability of cucurbit[8]uril as inducer of caps-9 dimerization (Fig. 2). Cucurbit[8]uril induces caps-9 dimerization at low concentration with concomitant strong enzymatic activation. The supramolecular induced caps-9 protein dimerization can be fully reversed and is superior to the constitutively dimeric caps-9 variant in terms of catalytic activity.

5.2 Results and Discussion

5.2.1 Protein engineering

Casp-9 (residues 140-416) was used for study of cucurbit[8]uril-induced caps-9 dimerization and activation. According to crystal structure of a native caps-9 dimer, the two N-termini of position 140 are closed together[7]. Therefore, caps-9 was provided with an FGG motif (without linker) at its N-terminus and this construct was genetically fused with an intein domain, as autocleavable N-terminal tag in a pTWIN1 plasmid[19]. The modified plasmid was transformed and expressed in BL21 (DE3) cells. The fusion protein was captured on a chitin beads column and the FGG-casp-9 protein was eluted after induced auto-cleavage of the intein at room temperature and pH 7.0. As reference an MGG-casp-9 construct was generated, featuring an N-terminal methionine in place of the phenylalanine. Additionally, a dimeric FGG-dimercasp-9 construct was generated, bearing the previously reported constitutively dimeric interface, consisting of the replacement of the Gly102-Cys-Phe-Asn-Phe106 sequence in the β6 strand with Cys264-Ile-Val-Ser-Met268 (Fig. 3). All proteins were isolated in purity, featuring a large and a small subunit, as evidenced by SDS-PAGE and LC-ESI-MS (Fig. 4, 10).

![Figure 3. Schematic representation of engineering FGG-casp-9 and FGG-dimercasp-9 (dimeric interface) from casp-9 (residues 140-416).](image-url)
5.2.2 Protein dimerization

Cucurbit[8]uril mediated casp-9 dimerization was studied using dynamic light scattering (DLS) (Fig. 5). The reference construct MGG-Casp-9 scattered with an average particle size of 2.9 nm. This corresponds, assuming a spherical particle, to a molecular weight of ~40 kDa and is confirmed with the calculated mass of 32 kDa. Addition of cucurbit[8]uril to this protein did not change the scattering function and size of the particles. The size of the individual FGG-casp-9 was determined to be 2.8 nm which corresponds to an approximate mass of 37 kDa, nicely in line with its calculated mass (32 kDa) and the results observed for MGG-casp-9. Upon addition of cucurbit[8]uril to this protein, an increase in the particle size to 3.5 nm was observed. This change correlates with an novel approximate protein mass of 63 kDa, which nicely mirrors the change seen in the calculated mass from monomeric to dimeric FGG-casp-9 (64 kDa). Cucurbit[8]uril thus induces an increase in the particle size, corresponding to a supramolecular-induced dimerization of FGG-incasp-9, in line with previously observed results on cucurbit[8]uril mediated dimerization of fluorescent proteins[19].
5.2.3 Protein activation

The correlation between supramolecular protein dimerization and enzyme activation was studied using enzyme activity tests. To this end, the catalytic activity of the different casp-9 constructs in cleaving the synthetic fluorescent casp-9 substrate Ac-LEHD-AFC (N-acetyl-Leu-Glu-His-Asp-7-amino-4-trifluoromethyl coumarin) was studied in the presence and absence of cucurbit[8]uril (Fig. 6a-b). The MGG-casp-9 showed a background activity of 104 Units per mg enzyme (U/mg). Addition of cucurbit[8]uril to this protein did not elicit a change in the enzyme activity (99 Units/mg). These results show that cucurbit[8]uril does not dimerize and activate a normal casp-9 variant and that it also does not affect the active site or the periphery of the enzyme. The FGG-casp-9 has a similar weak background activity, proving that the additional N-terminal phenylalanine has no intrinsic effect on the casp-9 properties. However, addition of cucurbit[8]uril to this protein leads to a ~50-fold increase in catalytic activity over the cucurbit[8]uril free state (Fig. 6a, table 1). Interestingly, the ternary formation constant of cucurbit[8]uril: casp-9 (1:2) complex showed about $2.7 \times 10^{12}$ M$^{-2}$, which is approximately 20-fold higher than that of cucurbit[8]uril:FGG: (ternary $K = 1.5 \times 10^{11}$ M$^{-2}$) complex$^{18}$. The increased ternary formation constant is results of supramolecular interaction (between cucurbit[8]uril and FGG peptide motif) and protein-protein interaction. This shows that the binding of cucurbit[8]uril to the FGG-casp-9 N-terminus and subsequent casp-9 dimerization allows for the formation of a highly enzymatically active casp-9 dimer.

![Figure 6. Cucurbit[8]uril-induced casp-9 enzymatic activity. a) Activity for the cleavage of casp-9 substrate Ac-LEHD-AFC of different casp-9 variants (at 0.15 μM) in the absence (grey bars) and presence (dark bars) of cucurbit[8]uril (4 μM). b) Concentration effect of cucurbit[8]uril on FGG-casp-9 (at 0.15 μM) activity on Ac-LEHD-AFC cleavage.](image)

The FGG-dimercasp-9 did show an increased activity in the absence of cucurbit[8]uril. This constitutive dimeric variant was 7-fold more active than the normal casp-9 in the absence of cucurbit[8]uril. The observed activation of casp-9 via the engineered dimerization interface is fully in line with the reported constructs of this protein$^{11}$. Addition of cucurbit[8]uril to the
FGG-dimercasp-9 construct induced an additional 3-fold increase in catalytic activity. The FGG-casp-9 activity enhancement by cucurbit[8]uril is concentration dependent (Fig. 6b), in line with a reversible recognition and binding event. Enzyme activity increases upon increased addition of the supramolecular inducer of dimerization until a maximal activity is reached when all protein is dimerized via cucurbit[8]uril.

In the apoptotic pathway, casp-9 catalyzes, amongst others, the regeneration and activation of caspase-3. The catalytic activity of the cucurbit[8]uril induced casp-9 dimerization was therefore determined using caspase-3 as natural substrate (Fig. 7, Table 1)[20]. The FGG-casp-9 requires approximately 22 hours to cleave half of the caspase-3 substrate (Fig. 7a) in its monomeric form in the absence of cucurbit[8]uril. Addition of cucurbit[8]uril to this solution results in a strong enhancement of catalytic activity against the natural substrate, resulting in a half-time of cleavage of only 20 minutes (Fig. 7b). This demonstrates that the catalytic activity on natural substrate for the cucurbit[8]uril induced FGG-casp-9 dimer is faster than the catalytic activity of the FGG-casp-9 without supramolecular inducer of dimerization (Table 1). The catalytic activity of the reference construct MGG-casp-9 is similar to that of FGG-casp-9 without cucurbit[8]uril, but in contrast does not increase upon addition of the host molecule (Fig. 12, Table 1). The FGG-dimercasp-9 again shows a higher basal activity due to the intrinsic dimerization behaviour. Addition of cucurbit[8]uril to this solution enhanced the activity, but does not reach the enzymatic activity level of the FGG-casp-9 in the presence of cucurbit[8]uril (Table 1).

Cucurbit[8]uril very strongly induces the formation of FGG-casp-9 dimers resulting in high enzymatic activity of these complexes. Also, supramolecular induced casp-9 dimerization and activation is significantly more efficient than engineering of the dimerization interface. This is illustrated by the additional increase in FGG-dimercasp-9 activity upon cucurbit[8]uril addition and by the overall higher activity of the FGG-casp-9 construct in the presence of cucurbit[8]uril. The significant increase in activity of cucurbit[8]uril induced FGG-casp-9 dimerization supports the notion that the mechanism of casp-9 activation is primarily related to protein dimerization[6,7]. However, the results also show that the molecular pathway via how protein dimerization is induced can significantly influence the activity of the resulting casp-9 assembly[21,22]. The lower catalytic activity of the FGG-dimercasp-9/cucurbit[8]uril complex compared to the FGG-casp-9/cucurbit[8]uril complex demonstrates that allosteric effects involved in protein dimerization could have an impact on the catalytic activity. In case of the caspase-9 constructs with a mutated dimerization interface[81], the active site might be less proximally, compared to the facilitated casp-9 dimerization with cucurbit[8]uril. The inserted mutations in the dimerization interface apparently lead to a molecular packing in the casp-9 dimer diverged from the native state. Additional facilitated dimerization in the FGG-dimercasp-9 by the cucurbit[8]uril does further enhance the activity of the enzyme, but not to the extent of the dimer which is
exclusively formed by the cucurbit[8]uril mediated dimerization and the natural casp-9 dimerization interface. These results therefore suggest that a dimerization event alone may not be sufficient for optimal casp-9 activation; rather an interplay of active site rearrangement induced by an appropriate protein dimerization event, is required for optimal activity.

Figure 7. Cleavage activity of FGG-casp-9 (0.15 μM) for casp-3 (4 μM) in the absence (a) and presence (b) of cucurbit[8]uril (4 μM) in buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl₂ and 2 mM TCEP, pH 6.5) at 37 °C. Caspase-3 fl (full length); ls (large subunit); ss (small subunit). c) Quantitation of substrate cleavage by FGG-casp-9 in the absence (triangles) and presence (circles) of cucurbit[8]uril.
Table 1. Catalytic activity of casp-9 on the synthetic and natural substrate.

<table>
<thead>
<tr>
<th></th>
<th>Ac-LEHD-AFC (U/mg)</th>
<th>Time required for 50% cleavage of caspase-3 by 0.15 μM of casp-9 (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–Q8</td>
<td>+Q8</td>
</tr>
<tr>
<td>MGG-casp-9</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>FGG-casp-9</td>
<td>50</td>
<td>2600</td>
</tr>
<tr>
<td>FGG-dimercasp-9</td>
<td>370</td>
<td>1000</td>
</tr>
</tbody>
</table>

*U/mg: One unit will cleave 1.0 nmole of substrate Ac-LEHD-AFC per hour at pH 6.5 at 37 °C
Q8: Cucurbit[8]uril
FE: Fold Enhancement

5.2.4 Reversible blockade

The supramolecular induction of FGG-casp-9 dimerization and activation critically depends on the supramolecular complexation with cucurbit[8]uril. Such a supramolecular assembly process provides the opportunity to reverse the protein dimerization with a competitor molecule. Concomittantly, the reversibility of the cucurbit[8]uril induced casp-9 activation was evaluated via inhibition of complex formation upon addition of a competitor peptide. For this, the enzymatic active complex of cucurbit[8]uril and FGG-casp-9 was treated with increasing amounts of the short FGG peptide (Fig. 8). Upon addition of the competitor peptide the enzymatic activity of the casp-9 was decreased in a dose-dependent fashion. At an excess of FGG peptide, the FGG-casp-9 is restored in its monomeric state (inactive form), featuring only background activity. The reversibility of the cucurbit[8]uril – FGG system illustrates the full control achievable over casp-9 dimerization and activation with the supramolecular approach and the potential to either induce of block protein dimerization at will with specific small host-guest molecules.
5.3 Conclusions

Cucurbit[8]uril is a supramolecular inducer of dimerization, which facilitates the formation of the casp-9 dimer at low concentrations and simultaneously sustains the correct enzyme rearrangement upon dimerization. Casp-9 comprising a short genetically-encoded N-terminal FGG motif can be brought to assemble in a highly catalytically active dimer by host-guest complex formation with cucurbit[8]uril. The activity of cucurbit[8]uril-induced casp-9 dimers is not only significantly greater than that of the isolated protein, but is also superior to the protein surface engineering approach. The results show that two monomers of casp-9 can be activated through facilitated proximity. The results support the hypothesis that the mechanism of casp-9 activation is through a combination of dimerization and rearrangement of an active site conformation. In addition, cucurbit[8]uril-induced FGG-casp-9 dimerization can be reversed by challenging the complex with a competitive FGG peptide. It is therefore possible for supramolecular host-guest chemistry to control the activity of casp-9, both via increasing casp-9 activity by supramolecular-induced dimerization and via decreasing its activity by reversal of the complex formation using a competitive ligand. Such a reversible supramolecular host-guest control over casp-9 activation could prove advantageous over protein engineering approaches yielding constitutively active caspase-9 variants\textsuperscript{[11,22]}. Controlling casp-9 activity using a supramolecular inducer of protein dimerization carries much potential for studying the biological mechanisms of this and related caspasers, for example in elucidating their exact molecular role and mechanism in the apoptotic pathway in mammalian cells. Cucurbit[8]uril-induced protein homodimerization thus provides a novel concept for inducers of protein dimerization which utilize only a small genetically encoded tripeptide motif in combination with a small synthetic molecule as inducer. In our opinion this system holds great promises, not only for studying casp-9 dimerization and activation as
shown above, but also for the huge array of other protein homodimerization events\cite{23}, such as for example observed in other dimerizing enzymes and membrane proteins\cite{24-29}.

5.4 Experimental section

5.4.1 Construction of plasmids

The human caspase-9 (casp-9) protein with N-terminal FGG motif was generated by using the N-terminal intein from plasmid pTWIN-1 (New England Biolabs)\cite{19}. DNA encoding for FGG-casp-9 (consisting of a large and a small subunits) was amplified by PCR using pET21-casp-9 containing human casp-9-encoding gene\cite{30} as a template and forward primer ON313, 5’-atcattgtaacattgtggcggccctctttgagttttagagggcgg-3’ and reverse primer ON 306, 5’-gttagcagcgggtcatcctgttagcgcgggtattctgag-3’. This PCR product was cloned into treated-pTWIN1 at SapI and BamHI using in-fusion enzyme (Clontech), resulting in plasmid FGG-casp-9-encoding gene fused with intein. The schematic representation for construction of FGG-casp-9 is shown in figure 9.

![Figure 9. Construction of plasmid for FGG-caspase-9 expression](image)

The plasmid MGG-casp-9 was generated from FGG-casp-9 template by single point mutation (F→M) using forward primer ON44, 5’-gacatcattgtaacattgtggcggccctctttgagttttagagggcgg-3’ and reverse primer, ON45, 5’-ctcaagagcaggctcatcgtttagtagctgac-3’. Plasmid FGG-dimercasp-9 (dimer interface) containing five-point mutation (GCFNF→CIVSM)\cite{11} was generated from FGG-casp-9 template using forward primer ON42, 5’-tataacagatcttgctttcatcgttctactgccggaaaaatttttct-3’ and reverse primer ON43, 5’-
ttcggagcatagaaacgatacaaggcatctgtttataaatccctttcac-3'. The amino acid sequences of proteins are as below.

**FGG-casp-9**

FGGGALESLRGNADLAYLSEMPCGHCLIINNVNFCREGLRTRTGSNIDCEKLRRRFSSL
HFMEVEKGDLTAKKMLVLALLELAQQDHGALDCVVVILSHGCQASHLQFPGAVYGTDG
CPVSVEKIVNIFNGTSCPSLGGKPCLFFIQACGGEQKDHGEVASTPEDESPGSNPEDATP
FQEGLRTFDQLDAISLPTPSIDFVSYSTFPGEVSRPKSGSFWYTLEDFEQWAHSEDLLQ
SLLLRVANAVSVKGYKQMPGCFNFLRKKLFFKTSAAALEHHHHHH

**MGG-casp-9**

MGGGALESLRGNADLAYLSEMPCGHCLIINNVNFCREGLRTRTGSNIDCEKLRRRFSSL
LHFMEVEKGDLTAKKMLVLALLELAQQDHGALDCVVVILSHGCQASHLQFPGAVYGTDG
GCPVSVEKIVNIFNGTSCPSLGGKPCLFFIQACGGEQKDHGEVASTPEDESPGSNPEDATP
TFQEGLRTFDQLDAISLPTPSIDFVSYSTFPGEVSRPKSGSFWYTLEDFEQWAHSEDLLQ
LQSLLRVANAVSVKGYKQMPGCFNFLRKKLFFKTSAAALEHHHHHH

**FGG-dimercasp-9**

FGGGALESLRGNADLAYLSEMPCGHCLIINNVNFCREGLRTRTGSNIDCEKLRRRFSSL
HFMEVEKGDLTAKKMLVLALLELAQQDHGALDCVVVILSHGCQASHLQFPGAVYGTDG
CPVSVEKIVNIFNGTSCPSLGGKPCLFFIQACGGEQKDHGEVASTPEDESPGSNPEDATP
FQEGLRTFDQLDAISLPTPSIDFVSYSTFPGEVSRPKSGSFWYTLEDFEQWAHSEDLLQ
SLLLRVANAVSVKGYKQMPCIVSMRLKKLFFKTSAAALEHHHHHH

5.4.2. Protein expression and purification

The plasmids FGG-casp-9, MGG-casp-9 and FGG-dimercasp-9 were transformed into *E. coli* strain BL21 (DE3). The bacteria were cultured in LB medium containing 100 μg/ml of ampicillin and the cells were grown at 37 °C, shaking 250 rpm to an *A*<sub>600</sub> of 0.7, then IPTG was added to a final concentration of 0.3 mM. The cells were continuously incubated overnight at 15 °C, shaking 250 rpm before being harvested. The pellet was resuspended into the BugBuster protein extraction reagent plus benzonase nuclease (to degrade DNA and RNA) and the insoluble material was removed by centrifugation at 20,000 rpm for 40 minutes at 4 °C. The soluble fraction was applied to a column filled with chitin beads (New England Biolabs) through gravity flow, and the column was washed with 40 volumes of sodium phosphate buffer (20 mM, 100 mM of sodium chloride, pH 7) and subsequently incubated about 6 hours at room temperature. The cleaved proteins were then collected in the flow-through using the phosphate buffer. The target proteins were analyzed by SDS-PAGE and LC-ESI-MS (Fig. 4, 10).
5.4.3 LC-ESI-MS analysis of proteins

The LC-ESI-MS experiments were performed using a LCQ Fleet (Thermo Scientific). Casp-9 consisting of large and small subunits were separated by C4 column in water/acetonitrile gradient (5% water: 95% acetonitril for 1 minute to 70% water: 30% acetonitril in 9 minutes and stay at 70% water: 30% acetonitril for 2 minutes) with 0.1% formic acid before they were introduced to the ion source and mass spectrometry. The mass of proteins were deconvoluted by Magtran software (Fig. 10).

![Figure 10a. LC-ESI-MS analysis of proteins. LC: large subunit at 7.5 minutes, small subunit at 7.9 minutes. A) MS for FGG-casp-9, large subunit: 19146 Da (found), 19143.7 Da (calc); small subunit: 12827 Da (found), 12826.4 Da (calc). B) MS for MGG-casp-9, large subunit: 19128 Da (found), 19131 Da (calc); small subunit: 12828 Da (found), 12826.4 Da (calc).]
Figure 10b. LC-ESI-MS analysis of proteins. LC: large subunit at 7.5 minutes, small subunit at 7.9 minutes. C) MS for FGG-dimercasp-9, large subunit: 19144 Da (found), 19143.7 Da (calc); small subunit: 12792 Da (found), 12791.5 Da (calc).

5.4.4 FGG tripeptide synthesis

The FGG peptide was synthesized by automated solid phase peptide synthesis (SPPS) techniques using standard 9-fluorenylmethoxycarbonyl (Fmoc)-coupling chemistry on a preloaded Fmoc-Gly-Wang resin. Fmoc deprotection was performed twice with 20% (v/v) piperidine in N-methyl-2-pyrrolidinone (NMP) for 5 min. The following Glycine (G) and Phenylalanine (F) amino acids were coupled twice for 20 minutes at room temperature using 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylouronium hexafluorophosphate (HBTU) and N,N-Diisopropylethylamine (DIPEA) as coupling reagents. Each amino acid coupling step and Fmoc deprotection step was followed by washing the resin with NMP for several times. After the Fmoc was removed from the N-terminus the peptide was cleaved from the resin by using 96% Trifluoroacetic acid (TFA), 2.0% Triisopropylsilane (TIS) and 2.0% H2O for 1 hour at room temperature.

After precipitation in diethyl ether the peptide was lyophilized from water resulting in a white fluffy powder. The peptide was characterized using liquid chromatography-mass spectrometry (LC-MS).

5.4.5 Dynamic light scattering

Dynamic light scattering was performed by using the Zetasizer µV (Malvern Instruments Limited, UK). FGG-casp-9 and reference MGG-casp-9 for measurement were diluted in 10 mM sodium phosphate, pH 7. The protein or the mixture of protein and cucurbit[8]uril was
filtered via 0.1 μm syringe filter into a clean cuvette. The cuvette then was inserted into the unit for 2 minutes of equilibration at 20 °C before a measurement was performed. The experiment was performed in duplicate.

5.4.6 Activity of casp-9 assay on synthetic substrate

The assay was based on the hydrolysis of the peptide substrate Ac-LEHD-AFC (Acetyl-Leu-Glu-His-Asp-7-amino-4-trifluoromethyl coumarin, Biomol) by casp-9 resulting in release of an AFC (7-amino-4-trifluoromethyl coumarin) moiety. The product AFC was detected by using an excitation wavelength of 400 nm and an emission wavelength of 505 nm.

The activity assays were carried out at 0.15 μM of casp-9 (FGG-casp-9, MGG-casp-9, FGG-dimercasp-9 in the absence or presence of cucurbit[8]uril) in 100 μL of assay buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl₂ and 2 mM TCEP, pH 6.5) with 100 μM of Ac-LEHD-AFC substrate. Kinetic casp-9 activity measurements were performed in duplicates using a 96 well plate and measured by Tecan Safire II at 37 °C. Specific activity of casp-9 was then calculated by using AFC calibration curve. Titration of cucurbit[8]uril to FGG-casp-9 was performed under similar conditions, where concentration of cucurbit[8]uril was titrated from 0 μM to 4.5 μM.

The AFC calibration curve was prepared in a series of AFC dilution with assay buffer in the range of concentration 0.05 μM to 10 μM (0.005-1 nmole per 100 μL) (Fig. 11).

![AFC calibration curve](image)

**Figure 11. AFC calibration curve**

5.4.7 Casp-9 assay on natural substrate

The experiments were carried out in mixtures of 0.15 μM casp-9 and 4 μM substrate (caspase-3) in the absence or presence of 4 μM cucurbit[8]uril diluted in assay buffer of 25 mM HEPES, 100 mM NaCl, 4 mM MgCl₂, 2 mM TCEP (pH 6.5). Reactions were incubated at 37 °C. Samples were collected in time intervals and reactions were stopped by addition of SDS-loading buffer. All samples were separated on 8 to 18% SDS-PAGE and proteins
detected with blue staining. The extent of caspase-3 (C163A) cleavage was determined by densitometry using an AlphaEaseFC.

**Figure 12A.** Catalytic activity of FGG-dimercasp-9 (0.15 μM) for natural substrate, caspase-3 (4 μM): a) in the absence of cucurbit[8]uril, b) in the present of cucurbit[8]uril (4 μM). Caspase-3 fl (full length); ls (large subunit); ss (small subunit).

**Figure 12B.** Catalytic activity of MGG-casp-9 (0.15 μM) for natural substrate, caspase-3 (4 μM): a) in the absence of cucurbit[8]uril, b) in the present of cucurbit[8]uril (4 μM). Caspase-3 fl (full length); ls (large subunit); ss (small subunit).

### 5.4.8 Fitting curve of cucurbit[8]uril titration to casp-9 activity

Cucurbit[8]uril titration to casp-9 activity could be fit by using a relatively simple model. The assumption in the fit is that casp-9 and cucurbit[8]uril exclusively form a (casp-9):cucurbit[8]uril complex in this concentration regime, which is a reasonable assumption.
based on the earlier finding that FGG binding is cooperative. The equation is used for fitting shown below.

\[
Y = \frac{P1^* ((4*P2) + 1/((\beta)^X)) - \text{SQRT}((-4*(P2)-1/((\beta)^X))^2 - 16*(P2)*(P2)))}{8}
\]

\(\beta\): the ternary formation constant
\(X\): cucurbit[8]uril concentration (M)
\(P2\): casp-9 concentration (M)

5.5 References

Chapter 6

Enhancement and reactivation of caspase-8 activity by cucurbit[8]uril induced dimerization

Abstract

Cucurbit[8]uril is brought forward as a macrocyclic host molecules competent to reversibly control dimerization and activation of caspase-8. Caspase-8 was genetically fused with a very short FGG peptide tag at the N-terminus (FGGCaspase-8) allowing cucurbit[8]uril to selectively recognize and induce FGGcaspase-8 dimerization, resulting in an increase of its activity. The activity of the cucurbit[8]uril-induced FGGcaspase-8 dimer was increased in a cucurbit[8]uril dose-dependent manner. In addition, two mutated fully monomeric FGGcaspase-8 analogues, FGGcaspase-8(D384A) and FGGcaspase-8 (F468A), were generated. Both FGGcaspase-8(F468A) and its cucurbit[8]uril-induced dimeric form did not show any enzymatic activity. The activity of the monomeric FGGcaspase-8(D384A) was 2.5 fold lower compared to wild type caspase-8 for the synthetic substrate (Ac-IETD-AFC). This FGGcaspase-8(D384A) protein, however, did not show any enzymatic activity for the natural substrate, caspase-3. Cucurbit[8]uril induced FGGcaspase-8(D384A) dimerization and fully reactivated the FGGcaspase-8(D384A) activity for both the synthetic and natural substrates. The enzymatic activity of cucurbit[8]uril-induced caspase-8 dimers could be reversed back to their basal activity via addition of a competitor FGG peptide. Control over protein dimerization by a supramolecular host guest system provides a flexible approach to control and modulate caspase-8 dimerization and activation and elucidate the molecular mechanisms of caspase-8 activity.
6.1 Introduction

Caspases are synthesized in the cell as inactive precursors, so called procaspases and consist of an N-terminal prodomain and a C-terminal protease domain. Their activation occurs when the prodomain is removed and the protease domain is proteolytically cleaved at specific aspartic acids into a small subunit and a large subunit. These two subunits dimerize very strongly forming a heterodimer and subsequently, two of such heterodimers together form an active complex\(^{[1-3]}\). An activated caspase can cleave other procaspases resulting in a cascade of activated caspases. There are numerous caspases, all with a different function in apoptosis. Initiator procaspases are activated by an intrinsic or extrinsic signal. Once activated, initiator caspases cleave and activate executioner procaspases such as procaspase-3 and procaspase-7 or target proteins inside the cell, resulting in apoptosis\(^{[1,2,4,5]}\).

![Schematic representation of the extrinsic apoptotic pathway](image)

**Figure 1.** Schematic representation of the extrinsic apoptotic pathway. The Fas ligand binds to the Fas receptor which allows FADD, Fas receptor and procaspase-8 to form a DISC complex. After DISC formation, procaspase-8 can be autocleaved, thereby generating an active caspase-8. The active caspase-8 cleaves and activates caspase-3, resulting in apoptosis.

Procaspase-8 is an initiator caspase which features two death effector domains (DEDs) and a catalytic domain that can be divided into a large and a small subunit with a linker in between. When a death signal from outside the cell, such as a Fas ligand, binds to a Fas death receptor (Fasr), it enables a Fas death domain (FADD) to bind to the Fas receptor inside the cell (Fig. 1). The FADD works as a link between the Fas death receptor and the death effector domains (DEDs) of procaspase-8. All these components together, (FADD, Fasr, procaspase-8) are called the death-inducing signaling complex (DISC)\(^{[6,7]}\). After DISC formation,
procaspase-8 can be autocleaved at proteolytic sites Asp216, Asp374 and Asp384, thereby generating an active caspase-8[8]. One hypothesis states that cleavage is not a requirement for the formation of an active site. The cleavage is merely a way to provide stability to the dimer generated during DISC formation. The most important step in caspase-8 activation is the dimerization of the monomers and involves a number of conformational changes. Upon death receptor stimuli, the procaspase-8 undergoes a proximity driven dimerization. The formed dimers are susceptible to cleavage by other procaspase-8 dimers at the region linking the large and small subunits at Asp374 and Asp 384. Thereafter, the site between the large subunit and the prodomain at Asp216 is accessible for cleavage. To ensure stabilization of caspase-8, the region between the two DEDs at Asp126 can be cleaved by caspase-8 itself, resulting in a mature caspase-8 that can be released into the cytosol[8].

Figure 2. Chemical induced proximity of caspase dimerization and activation. FKBP (rapamycin associated protein), FK1012 (dimeric FK506 analog).

Proximity driven dimerization is a crucial mechanism of caspase-8 activation[3]. A number of studies on designed caspase-8 constructs investigated this in more detail. Fusion of caspase-8 to a chemical binding domain protein (FKBP-rapamycin associated protein) allowed a dimeric FK506 analog to induce the caspase-8 fusion dimer, resulting in increased caspase-8 activation (Fig. 2)[9,10]. Structural and biochemical studies on caspase-8 provided new insights via an caspase-8 engineered fully monomeric caspase-8, which featured reduced or abolished caspase-8 enzymatic activity[11,12]. For all these cases either an activated caspase-8 was studied or a deactivated variant via engineering of the protein dimerization properties. However, the possibility for concomitant deactivation and reactivation of the same caspase-8 protein, allowing for control and comparison of the same constructs, has not been described. In this study, we demonstrate that a novel inducible dimerization approach via facilitated dimerization, known as supramolecular induced dimerization, can be smartly utilized not only to understand the mechanism of caspase-8 activation but also to reversibly control caspase-8 dimerization and reactivate the enzymatic properties of a monomeric variant.
Supramolecular chemistry has recently emerged as a powerful tool for reversible control protein dimerization\textsuperscript{[13-16]}. Control over protein dimerization by supramolecular systems requires introduction of a small supramolecular motif in a protein domain of interest. Selective recognition of this motif by a second supramolecular partner molecule through non-covalent interaction is a key factor to induce the close proximity of two proteins. Depending on the design of the supramolecular system, both homodimerization or heterodimerization can be achieved. For example, a lithocholic acid-conjugated cyan fluorescent protein can bind a β-cyclodextrin-conjugated yellow fluorescent protein leading to the formation of a protein heterodimer via the host-guest interaction between β-cyclodextrin and lithocholic acid\textsuperscript{[14,16]}. Controlled protein dimerization by means of a very small, genetically encoded peptide motif has been also demonstrated. Introduction of an N-terminal phenylalanine-glycine-glycine (FGG) motif in N-terminal fluorescent proteins, allows cucurbit[8]uril to selectively recognize and induce protein dimerization\textsuperscript{[13]}. The molecular mechanism of this system involves the selective recognition of the two N-terminal phenylalanines via hydrophobic interactions of the phenyl groups inside the cucurbit[8]uril.
cavity and interactions of the protonated N-terminal amine functionality with the carbonyl rim of the cucurbit[8]uril[17,18]. In all of the above described cases, the supramolecular elements stabilize an intrinsic protein interaction via an increase of the local concentration due to the supramolecular interaction of the appended motifs. This facilitated dimerization is significantly different from engineering the dimerization interface of proteins, which leads to a permanent change in dimerization affinity[19,20]. In this study, we describe the generation of wild type caspase-8 and monomeric-mutants of caspase-8 featuring an N-terminal FGG tag and their cucurbit[8]uril-induced dimerization (Fig. 3). Cucurbit[8]uril induced dimerization shows enhancement of wild type caspase-8 activity and full enzymatic reactivation of the inactive monomeric mutant caspase-8. The induction of caspase-8 dimerization and activation can be fully reversed by addition of a synthetic competitor for cucurbit[8]uril.

6.2 Results

6.2.1 Protein generation

Caspase-8(wt) exists in equilibrium between a monomeric and a dimeric state in buffered solutions (Fig. 3)[21]. Monomeric caspase-8 variants, caspase-8(D384A) (replacing aspartic acid with alanine at position of 384caspase-8) and caspase-8(F468A) (replacing phenylalanine with alanine at position of 468caspase-8) are fully monomeric in buffered solution, even at high protein concentrations[11]. The aspartic acid to alanine mutation completely prevents the cleavage process at amino acid site 384 of caspase-8(D384A), resulting in the small subunit still featuring the central 10 amino acid of the intersubunit linker (Fig. 4), which may sterically interfere with the dimer formation of caspase-8(D384A)[11,12]. Caspase-8(F468A) cannot cleave itself at all (Fig. 4). The phenylalanine to alanine mutation removes cross-strand intermolecular interactions between phenylalanine468 and proline466, making caspase-8(F468A) incompetent to dimerize[12]. The three proteins caspase-8(wt) and the monomeric mutants caspase-8(D384A) and caspase-8(F468A) (residue 217-479, catalytic domain of caspase-8) were selected to study the possibilities for cucurbit[8]uril induced facilitated caspase-8 dimerization and activation.

According to crystal structure of catalytic domain of caspase-8 dimer, the two N-temini of position 217 are closed together[31]. Therefore, caspase-8(wt), caspase-8(D384A) and caspase-8(F468A) were provided with an FGG motif or an MGG motif (without linker) at their N-terminus and these construct were genetically fused with an intein domain, as autocleavable N-terminal tag in a pTWIN1 plasmid[13]. The plasmid transformation and protein expression are described in experimental section. The intein-FGG/MGGcaspase-8 constructs were additionally fused with a chitin binding domain (CBD) which allows the CBD-intein-FGG/MGGcaspase-8 fusion proteins to bind to chitin beads for purification. The target caspase-8 featuring N-terminal FGG or MGG motif was then autocleaved by intein splicing.
at room temperature and pH 7 conditions. After isolation, the proteins were analyzed by SDS-PAGE and LC-ESI-MS, confirming their purity and integrity (experimental section).

![DNA constructs diagram](image)

**Figure 4.** Overview of caspase-8 proteins used in this study and their protein sequence characteristics. Caspase-8(wt) is fully processed into two subunits; Caspase-8(D384A) features only a single cleavage resulting in the linker sequence to remain attached to the small subunit; Caspase-8(F468A) cannot dimerize and be processed.

### 6.2.2 Caspase-8 activity on synthetic substrate

The enzymatic activity of isolated caspase-8 and cucurbit[8]uril-induced caspase-8 dimers was firstly determined for a synthetic substrate, Ac-IETD-AFC\(^{[22]}\). The active caspase-8 cleaves the substrate at the site next to aspartic acid, thereby releasing fluorescent AFC. The released AFC was detected by using an excitation wavelength at 400 nm and an emission wavelength at 500 nm. The rate of caspase-8 activity was determined by the time-dependent increase in AFC concentration (Fig. 5). The specific activity of the caspase-8 under study was subsequently calculated by using an AFC calibration curve (experimental section).
Enhancement and reactivation of caspase-8 activity by cucurbit[8]uril induced dimerization

Figure 5. Exemplary activity studies for FGGcaspase-8(wt) (0.15 μM) in the presence of cucurbit[8]uril (1 μM). Representative time-dependent spectra (A) and plot of the linear phase (B) of AFC production via caspase-8 cleavage of Ac-IETD-AFC.

In the absence of cucurbit[8]uril, FGGcaspase-8(wt) and FGGcaspase-8(D384A) featured specific activities of 606 units/mg and 238 units/mg, respectively (Fig. 6 and Table 1). The FGGcaspase-8(D384A) was still active, in line with previous literature reports\[^{[11]}\], and showed an around 2.5 fold lower activity than FGGcaspase-8(wt) in cleavage for the synthetic substrate. The remaining activity is somewhat surprising since FGGcaspase-8(D384A) has been demonstrated to be completely monomeric in buffered solution\[^{[11;[12]}\]. Addition of cucurbit[8]uril to FGGcaspase-8(wt) and FGGcaspase-8(D384A) solution, resulted in significant dose-dependent increase of the enzymatic activity. In the presence of 1 μM of cucurbit[8]uril, both FGGcaspase-8(wt) and FGGcaspase-8(D384A) featured the same activity of 1000 units/mg. Addition of more cucurbit[8]uril to FGGcaspase-8(wt) and FGGcaspase-8(D384A) did not change the activity of the enzymes, supporting the notion that the facilitated dimerization leads to a stable complex by the interaction of both supramolecular motif and the intrinsic protein dimerization. The activity of the two reference proteins, MGGcaspase-8(wt) and MGGcaspase-8(D384A), was the same as for FGGcaspase-8(wt) and FGGcaspase-8(D384A), respectively, in the absence of cucurbit[8]uril. Addition of cucurbit[8]uril to these reference protein solutions, however, did not lead to a change in activity. The monomeric FGGcaspase-8(F468A) did not feature any enzymatic activity on the synthetic substrate in the absence or presence of cucurbit[8]uril.
6.2.3 Catalytic activity of caspase-8 for natural substrate (casp-3)

In the apoptotic pathway, the initiator caspase-8 is responsible for cleavage and generation of casp-3/7 activation. Subsequently, activated casp-3/7 causes damage of DNA and apoptosis\(^1\)[23][24]. In order to test the cucurbit[8]uril induced control over caspase-8 activity, casp-3 was evaluated as a natural substrate. The different caspase-8 constructs were tested for their cleavage efficiency of casp-3 in the presence and absence of cucurbit[8]uril (Fig. 7 and Table 1). The FGGcaspase-8(wt) required approximately 25 minutes to cleave half of the casp-3 substrate in the absence of cucurbit[8]uril. Upon addition of cucurbit[8]uril to this solution, the FGGcaspase-8(wt) showed a significant increase of enzymatic activity, since it required only about 7 minutes for this complex to cleave half of the casp-3 substrate. The catalytic efficiency of cucurbit[8]uril induced FGGcaspase-8(wt) dimer was thus higher than that of FGGcaspase-8(wt) in the absence of cucurbit[8]uril. FGGcaspase-8(D384A), which showed a significant enzymatic activity for the synthetic substrate (Ac-IETD-AFC), was completely incompetent to cleave casp-3. Interestingly, addition of cucurbit[8]uril to FGGcaspase-8(D384A) reinstated enzymatic activity for the mutated caspase-8 against the natural substrate. The cucurbit[8]uril-induced FGGcaspase-8(D384A) dimer required only 10 min to cleave half of the casp-3 substrate, almost as active as the wild-type caspase-8 in the presence of cucurbit[8]uril. FGGcaspase-8(F468A) did not show any catalytic activity for the natural casp-3 substrate in the absence or presence of cucurbit[8]uril (Table 1).
Enhancement and reactivation of caspase-8 activity by cucurbit[8]uril induced dimerization

Figure 7. Cleavage activity of FGGcaspase-8(D384A) (0.15 μM) for casp-3 (4 μM) in the absence (A), and presence (B) of cucurbit[8]uril (1μM). Casp-3 fl (full length); Is (large subunit); ss (small subunit).

Table 1. Enzymatic activity of caspase-8 for the synthetic substrate (Ac-IETD-AFC) and the natural substrate (casp-3).

<table>
<thead>
<tr>
<th></th>
<th>Ac-IETD-AFC (Unit/mg)</th>
<th>Time required for 50% cleavage of casp-3 by 0.15 μM of caspase-8 (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-Q8</td>
<td>+Q8</td>
</tr>
<tr>
<td>MGGcaspase-8 (wt)</td>
<td>713</td>
<td>645</td>
</tr>
<tr>
<td>FGGcaspase-8 (wt)</td>
<td>606</td>
<td>1024</td>
</tr>
<tr>
<td>MGGcaspase-8 (D384A)</td>
<td>223</td>
<td>210</td>
</tr>
<tr>
<td>FGGcaspase-8 (D384A)</td>
<td>238</td>
<td>1000</td>
</tr>
<tr>
<td>FGGcaspase-8 (F468A)</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

Unit: One unit cleaves 1.0 nmole of substrate Ac-LEHD-AFC per minute at pH 6.5 at 37 °C.
Q8: Cucurbit[8]uril
n.d.: no detection
FE: Fold enhancement
6.2.4 Reversible inhibition of caspase-8 dimerization and activity

A supramolecular cucurbit[8]uril-induced protein dimerization approach provides a powerful tool to reversibly control protein dimerization and activation[13]. The increase of enzymatic activity via cucurbit[8]uril-induced caspase-8 dimerization should be brought back to the basal enzymatic activity via inhibition of complex formation upon addition of a competitor peptide. In order to establish evidence for this concept, the FGG peptide was used as a competitor peptide to inhibit the formation of cucurbit[8]uril-induced FGGcaspase-8(D384A) dimerization and activation for the synthetic substrate, Ac-IETD-AFC. Upon addition of the FGG peptide to a solution of cucurbit[8]uril-induced FGGcaspase-8(D384A) dimer, a decrease of the enzymatic activity of caspase-8 was observed (Fig. 8). At an excess of FGG peptide, the activity of the solution was decreased to the basal background level of monomeric FGGcaspase-8(D384A). These results show that the supramolecular induced caspase-8 dimer is a reversible complex, allowing for both induction of enzymatic activity and subsequent blockage by replacing the cucurbit[8]uril bound FGG-tag of FGGcaspase-8(D384A) with a competitor FGG peptide.

![Figure 8](image)

**Figure 8.** A) Schematic representation of reversible inhibition of the enzymatic activity of cucurbit[8]uril-induced FGGcaspase-8(D384A) dimerization via addition of FGG peptide. B) A complex of cucurbit[8]uril-induced FGGcaspase-8(D384A) [0.15 μM of FGGcaspase-8(D384A): 1 μM of cucurbit[8]uril] was titrated with increasing amounts of FGG peptide.
6.3 Discussion

Dimerization has been demonstrated as a crucial mechanism for caspase-8 activation\[^{1;21;26;27}\]. The absence of the central 10 amino acids of the intersubunit linker of caspase-8(wt) provides an advantage for the arrangement of two subunits in close proximity and stabilization of the active site loop, resulting in enzymatic activity of caspase-8(wt). The mutant caspase-8(F468A) exists completely as a monomer even at high concentration\[^{11}\], not showing catalytic activity. The incompetence for dimerization of caspase-8(F468A) could be explained by the absence of a cross-strand intermolecular interaction between phenylalanine and proline, thus resulting in the absence of self-proteolytic processing\[^{12}\]. This protein thus lacks enzymatic activity connected to its impaired dimerization capacities, supporting the notion for the need of enzyme active-site rearrangement upon dimerization\[^{11;12}\]. In contrast to caspase-8(F468A), caspase-8(D384A) surprisingly still is enzymatically active for the synthetic substrate, even though caspase-8(D384A) has been demonstrated to exist in monomeric form in buffered solution\[^{11;12}\]. The conflicting results are difficult to reason or further interrogate using the existing point-mutated caspase-8 analogs. Point mutations in the catalytic protein domain could lead to a plethora of biological effects, including loss of dimerization affinity and impaired active site rearrangement. Which of these processes is detrimental for the enzyme activity is difficult to establish, since they are typically coupled.

The complexity of the molecular process of caspase-8 activation is further illustrated by the activity of caspase-8(D384A) towards the synthetic Ac-IETD-AFC substrate, but inactivity to the natural caspase-3 substrate featuring the same substrate motif (Fig. 9). An orthogonal means to control caspase-8 dimerization, uncoupled from receptor rearrangement, via an allosteric, facilitated dimerization would be able to shed light on the molecular processes involved.

Cucurbit[8]uril-induced caspase-8 dimerization provides an important, novel biochemical tool to study dimerization active site rearrangement of caspase-8. Addition of cucurbit[8]uril to FGGcaspase-8(wt) and FGGcaspase-8(D384A) increases the enzymatic activity of both proteins to the same end-value caspase-8 for the synthetic substrate. The selective recognition of the FGG tag by cucurbit[8]uril is a crucial mechanism to facilitate the dimerization of both FGGcaspase-8(wt) and FGGcaspase-8(D384A)\[^{13;17}\]. The selectivity of cucurbit[8]uril in this process is exemplified by the absence of an effect on both MGGcaspase-8(wt) and MGGcaspase-8(D384A) . Interestingly, the maximum activity of the cucurbit[8]uril-induced FGGcaspase-8(D384A) dimer has the same level as that of cucurbit[8]uril-induced FGGcaspase-8(wt) dimer. This demonstrates the ability of cucurbit[8]uril not only to induce full dimerization for both proteins, but also to enable the rearrangement of the active site of caspase-8 for enzymatic activation. Overall, cucurbit[8]uril induces dimerization of the mutant caspase-8, which leads to full regeneration of its enzymatic activity. Addition of cucurbit[8]uril to FGGcaspase-8(F468A) did not induce any enzymatic activity, showing that
dimerization alone is not sufficient for caspase-8 activation. For the FGGcaspase-8(F468A), the point mutation not only affects the capacity for dimerization, but also hinders effective active-site rearrangement. The results support the hypothesis that the self-proteolysis process and the cross-strand intermolecular interaction are both critically important for activation of caspase-8[12].

Notable differences were observed for the cleavage of the different caspase-8 variants in the absence and presence of cucurbit[8]uril for the synthetic and the natural substrates. The catalytic activity of cucurbit[8]uril induced FGGcaspase-8(wt) for the natural substrate casp-3 is higher than that of FGGcaspase-8(wt) in the absence of cucurbit[8]uril. This notion is further strengthened by the surprising observation that monomeric FGGcaspase-8(D384A) is still substantially active towards the synthetic substrate but does not show any enzymatic activity for the natural substrate (Fig. 8A). These results can be reasoned considering the size and structure of the two different substrates. The small synthetic peptide substrate (Ac-IETD-AFC) might easily access the active site pocket of monomeric FGGcaspase-8(D384A). In contrast, the quaternary structure of casp-3 might not obtain access to the active site pocket of the monomeric FGGcaspase-8(D384A). These results strongly suggest differences in substrate processing by caspases, and potentially other enzymes, and provide an explanation for the previously observed, unexpected, catalytic activity of monomeric caspase-8(D384A)[11]. The caspase-8(D384A) still retains the 10 amino acids intersubunit linker on its small subunit that strongly affects dimer formation and activation of caspase-8(D384A) even at high caspase concentration[11].

FGGcaspase-8(D384A) alone is completely inactive against the caspase-3 substrate. Addition of cucurbit[8]uril induces the dimerization of the protein and remarkably imparts full enzymatic towards the natural substrate. The cucurbit[8]uril thus fully reinstates the enzyme activity of the point-mutated inactive monomer. Apparently, the cucurbit[8]uril induced dimerization of FGGcaspase-8(D384A) rearranges the loop of the active site which allows caspase-3 to access the active site pocket (Fig. 9). Within the dimerized form of caspase-8(D384A), the intersubunit linker is thus not hindering the access to the active site anymore. Overall these results show that there is a finely-balanced interplay between caspase-8 dimerization and activation and that outcomes of the activity studies depend strongly on the type of substrate under study. For caspase-8, it becomes clear that dimerization is fully required to achieve activity toward its natural substrate, and that this dimerization needs to occur via its natural recognition motif. Facilitation of the dimerization by cucurbit[8]uril has shown the need for a specific active-site rearrangement, since the cucurbit[8]uril induced dimer of caspase-8(F468A) does not feature any activity, whereas the cucurbit[8]uril induced dimer of caspase-8(D384A) does. Caspase-8 dimerization is therefore closely coupled and plays a key role in reorganization of the active site for enzymatic activation.
Finally, the selective recognition of the FGG tag of FGGcaspase-8(D384A) by cucurbit[8]uril can be inhibited by a competitor FGG peptide. This results in the restoration of the basal enzymatic activity of FGGcaspase-8(D384A). The cucurbit[8]uril-FGG system thus provides an entry to reversibly switch on and off caspase-8 dimerization and activation.

Figure 9. Enzymatic activity of monomeric FGGcaspase-8(D384A) and cucurbit[8]uril/FGGcaspase-8(D384A) complex for the synthetic substrate, Ac-IETD-AFC and natural substrate, caspase-3. A) Monomeric FGGcaspase-8(D384A) cleaves Ac-IETD-AFC, thereby releasing free AFC but does not cleave caspase-3. B) Cucurbit[8]uril-induced FGGcaspase-8(D384A) dimerization cleaves both Ac-IETD-AFC and caspase-3, thereby releasing free AFC and two fragments of caspase-3 (large and small subunits), respectively.

6.4 Conclusions

Cucurbit[8]uril selectively recognizes and induces dimerization of caspase-8 variants having an FGG tag at their N-terminus in an orthogonal, facilitated fashion. Supramolecular induced dimerization is a crucial mechanism to enhance the enzymatic activity of these caspases. Profound differences were observed for a number of mutant caspase-8 variants with respect to activity in monomeric versus dimerized state and regarding activity towards synthetic and natural substrate. Caspase-8(F468A) is inactive against any substrate both in monomeric form and in cucurbit[8]uril dimerized form. Caspase-8(D384A) in monomeric form cleaves the synthetic substrate, but is incompetent to cleave the natural caps-3 substrate. Cucurbit[8]uril-induced dimerization of caspase-8(D384A) fully reactivates the enzymatic
activity of the inactivated FGGcaspase-8(D384A) to the level of the wild-type caspase-8 for both the synthetic substrate and natural substrate. The results have shown that caspase-8 dimerization is tightly coupled with active site-rearrangement and that protein dimerization is the driving force for this active-site rearrangement.

Classical protein engineering approaches via introduction of point mutations leads to caspase proteins that can only be studied in a single state, e.g. monomeric\textsuperscript{[11]}, or dimeric\textsuperscript{[19]}. The application of cucurbit[8]uril induced dimerization allows the study of these caspases both in monomeric and dimerized state, leading to novel insights in the caspase-8 activation mechanism as shown above. Control over enzyme dimerization by cucurbit[8]uril and a very small genetically encoded FGG peptide motif provides a powerful approach to study molecular mechanism of enzyme dimerization and activation. It can be envisioned that this concept can also be applied to other relevant protein dimerization and activation events in biologically relevant processes\textsuperscript{[28-30]}.

6.5 Experimental section
6.5.1 Construction of plasmids

The human caspase-8 (amino acids 217-479) having an FGG motif at the N-terminus was generated by using the N-terminal intein from plasmid pTWIN1 (Fig. 10) (New England Biolabs). DNA encoding for FGGcaspase-8 was amplified by PCR using caspase-8 gene (Invitrogen) as a template and forward primer ON143, 5'- atc att gta cac aac ttt ggt gcc ggt tgt gac aaa gtt tac caa atg aaa ag -3' and reverse primer ON 144, 5'- gtt agc agc cgg atc tca gtt gtt gtt gtt gtg gtg gtg gtg gtg gtg cag -3'. This PCR product was cloned into SapI and BamHI treated-pTWIN1 using in-fusion enzyme (Clontech), resulting in the FGGcaspase-8 plasmid. The schematic representation for the construction of FGGcaspase-8 is shown in Figure 9. The plasmid aspase-8 having an MGG motif at the N-terminus (MGGcaspase-8) was generated by using the same intein system method with forward primer ON145, 5'- atc att gta cac aac atg ggt gcc ggt tgt gac aaa gtt tac caa atg aaa ag -3' and reverse primer, ON144, 5'- gtt agc agc cgg atc tca gtt gtt gtt gtt gtt gtt gtg atc ag -3'.

The engineered monomeric variants were generated by replacing aspartic acid with alanine at position of 384, resulting in caspase-8(D384A) or replacing phenylalanine with alanine at position of 486, resulting in caspase-8(F486A). The substitution of the amino acid codon was achieved by using Quikchange mutagenesis (Stratagene). FGGcaspase-8(D384A) and MGGcaspase-8(D384A) were generated by using the templates of FGG-caspase8 and MGGcaspase-8, respectively with a pair of primers ON150: agc aac cct att tag aaa tgg ctt tat cat cac ctc aaa cga g and ON151: ctc gtt tga ggt gat gat aaa gcc att tct aaa tag ggt tgt tgc. FGGcaspase-8(F486A) was generated by using the templates of FGGcaspase-8 with a pair of primers ON152: gaa aca gat gcc tca gcc tac tgc cac act aag aaa aaa act tgt c and ON153: gac
aag ttt tct tag tgt ggc agt agg ctg agg cat ctg ttt c. The amino acid sequences of the resulting proteins are shown below.

**FGGcaspase-8(wt)**

\[
\text{FGGGLDKVYQMKSKPRGCYLIIINHNFAKAREKVPKLHSIRDRNTHLDAGALTTFEE} \\
\text{LHFEIKPHDCTVQIEILKIYQLMHDHSNMDCFCCILSHGDKGIYGTGDQEAPIYETSLQ} \\
\text{FTGLKCPSLAGKPKVFFIQCQGDNQYKGPETDSEEQPYLEMLSSPQTRYIPDEADFLL} \\
\text{GMATVNNCYSYRNPAGETYWYQSLQSLRERCPRGDDILTLTEVNYEVSNKDDKNNMG} \\
\text{KQMPQPTFTLRKKLVFPSDHHHHHH}
\]

**MGGcaspase-8(wt)**

\[
\text{MGGGLDKVYQMKSKPRGCYLIIINHNFAKAREKVPKLHSIRDRNTHLDAGALTTFEE} \\
\text{LHFEIKPHDCTVQIEILKIYQLMHDHSNMDCFCCILSHGDKGIYGTGDQEAPIYETSLQ} \\
\text{QFTGLKCPSLAGKPKVFFIQCQGDNQYKGPETDSEEQPYLEMLSSPQTRYIPDEADFLL} \\
\text{LGATVNNCYSYRNPAGETYWYQSLQSLRERCPRGDDILTLTEVNYEVSNKDDKNNMG} \\
\text{KQMPQPTFTLRKKLVFPSDHHHHHH}
\]

Figure 10. Schematic construction of the plasmid for FGGcaspase-8 expression
FGGcaspase-8(D384A)

FGGGLDKVYQMKSKPRGYCLIINNHNFAKAREKVPKLHSIRDRNGTHLDAGALTTFEE LHFEIKPHDDCTVEQIYEILKIYQLMDHSNMDCFICCILSHGDKGIIYGTGDQEAPIYELTSQ FTGLKCPSLAGPKVFFIQACQGDNYQKGIPVETDSEEQPYLEMALSSPQTRYIPDEADFLL GMATVNNCVSYRNPÆGETWYIQSLCQLRSRCPRGDDILTILTEVNYEVSNKDDKKNMG KQMPQPTFTLRKKLVFPSDHHHHHH

MGGcaspase-8(D384A)

MGGGLDKVYQMKSKPRGYCLIINNHNFAKAREKVPKLHSIRDRNGTHLDAGALTTFEE ELHFEIKPHDDCTVEQIYEILKIYQLMDHSNMDCFICCILSHGDKGIIYGTGDQEAPIYELTS QFTGLKCPSLAGPKVFFIQACQGDNYQKGIPVETDSEEQPYLEMALSSPQTRYIPDEADFLL LGMATVNNCVSYRNPÆGETWYIQSLCQLRSRCPRGDDILTILTEVNYEVSNKDDKKNMG KQMPQPTFTLRKKLVFPSDHHHHHH

FGGcapase-8(F468A)

FGGGLDKVYQMKSKPRGYCLIINNHNFAKAREKVPKLHSIRDRNGTHLDAGALTTFEE LHFEIKPHDDCTVEQIYEILKIYQLMDHSNMDCFICCILSHGDKGIIYGTGDQEAPIYELTSQ FTGLKCPSLAGPKVFFIQACQGDNYQKGIPVETDSEEQPYLEMALSSPQTRYIPDEADFLL GMATVNNCVSYRNPÆGETWYIQSLCQLRSRCPRGDDILTILTEVNYEVSNKDDKKNMG KQMPQPTATLRKKLVFPSDHHHHHH

6.5.2 Protein expression and purification

The plasmids were transformed into E. coli strain BL21 (DE3). The bacteria were cultured in LB medium containing 100 µg/ml of ampicillin and the cells were grown at 37 °C, shaking 250 rpm to an A600 of 0.7, then IPTG was added to a final concentration of 0.3 mM. The cells were continuously incubated overnight at 15 °C, shaking 250 rpm before being harvested. The pellet was resuspended into the bugBuster protein extraction reagent plus benzonase nuclease (to degrade DNA and RNA) and the insoluble material was removed by centrifugation at 20000 rpm for 40 minutes at 4 °C. The soluble fraction was applied to a column filled with chitin beads (New England Biolabs) through gravity flow, and the column was washed with 40 volumes of sodium phosphate buffer (20 mM, 100 mM of sodium chloride, pH 7) and subsequently incubated about 6 hours at room temperature (Fig. 11). The cleaved proteins were then collected in the flow-through using phosphate buffer. The target proteins were analyzed by SDS-PAGE and LC-ESI-MS (Fig. 11B and Table 2).
**Figure 11.** A) Schematic representation of generation of FGGcaspase-8 by intein splicing. B) SDS-page of caspase-8: 1) FGGcaspase-8(wt) consisting of a large (17.7 kDa) and a small (11.7 kDa) subunits. 2) MGGcaspase-8(wt) consisting of a large (17.7 kDa) and a small (11.7 kDa) subunits; 3) FGGcaspase-8(D384A) consisting of a large (17.7 kDa) and small (12.9 kDa, including 10 central amino acids, S375-A384) subunits; 4) MGGcaspase-8(D384A) consisting of a large (17.7 kDa) and small (12.9 kDa, including 10 central amino acids, S375-A384) subunits. 5) FGGcaspase-8(F468A) is uncleaved consisting of one unit of caspase-8 (30.5 kDa).

**6.5.3 LC-ESI-MS analysis of proteins**

The LC-ESI-MS experiments were performed using a LCQ Fleet (Thermo Scientific). Proteins went through the C4 column under elution with a water/acetonitrile gradient (5% water: 95% acetonitril for 1 minute to 70% water: 30% acetonitril in 9 minutes and stay at 70% water: 30% acetonitril for 2 minutes) with 0.1 % formic acid before they were introduced to the ion source and mass spectrometry. The mass of the proteins was deconvoluted by Magtran software (Table 2)
### Table 2. Calculated and measured mass of proteins

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Large subunit</th>
<th>Small subunit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calculated</td>
<td>Measured</td>
</tr>
<tr>
<td></td>
<td>mass (Da)</td>
<td>mass (Da)</td>
</tr>
<tr>
<td>FGGcaspase-8</td>
<td>17688</td>
<td>17687 -1</td>
</tr>
<tr>
<td>(wt)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGGcaspase-8</td>
<td>17672</td>
<td>17670 -2</td>
</tr>
<tr>
<td>(wt)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGGcaspase-8</td>
<td>17688</td>
<td>17687 -1</td>
</tr>
<tr>
<td>D384A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGGcaspase-8</td>
<td>17672</td>
<td>17670 -2</td>
</tr>
<tr>
<td>D384A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGGcaspase-8</td>
<td>30528</td>
<td>30528 0</td>
</tr>
<tr>
<td>F468A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**6.5.4 Activity assay for caspase-8 on synthetic substrate**

The assay was based on the hydrolysis of the peptide substrate Ac-IETD-AFC (Acetyl-ILe-Glu-Thr-Asp-7-amino-4-trifluoromethyl coumarin,) by caspase-8 resulting in release of an AFC (7-amino-4-trifluoromethyl coumarin) moiety. The product AFC was detected by using excitation wavelength, 400 nm and emission wavelength, 500 nm.

The activity assays were carried out at 0.15 μM of caspase-8 in 100 μL of assay buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl₂ and 2 mM TCEP, pH 6.5) with 100 μM of Ac-IETD-AFC substrate. Kinetic caspase-8 activity measurements were performed in duplicates using quartz cuvettes of 0.5 cm pathlength and measured by Cary Eclipse fluorescence spectrophotometer (Varian) at 37 °C. Specific activity of caspase-8 was then calculated by using an AFC calibration curve. Titration of cucurbit[8]uril to caspase-8 was performed under similar conditions, where the concentration of cucurbit[8]uril was varied from 0 μM to 4 μM.

The AFC calibration curve was prepared via an AFC dilution series with assay buffer in the range of concentration 0.08 μM to 10 μM (Fig. 12).
Enhancement and reactivation of caspase-8 activity by cucurbit[8]uril induced dimerization

6.5.5 Caspase-8 assay on natural substrate

The experiments were carried out in the mixtures of 0.15 μM caspase-8 and 4 μM substrate (casp-3) in the absence or presence of 4 μM cucurbit[8]uril in assay buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl₂, 2 mM TCEP, pH 6.5) Reactions were incubated at 37 °C. Samples were collected at time intervals and reactions were stopped by addition of SDS-loading buffer. All samples were separated on 8 to 18% SDS-PAGE and proteins detected with blue staining (Fig. 13). The extent of casp-3 (C163A) cleavage was determined by densitometry using an AlphaEaseFC.

Figure 12. AFC calibration curve

Figure 13. Cleavage activity of FGGcaspase-8(wt) (0.15 μM) for casp-3 (4 μM) in the absence (A), and presence (B) of cucurbit[8]uril (1μM). Casp-3 fl (full length); ls (large subunit); ss (small subunit).
6.6 References

Chapter 7
Exploring cucurbit[8]uril-induced protein dimerization in living cells

Abstract

Control over membrane protein dimerization using a host-guest system holds a great promise for the regulation of many cellular events. The recognition of the FGG peptide motif by cucurbit[8]uril would provide a means to induce membrane protein dimerization in living cells. Localization of the membrane protein EGFR (epidermal growth factor receptor), containing a C-terminal green fluorescent protein (GFP) domain, was monitored by fluorescence microscopy. Replacement of the native leucine-glutamic acid-glutamic acid (LEE) motif, located behind the cleavage signal peptide sequence, with the phenylalanine-glycine-glycine (FGG) motif could allow the generation of EGFP-GFP bearing the free N-terminal FGG motif. GFP internalization is observed by adding EGF ligand to FGG-EGFR-GFP under wide-field microscopy and confocal microscopy. Similar EGFR features were observed after the addition of cucurbit[8]uril to the FGG-EGFR-GFP protein when observed with wide-field microscopy but not when observed by confocal microscopy. The exact molecular mechanism of cucurbit[8]uril-FGG-EGFR-GFP recognition and subsequent binding requires the generation of model proteins. Model proteins bearing an N-terminal FGG motif liberated by caspase-9 cleavage were generated and showed effective liberation of the desired N-terminal motif, thus posing a promising system for follow-up studies on membrane protein dimerization in live cells.
7.1 Introduction

Dimerization of membrane proteins is a crucial mechanism for the activation of these proteins in many cellular events such as cell growth, differentiation, migration and apoptosis\(^{[1-7]}\). A highly pronounced role of membrane protein dimerization comes to play in the regulation of signal transduction\(^{[5,6,8-10]}\). A large number of protein surface receptors dimerize in the membrane upon signal binding, subsequently leading to effective intracellular signal transduction. The dimerization of the surface receptor proteins is required for the activation of the intracellular kinase activity of the membrane protein itself or of protein kinases associated with the transmembrane protein\(^{[1,3,6,8]}\). These transmembrane proteins typically consist of a helical transmembrane domain between an extracellular domain and a cytoplasmic domain. The extracellular domain is responsible for the recognition of an extracellular ligand, which leads to changes in protein conformation. The conformation changes result in an enhancement of the proximity of the extracellular domains of two or more proteins, so-called ligand-induced protein dimerization, followed by the interactions of the intracellular domains of these proteins.

Protein tyrosine kinase receptors include the platelet-derived growth factor receptor (PDGFR) family, the epidermal growth factor receptor (EGFR) family, and the fibroblast growth factor receptor (FGFR) family which are activated upon growth factor binding\(^{[1,4,11-13]}\). The ligand-induced dimerization brings two intracellular kinase domains in close proximity to allow multiple phosphorylation events to occur in which one receptor of the dimer cross-phosphorylates the other. Phosphorylation on a tyrosine residue inside the kinase domains leads to increased activity of the kinase and precedes phosphorylation of other sites in the receptor. Phosphorylation outside the kinase domains serves the important function as docking sites for downstream signal transduction molecules featuring SH2 domains\(^{[14]}\). The ligand-induced membrane protein dimerization can result from different types of molecular mechanisms. Several of the extracellular ligands are themselves assembled into dimeric form and thus contain two sites for receptor binding. For example, PDGF and CSF-1 are disulfide bonded dimers which bind simultaneously two receptors, resulting in stable receptor dimers\(^{[15,16]}\). A monomeric hGH which possesses bifunctional surfaces can bind and dimerize two receptor molecules, thereby forming a complex of 1 ligand:2 receptors\(^{[17]}\). Two ligand molecules can also bind to two receptor molecules, forming a 2:2 (receptor:ligand) complex. For example, the crystal structure of the EGF-EGFR complex (only the extracellular domain) shows EGFR dimerization requires the binding of two EGF ligands to two EGFR receptors in a 2EGF:2EGFR complex\(^{[18]}\).

Control over membrane protein dimerization using low molecular weight organic molecules instead of peptide or protein ligands is of high interest for the design of new approaches to manipulate and regulate signal transduction pathways in biological processes. One class of such molecules are known as chemical inducers of dimerization such as
rapamycin, coumermycin, FK1012 (a semisynthetic dimer of FK506) and [(CsA)_2] (a semisynthetic dimer of cyclosporine). The chemical featuring a bifunctional interface which can selectively bind and bring two specific proteins together to form a dimeric complex\[^{19,20}\]. These small molecules have for example been used to regulate cell surface receptors that lacked their transmembrane and extracellular domains, showing that dimerization of the intracellular domains alone can lead to signal transduction\[^{21-23}\]. This general approach has been demonstrated using fusion proteins of a kinase domain with a protein domain binding molecular inducers of dimerization. For example, a membrane-associated fusion protein featuring an FKBP domain can recruit a protein kinase of interest fused to an FRB domain in the presence of rapamycin. The protein kinase is recruited to the plasma membrane by a small molecule mediated FKBP-rapamycin-FRB interaction, leading to activation of the kinase and resulting in signal transduction\[^{24,25}\]. The synthetic small molecules (CsA)_2 and FK1012 have been shown to be inducers for the homodimerization of cyclophiline and FKBP respectively, allowing to control programmed cell death. The Fas cytoplasmic domain is fused with cyclophiline (or FKBP) and a myristoyl membrane anchor. The addition of the molecular inducer of dimerization mediates the dimerization of the Fas receptor leading to activation of Fas signal transduction, resulting in cell death (Fig. 1)\[^{22,24}\]. These examples demonstrate that chemical inducers of dimerization provide a powerful tool to induce membrane protein dimerization and control biological responses in signal transduction pathways. However, there are still many open questions related to membrane protein dimerization and activation, especially regarding the interplay of the extracellular and intracellular domains. An orthogonal approach for the controlled dimerization of membrane proteins would possibly address this need. Here a supramolecular-induced membrane protein dimerization system has been explored.

![Figure 1. Chemical induced membrane protein dimerization. A) (CsA)_2 induces dimerization of a Fas-cyclophilin receptor. Cyclophilin is fused between a myristoyl anchor and a Fas domain, the presence of (CsA)_2 induces the dimerization of the Fas-cyclophilin receptor, leading to activation of the Fas domain, resulting in cell death; B) FK1012 induced dimerization of the Fas-FKBP receptor.](image-url)
Supramolecular host-guest chemistry has recently emerged as a versatile tool to reversibly control protein dimerization \[26-29\]. The specific recognition and dimerization of naphthalene and methylviologen by cucurbit[8]uril allows for inducing dimerization of protein-conjugated with these two guest elements\[28\]. Alternatively, proteins consisting of a very small genetically encoded phenylalanine-glycine-glycine (FGG) peptide motif at the N-terminus allows cucurbit[8]uril to selectively recognize and induce their dimerization\[26\]. The cucurbit[8]uril binds the FGG motif via a combination of hydrophobic interactions of the phenyl groups inside the cucurbit[8]uril cavity and interactions of the protonated N-terminal amine functionality with the carbonyl rim of the cucurbit[8]uril\[26;30\]. In this chapter, we envisaged to use the cucurbit[8]uril-FGG approach to induce membrane protein dimerization. We firstly generated fluorescent EGFR membrane proteins featuring an N-terminal FGG motif (Fig. 2). The cucurbit[8]uril-induced FGG-EGFR-GFP dimerization was preliminary evaluated with fluorescence microscopy. Subsequently, model proteins, featuring a specific protease sensitive tag were generated for the controlled liberation of an N-terminal FGG motif in membrane proteins. These studies led to the formulation of a novel model membrane protein to test cucurbit[8]uril induced protein dimerization.

![Figure 2](image.png)

*Figure 2. Schematic representation of the concept of cucurbit[8]uril induced FGG-EGFR-GFP dimerization in living cells.*

### 7.2 Results and discussion

A protein requires the action of two distinct elements to end up in the plasma membrane after protein expression: an N-terminal signal sequence motif and a transmembrane stop-transfer sequence. The signal peptide, which typically consists of hydrophobic amino acids, is cleaved by a signal peptidase during transportation of the protein sequence through the
endoplasmic reticulum (ER) membrane\textsuperscript{31}. The helical transmembrane sequence is blocked on the ER membrane, and thus the N-terminal fragment is located inside ER and the C-terminal fragment is located in the cytosol. The protein element inserted in the ER membrane is then exposed on the cell surface of the plasma membrane via transportation through the Golgi apparatus. The N-terminus of this membrane protein is thus finally present at the extracellular side of the plasma membrane and the C-terminus at the intracellular side of the plasma membrane\textsuperscript{32}.

7.2.1 Cucurbit[8]uril-induced FGG-EGFR-GFP dimerization

Epidermal growth factor receptor (EGFR) is a member of the receptor tyrosine kinase family and is connected to many human cancers\textsuperscript{11;12;33}. In the cellular system, growth factor binding to the extracellular domain of EGFR leads to dimerization of EGFR and activates the tyrosine kinase of its intracellular domain. The structural mechanism of the growth factor ligand-induced EGFR dimerization is well established\textsuperscript{18;34}. The ligands such as EGF (epidermal growth factor) and TGF (transforming growth factor) contact two of the four subdomains (domains I and III) of EGFR at the extracellular side, resulting in a complex of 2ligands:2EGFRs (Fig. 3). The binding of the growth factor ligand to EGFR changes and stabilizes the conformation of EGFR, which is required for dimerization.

\textbf{Figure 3. EGF-induced EGFR dimerization. A) the tethered EGFR conformation with 4 domains I, II, III and IV. B) EGF (cyan) contacts domains I and III which leads to the change in conformation of domain II, resulting in stabilization of the complex (2ligands:2EGFRs). C) Model for EGF (E) induced EGFR dimerization and activation. Image reproduced from reference\textsuperscript{18}.}
In this study, the specific recognition and dimerization of two FGG peptide motifs by cucurbit[8]uril was explored to control EGFR dimerization in living cells. The generation of EGFR constructs bearing an N-terminal FGG motif was based on its insertion after the signal peptide cleavage site. The EGFR signal peptide consists of 24 amino acids (MRPSGTAGAALLALLAALCPASRA) which targets EGFR for translocation across the membrane of ER. Many methods are available for predicting signal peptides from membrane proteins and their cleavage site. The current best signal peptide predictor, SignalP 4.0\textsuperscript{[35]}, was used to predict the cleavage site between the signal peptide and processed native EGFR. The result predicted the cleavage site to be between positions 24 and 25 (SRA↓LEE) with the highest C-score (0.562) (the predicted cleavage site value) (Fig. 4 left). This may provide EGFR bearing an N-terminal leucine-glutamic acid-glutamic acid (LEE) motif. Triple replacement of this LEE motif with phenylalanine-glycine-glycine (FGG) resulted in a mutant EGFR bearing an FGG motif directly behind the cleavage signal peptide site (Fig 5A). The prediction for this mutant EGFR using SignalP 4.0 showed that the highest C-score value is 0.413 for the cleavage site between position 24 and 25 with the local sequence context SRA↓FGG. Based on these calculations, the designed constructs was created as such, to provide the highest possibility to generate an EGFR protein bearing an N-terminal FGG motif at the outside of the cell (Fig. 4 right, Fig. 5B).

![Graph](image1)

**Figure 4.** Prediction from the SignalP 4.0 for the sequence wild type EGFR (left) and mutant EGFR (right). “C-score” is the predicted cleavage site value. Both the wild type and mutant EGFRs showed the predicted cleavage site between position 24 and 25, resulting in EGFR bearing an N-terminal LEE or FGG motif, respectively.
Exploring cucurbit[8]uril-induced protein dimerization in living cells

Figure 5. A) Mutation of three amino acids LEE to FGG. B) The signal peptidase may recognize and cleave the signal peptide of EGFR, resulting in generation of wild type LEE-EGFR [35] and mutant FGG-EGFR.

EGFR internalization by EGF-induced by EGFR dimerization is an element in EGFR kinase activation. The EGFR internalization can be elucidated via the localization of chimeric receptors bearing a C-terminal green fluorescent protein (GFP). The activation of EGFR-GFP by EGF results in redistribution of GFP via cellular internalization/aggregation[36,37] (Fig. 6)

Figure 6. Spatial and temporal dynamics of wild type EGFR-GFP are observed under the confocal microscopy. EGF (100 ng/ml) induced EGFR-GFP internalization. The long arrow indicates EGFR-GFP located initially at the cell plasma membrane (HEK293). The short arrows indicate redistribution of GFP via cellular internalization. Image reproduced from reference[37].

To determine whether cucurbit[8]uril can induce FGG-EGFR dimerization and internalization, an FGG-EGFR-GFP construct was generated based on the EGFR-GFP positive control[38]. EGF and cucurbit[8]uril-induced EGFR dimerization was studied using wide-field fluorescence microscopy. The wild type and mutant EGFR-GFP (Fig. 7A, B, C, D) were strongly expressed in Hela cells under the standard cell culture medium conditions. Addition of EGF to EGFR-GFP (wt) and the FGG-EGFR-GFP led in both cases to EGF-induced redistribution of the EGFR constructs, possibly via cellular internalization (Fig. 7E, F). This result indicates that the FGG-EGFR-GFP is correctly processed, like its natural counterpart, and is still active/responsive to EGF. The FGG mutation close to the signal
peptide sequence might not therefore effect processing and activity. Interestingly, after addition of cucurbit[8]uril to FGG-EGFR-GFP, redistribution was similarly observed for some cells after 1 hour (Fig. 7G). Addition of the same amount of cucurbit[8]uril to the native EGFR-GFP (reference) did not change its distribution (Fig. 7H).

Figure 7. EGF and cucurbit[8]uril-induced FGG-EGFR-GFP internalization observed by widefield fluorescence microscopy. A, B, C, D) Images of FGG-EGFR-GFP and wild type EGFR-GFP (wt) expressed in Hela cells after 48 hours. E, F) Upon addition of EGF (100 ng/ml) to EGFR-GFP (wt) and FGG-EGFR-GFP redistribution into particles was observed. G) Addition of cucurbit[8]uril (5 μM) to FGG-EGFR-GFP also resulted in redistribution into particles. F) Addition of cucurbit[8]uril to EGFR-GFP did not lead to a change in redistribution. Long arrows indicate staining at the plasma membrane, and short arrows indicate possible regions of intracellular internalization.

EGF and cucurbit[8]uril-induced EGFR redistribution were also studied using confocal microscopy in order to determine the exact localization of the protein constructs within the cell and after stimulation. Results showed that both EGFR-GFP (wt) and FGG-EGFR-GFP located at the plasma membrane. This demonstrates again that the mutation to FGG-EGFR does not affect the protein synthesis and final cellular processing and localization of the EGFR. Upon addition of EGF to EGFR-GFP, cellular internalization is observed, typical for EGF-induced EGFR-GFP dimerization [37,39]. Unfortunately, addition of EGF or cucurbit[8]uril to FGG-EGFR-GFP resulted in unclear results. EGF-stimulated FGG-EGFR-GFP possibly still resulted in a certain degree of cellular internalization (Fig. 8E). Cucurbit[8]uril-induced FGG-EGFR-GFP hardly showed any internalization under these conditions (Fig. 8F). The results obtained for cucurbit[8]uril-induced FGG-EGFR-GFP dimerization as observed using wide-field and confocal fluorescence microscopy are thus conflicting to a certain extent. The
protein expresses well and also localizes nicely in the plasma membrane, but cellular internalization cannot be established. This remains a highly interesting topic to further elucidate in more detail, as stimulation of FGG-EGFR-GFP might induce dimerization but possibly not result in protein internalization, which implies that these two processes are not directly coupled. The different observations of cucurbit[8]uril-induced FGG-EGFR-GFP dimerization between the widefield and confocal microscopy needs to be further investigated. It might depend on experimental conditions such as tranfection reagent or living cell behavior. Therefore, optimal conditions need to be established to provide a constant result.

Figure 8. Confocal microscopy studies of EGF and cucurbit[8]uril-induced EGFR-GFP redistribution A, B, C) FGG-EGFR-GFP and EGFR-GFP (wt) expressed in Hela cells after 36 hours. D) Addition of EGF (100 ng/ml) to EGFR-GFP (wt) resulted in redistribution of EGFR-GFP via cellular internalization. E) Addition of EGF (100 ng/ml) to FGG-EGFR-GFP potentially showed redistribution of EGFR-GFP. F) Addition of cucurbit[8]uril (740 nM) to FGG-EGFR-GFP resulted in only very minor changes of EGFR-GFP distribution. Long arrows indicate staining at the plasma membrane, and short arrows indicate regions of intracellular internalization.

The use of cucurbit[8]uril as an inducer of EGFR dimerization and activation is confronted with a number of scientific challenges. Cleavage of the signal peptide sequence is still a difficult to predict feature and limits exact molecular control over the N-terminal protein composition\[^{38}\]. The extracellular domain of EGFR features a complicated 4 subdomain organisation, whose EGF-induced dimerization mechanism, including a protein reorganization, might differ substantially from the facilitated dimerization induced by cucurbit[8]uril\[^{18}\]. Whether EGFR dimer formation drives the internalization or activation is
still not fully understood\textsuperscript{[40]}. In order to obtain more molecular control over cucurbit[8]uril-induced protein dimerization in the membrane, fluorescent protein models will be studied first, because of their easier manipulation and observation.

7.2.2 Generation of proteins bearing an N-terminal FGG motif via caspase mediated cleavage

Cucurbit[8]uril mediated membrane protein dimerization strictly requires the proteins to bear an N-terminal FGG motif. Due to the limited knowledge and molecular control over the cleavage of the signal peptide, an orthogonal approach capable of generating N-terminal FGG motifs was explored. The enzyme caspase-9 selectively recognizes and cleaves peptides and proteins after the aspartic acid (D) residue of an LEHD peptide sequence\textsuperscript{[41,42]}. However, whether caspase-9 can specifically cleave a LEHDFGG motif to generate proteins bearing an N-terminal FGG motif has not yet been explored. Therefore, a YFP protein featuring an LEHD-FGG sequence was generated as a model protein to explore the substrate tolerance of the caspase-9 enzyme. The YFP protein was incubated with caspase-9 at 37 °C in buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl\textsubscript{2} and 2 mM TCEP, pH 6.5). LC-MS analysis of the resulting protein mixture revealed that caspase-9 indeed can selectively recognize and cleave this protein after the LEHD sequence, leading to the generation of YFP bearing an N-terminal FGG motif (Fig. 9).

![Figure 9. Deconvoluted mass spectra of LC-ESI-MS runs of LEHD-FGG-YFP (left) and caspase-9-cleaved LEHD-FGG-YFP (right). Caspase-9 selectively recognizes and cleaves LEHD-FGG-YFP, resulting in the generation of FGG-YFP in buffer (25 mM HEPES, 100 mM NaCl, 4 mM MgCl\textsubscript{2} and 2 mM TCEP, pH 6.5). Reaction conditions: 20 μM LEDH-FGG-YFP, caspase-9 (50U), incubated at 37 °C for 10 hous.](image-url)
7.3 Conclusions and outlook

The cucurbit[8]uril-induced dimerization system holds great promise in controlling membrane protein dimerization in living cells. EGFR bearing an N-terminal FGG motif, by cleavage signal peptide engineering, could successfully be expressed in the cellular membrane. Stimulation of this protein construct with EGF or cucurbit[8]uril led to interesting observations, which require additional studies to elucidate the interplay of EGFR dimerization and internalization. The usage of caspase-9 for the cleavage of LEHDFGG sequence was successfully shown and provides a useful new entry for the generation of proteins bearing an N-terminal FGG motif on the surface of live cells.

In order to further explore the potential of cucurbit[8]uril-induced membrane protein dimerization, a novel set of membrane surface fluorescent proteins is proposed. Both EGFP and mCherry bearing an N-terminal LEHD-FGG motif could be generated to be expressed on the cell surface by fusing these two proteins to a transmembrane protein domain (from pDisplay Tm vector, Invitrogen). The N-terminal FGG motifs will be set free after incubation of the cells with caspase-9 (Fig, 10). The caspase will typically not enter the cell and will only cleave after the specific LEHD on the outside of the cell. Liberation of the FGG-motifs allows cucurbit[8]uril to induce random heterodimerization (and homodimerization) of the EGFP with the mCherry protein, which can be elucidated by energy transfer studies between the donor EGFP and the acceptor mCherry. Hopefully, this novel concept of cucurbit[8]uril-induced membrane protein dimerization will subsequently open up the possibility to regulate signal transduction of membrane proteins in biological processes.

![Schematic representation of transmembrane EGFP bearing an extracellular protected N-terminal FGG motif, and the liberation of the FGG via caspase-9 mediated cleave on the cell surface.](image)

![Schematic representation of cucurbit[8]uril-induced random heterodimerization of FGG-tagged EGFP and mCherry on the cell surface.](image)
Chapter 7

7.4 Experimental section

7.4.1 Construction of plasmids

Construction of FGG-EGFR-GFP

DNA encoding for FGG-EGFR-GFP was generated from an EGFR-GFP template[38] by using the QuickChange Site-Directed Mutagenesis kit (Qiagen) to replace the three amino acids LEE, behind the cleavage signal peptide site, with FGG.

Amino acid sequence of EGFR-GFP

- Underlined sequence: Signal peptide
- Bold sequence: extracellular domain (I, II, III, and IV)
- Bold and underlined sequence: transmembrane region
-Italic sequence: intracellular domain
-Italic and underlined sequence: green fluorescent protein (GFP)

MRPSGTAGAAALLAALCPASRALLEEKVCQGTSNKLTQLTQLGTFEDHFHLSSLQRMFNN
CEVVLGNLITYVQRNYDLSFLKTIQEVAGYVLIALNTVERPLENQLIIRGNNMYENS
ALAVLSYANKTGLKELPMRLQELHGAVRFSNPNALCNESEIQWRDVISDFLSN
MSMDFQNHLSCQKCDPSCPNGSCWGAGEENCQLKTIIAQAQCSCGRGKSPSDC
CHNQCAAGCTGPRSDCLVRCKFRDEATCKDTCPPLMLYNPTTMYQDMEGKYSF
GATCVKKCPMRNYVVDHTHSCVRACGADSYEMEEDGVKRCKCCEGPRCRKVCNGIGG
EKDFSLINATNKHFKNCSTISGDLHIPARQGDFSFTHTHPDQLDLKTVKETIGF
LIQAWPENRTDLHAFENLEIIIRGRTKQHQFQSLAVVSLNITSGLRSLKSEIDGDIIS
GNKNLICYANTINWKFLFGTSGKQTIIKNRGENSCKATQVCVALCSPEGCGWPEPR
DCVSCRVNVRSGRECVDKCNLLEGEFQVSEICHQPELQPAMNITCGRPDC
IQCAHYIDPHCVKTCPAGVMENNTLVKWYADAGHVCHELCHPNCTYGTCPGGE
GCPNTGIPKISSLATGMVGAALLLWVAALGIGLFRMRRRHIVKNTRRLRLQERELVEPLTSG
EAPNQALLRLIKETEFKKIKVLGSGAGFTVYKGLWPEGEFKVIPVAIKELREATSPKANKEILD
AYVMASVDNPHVCRLLGICLSTVQLTQLMPFGCLLDYVREHDNIGSQQYLLNWCQIAKG
MNYLEDRRLVHRDLAARVNLVKTPQHVITDGLAKLGAEEKEHYAEGKVPIKMALESI
LHRIYTHQSDVWSYGVTVWEMFTGSKPYDGIAPASEISSLEKGERLQPQPPCTIDVYMIMVKC
WMIDASRPKRELIEFSKMARDPQRYLVIQGDERMLPSPTDNSYRALMDDEMDDDVV
DADEYLIPQQGFSSPSRTPLLLLSSLATSNNSTVACIDRNLQSCPKEDSFLQRYSSDPTGA
LTEDSIDDTLFPVEYNQSPKVPSQVNPVYHNQPNLPSAPRDPHYQPDHSTAVGNEYPY
LNTVQPTCVNSTFSDFPAWHAQKGHSIQSLDNPDYQQDFPPKEAKPNGFLGKSTAEANAYLRV
APQSCEFIGAAIGGQGGMVSKGEELETGVVPIYELDGDVNGHFEVSGEGEDDATYKLT
LKFICTTGGKLVPWPTLVTTLTYGVQCFSRYPDHMKQHDFKSAHMPVEVQERTIFFKDDGN
YKTRAEVKFEGDLTNRIELKGIDFEDGNLGHKLEYNSSHNVYIMADKQKNGIKVNFKIR
Exploring cucurbit[8]uril-induced protein dimerization in living cells

HNIEDGSVQLADHYQQNTPIDGPVLPPDNHYLSTQSAKDPNEKRHDHMVLLEFVTAAIG
LGMDELYK

Amino acid sequence of FGG-EGFR-GFP
- Underlined sequence: Signal peptide
- Bold sequence: extracellular domain (I, II, III, and IV)
- Bold and underlined sequence: transmembrane region
- Italic sequence: intracellular domain
- Italic and underlined sequence: green fluorescent protein (GFP)

MRPSGTAGAALLAALCPASRAFGGKVCQGTSNKLTQLGTFEDHFLSLQRMFN
NCEVVLGNLEITYVQRNDILKTIQEVAGYVLIALNTVERPLENAQLIIRGNMYENS
YALAVSLNVDANKTGLKELPNQLEIHLGAVRFSNPNALCNVESIQWRDIVSSDFLS
NMSMDFQNHLSGSCQKCDPSCPNHSCGWAGEEENCQKLTIIACAQQCSCRGKSPSD
CCHNQCAAGCTGPRESDCLVCRKFRDEATCKDTCPPMLYNTPTYQMDVMNPBEKYSF
GATCVKVKKPRNYVVTDDHSCVRCAGSYEMEDGVRKCKKCEPCKRCVKNGIGIC
EFKDSLSINATIKHKFNCTSISGDHLIPVAFRGSFITHTPPLDPQELIDLKTVEITGF
LLIQAWPENRTDLHAFENLEIIRGRTQHQGQFLSAVVSNSLGLRKLKEISGDVII
GKKNLKCANTINWKKLFTSGKQKTIIISNREGSCKATGQVCALSPEGCWWPEPR
DCVSCRNSRGRCEVDKCNLLEGEPRECIQCHPECLPQAMNITCTGRGPDNC
IQCAHYIDGPHCVKTCPAGVMGENNTLWVKYADAGHVCHLCHPNCTYGCTGPGLE
GCPTNGPKIPSATGVMGALLLLLLLVALGIGILFMRRRHIRKRTLRLRLLQERELVEPLTPSG
EAPNQALLRLILKETEFKKIKVLGSGAFGTGVYKGLWIEPEGVKIPvAIKELEATS.PIKAN.ELDE
AYVMASVDPNHVCRLLGICTSTVQLITQLMPFGLIDYVREHKDNISQYLLNWCVIAK
MNYLEDRLRVLHRDLAAANVVLKTPHQHVTDFGLAKLLGAEEKYAEGGKVKIPKMALESI
LHRIYTHQDSDWWSYGTVWEMTGSKPDYDIPASEISSILEKGERLPQPICIDYVIMIVKC
WIMADSRPKFRELIEFSKMPDRQYVLIQGDRMIHPLPTDSNFYRALMEDMDDDV
DADEYLIQPGGSSPPSTRTPLSSSAATSNSTVACIDRNLQSCPITKEDSFLQRYSIDPTGA
LTEDSIDITFLPQVINGYSKRPAGSVQNPYVHNQPLNPAPRDPHYQDPHSTAVGNPEY
LNTVQPTCVNSTDFSDPAHWAQKGSQISLDPQYCQDFPFKEAKPNGIFGSTAENAELRV
APQSSSEFGIALGGGGGMVSKGEELETFGVVPIVELEDGVNGHFSVSGEGDAYSXKLT
LFICTTGGKLVPWPWTTLTVTLTVGQVCFSRYPDHMKOHDFKSAMPEGYVQERTIEFKKDDGN
YKTRAEVKFEGLTLVNRIELKGIDFKEDGNNILGHKLEYNISHNVYIMADKQNKGKINFKIR
HNEDGSVQADHYQQNTPIDGPVLPPDNHYLSTQSAKDPNEKRHDHMVLLEFVTAAIG
LGMDELYK
Design of two sets of EGFP and mCherry bearing N-terminal FGG motif on the surface.

The C-terminus of the surface proteins, LEHD-FGG-EGFP and LEHD-FGG-mCherry, containing a caspase-9 cleavage site (LEHD↓), were fused to the platelet derived growth factor receptor transmembrane domain (PDGFR-Tm) (Fig. 11). The N-terminus of these proteins was fused to a murine Ig signal peptide. DNAs encoding for LEHD-FGG-mCherry and LEHD-FGG-EGFP on the cell surface were amplified from mCherry and EGFP templates by PCR, respectively. The forward primer, 5’- GGC CAG ATC TGG CGG CAG CCT GGA GCA CGA CTT CGG CGG CGG CAG CGG CGG CGT GAG CAA GGG CGA GGA GCT GCT CTT CTT CGG CGG CGG CGG CGG CGG CGG CGT GAG CAA GGG CGA GGA GCT GCT CTT C3’, contained a caspase-9 cleavage site between N-terminal Ig signal peptide and FGG-EGFP/mCherry, and contained a BglII site for cloning. The reverse primer, CGA CCT GCA GGC CGC CGC TGC CGC CGC CGC CGA ACT GGA AGT ACA GGT TCT CGC TGC CGC CCT TGT ACA GCT CTG CCA TGC CGA GAG TG, contained a short GGGS linker and a PstI site for cloning. PCR product fragments were purified with PCR purification kit (Qiagen), treated with restriction enzymes (BglII and PstI), and ligated into pDisplay™ vector (Invitrogen).

Figure 11. Construction EGFP bearing N-terminal FGG motif on the living cell surface.

7.4.2 Cell culturing and DNA transfection

Hela cells were cultured in DMEM growth medium (Dulbecco’s Modified Eagle Medium) with 10% FBS (Fatal Bovine Serum) and 1% antibiotics, penicillin/streptomycin and incubated at 37 °C in 5% CO2. The cells were split on a 6-well plate one day before DNA transfection. DNAs (EGFR-GFP, FGG-EGFR-GFP) were transfected by using GenJet™ reagent for Hela cells (1 μg DNA:3 μl reagent) (SignaGen) and incubated at 37 °C in 5% CO2.
7.4.2 EGF/cucurbit[8]uril-induced FGG-EGFR-GFP dimerization and cell image

Cells were imaged on a widefield Zeiss Axio microscopy with GFP filter set 38 HE (excitation 470/40; emission BP 525/50) and on a Leica TCS SP5 AOBS (confocal microscopy) equipped with an HCX PL APO ×60/1.4 NA oil immersion lens and a temperature-controlled incubation chamber maintained at 37 °C and 5% CO2. GFP was excited using the 488 nm Ar laser line, emission band was detected in the range of 490-550 nm. The cells containing EGF (100 ng/ml) were incubated for 15 minutes. The cells containing cucurbit[8]uril (740 nM) were incubated for 1 hour. The pictures were imaged before and after addition of EGF or cucurbit[8]uril to the cells.

7.5 References


Supramolecular cucurbit[8]uril induced protein dimerization

Protein-protein interactions and protein dimerization play a key role in many biological processes. Understanding and controlling the molecular mechanisms that regulate protein dimerization is crucial for biomedical research. In the first chapter, different examples of molecular control over protein dimerization and opportunities for supramolecular chemistry in this respect are discussed. Approaches that allow induction of protein dimerization by small molecules provide powerful tools to modulate the functionality of dimerized proteins in gene expression, post translational processing and signal transduction. Supramolecular chemistry, as an orthogonal approach to induce protein dimerization, represents a novel and promising system in this respect. In this thesis, specific focus is on cucurbit[8]uril-based host-guest chemistry as a new, robust and versatile supramolecular approach to explore and modulate the molecular mechanisms of protein dimerization. Proteins bearing an N-terminal phenylalanine-glycine-glycine (FGG) motif were genetically generated by an intein-splicing system. A very short, genetically encoded, N-terminal FGG peptide motif allowed control over the dimerization of proteins such as fluorescent proteins, caspase-8, and caspase-9. The dimerization and activation of surface membrane proteins were also explored. Supramolecular control of protein dimerization thus provides an attractive novel concept for generating well-defined synthetic protein assemblies and for the regulation of biological events that depend on protein oligomerization.

In chapter 2, the generation of a pair of monomeric cyan and yellow fluorescent proteins (mCFP, mYFP) bearing a genetically encoded N-terminal FGG motif is described. Cucurbit[8]uril selectively recognized and induced both homo- and hetero-dimerization of these proteins, resulting in a decrease of fluorescence anisotropy (due to homo-FRET) in case of homodimerization and the occurrence of strong energy transfer from the CFP donor to the YFP acceptor (hetero-FRET) in case of random heterodimerization. The supramolecular induced protein dimers were stable and could be separated on a size exclusion chromatography column. The supramolecular interaction between cucurbit[8]uril and two FGG peptide motifs could also be reversed by the addition of a small competitor ligand (methyl viologen), which overall leads to inhibition of the protein dimerization.

The molecular mechanism of cucurbit[8]uril-induced fluorescent protein dimerization and their resulting complexes was further investigated by using dynamic light scattering and small angle X-ray scattering (SAXS) in chapter 3. Dynamic light scattering showed a 2-fold increase in apparent molecular size upon addition of cucurbit[8]uril to a protein with an N-terminal FGG sequence, whereas the size of a protein particle with an N-terminal methionine did not change in the presence of cucurbit[8]uril. SAXS data showed the two subunits of a
cucurbit[8]uril-induced FGG-mYFP dimer complex to have a z-shaped structure. The two fluorescent protein domains are in very close proximity forming a very compact protein dimer, despite being connected by a flexible peptide linker dimerized via cucurbit[8]uril.

In chapter 4, cucurbit[8]uril has been used as an inducer of protein tetramerization, by combining the twofold binding of an FGG motif to cucurbit[8]uril with intrinsic affinities between the proteins domains as a stepwise assembly process. The combination of a dimerizing interface on the fluorescent proteins under study (dYFP, dCFP), with an encoded N-terminal phenylalanine-glycine-glycine (FGG) peptide motif allowed cucurbit[8]uril to selectively recognize and induce FGG-dYFP or FGG-dCFP homotetramerization. The concept of cucurbit[8]uril-induced protein homotetramerization was elucidated by using a combination of dynamic light scattering and size exclusion chromatography experiments. Upon addition of cucurbit[8]uril to a solution of for example FGG-dYFP, pre-dimerized in solution, the tetrameric protein assembly was automatically generated.

Application of the supramolecular protein dimerization approach to control enzyme activation is described in chapters 5 and 6. Caspase-9 bearing a short genetically-encoded N-terminal FGG motif was generated by the intein splicing system in chapter 5. Formation of the highly catalytically active dimer could be induced by host-guest complex formation with cucurbit[8]uril. The activity of the cucurbit[8]uril-induced caspase-9 dimers was not only significantly greater than that of the isolated protein, but was also better than reported previously using a protein surface engineering approach. In addition, the cucurbit[8]uril-induced caspase-9 dimerization could be reversed by addition of a competitive FGG peptide to the dimeric complex. Supramolecular host-guest chemistry thus allows control over the activity of caspase-9 in two ways, an increase in caspase-9 activity by supramolecular-induced dimerization and a decrease in activity by reversal of the complex formation using a competitive ligand for cucurbit[8]uril.

In chapter 6 it is shown that the application of cucurbit[8]uril induced dimerization allows the study of caspase-8 proteins both in their monomeric and dimerized state, not achievable via classical point mutation engineering alone. This study provided novel insights in the caspase-8 activation mechanism. Monomeric caspase-8 variants were firstly generated via introduction of specific point mutations into the intrinsically dimerizing caspase-8 protein. As an example, the monomeric caspase-8(D384A) was found to be active against a short synthetic peptide substrate, but was incompetent to cleave the natural substrate caspase-3. The cucurbit[8]uril-induced dimer of caspase-8(D384A) was reactivated to full enzymatic activity back to the level of the wild-type caspase-8 for both the synthetic substrate and natural substrate. These results have shown that caspase-8 dimerization is tightly coupled with active site-rearrangement and that protein dimerization is the driving force for this active-site rearrangement.
In chapter 7, cucurbit[8]uril-induced dimerization of membrane proteins on the surface of living cells is explored. Firstly, an EGFR construct bearing an N-terminal FGG motif was engineered, based on engineering of the signal peptide cleavage site. Stimulation of this protein construct with the EGF ligand and cucurbit[8]uril led to interesting observations, which requires additional studies to elucidate the interplay of EGFR dimerization and internalization. Secondly, studies to explore cucurbit[8]uril-induced membrane protein dimerization using membrane proteins tagged with fluorescent domains were initiated. Molecular mechanisms to provide these proteins with an N-terminal FGG motif are investigated.

In conclusion, the combination of the attractively, very short FGG peptide motif with the easy accessibility of both the genetically encoded proteins and the synthetic supramolecular cucurbit[8]uril molecule holds great promises for the application as supramolecular inducer of protein dimerization, not only for the here reported fluorescent protein assemblies and caspases, but also for example for other multimeric enzymes and membrane proteins.
Dung Thanh Dang (Đặng Thanh Dũng) was born on March 29th 1982 in Dong Thap, Vietnam. In 2004, he received his Bachelor’s degree at University of Sciences-Vietnam National University in Ho Chi Minh, Vietnam. In 2005, he went to South of Korea and studied for his Master’s degree in the group of Prof. Yoon-Mo Koo at Inha University in Incheon and he obtained his Master in 2007 at this University. His Master’s subject focused on “lipase-catalyzed bioconversion in ionic liquids” which successfully achieved some publications and one patent on this subject. To improve his knowledge and education, he started an internship for three months at the Max Planck Institute of Molecular Physiology in Dortmund, Germany. At that time, he worked on the project of protein engineering in the group of Prof. dr. Luc Brunsveld. In September 2008, he went to The Netherlands and started as a PhD student in Chemical Biology at Biomedical Engineering Department, Eindhoven University of Technology in Eindhoven. His PhD’s project focused on “supramolecular cucurbit[8]uril induced protein dimerization” under supervision of Prof. dr. Luc Brunsveld. The main goal of this project was to use cucurbit[8]uril to reversibly control protein dimerization and activation. The most important results of his PhD research are described in this thesis.
My PhD project would not have been accomplished within 4 years without those people whom I would like to thank in this acknowledgement section.

First of all, I would like to express appreciation and thanks to my supervisor, Prof. Luc Brunsveld for giving me the chance to do the research first at the Max Planck Institute in Dortmund, Germany, and then at the Eindhoven University of Technology, The Netherlands.

Dear Luc! It has been almost 4 years since I started working under your supervision. You provided me a promising project with good working conditions and time for independent development. I would say my scientific research and communication skills were limited at the beginning; you are the one who helped me to improve these skills step by step. Especially, you have devotedly provided a lot of suggestions and corrections on my reports, manuscripts, presentations and this thesis; I have really learned a lot from you. Your kindness creates a comfortable atmosphere to do the research, and I really appreciate that.

I would like to thank the members of the thesis reading committee, Prof. Dr. R. P. Sijbesma, Dr. M. Merkx, Prof. Dr. J. A. Killian and Prof. Dr. A. R. Urbach for their critical comments and corrections which helped to improve the quality of my thesis. Especially, I would like to thank Dr. M. Merkx for useful discussions and suggestions throughout my PhD studies on my research and manuscripts. In addition, I would like to thank Prof. Dr. E. W. Meijer and Dr. P. Y. W. Dankers for kindly taking part in the PhD-defense committee.

I would like to thank Dr. Hoang and Dr. Trang for guiding and helping. Anh chị thân mến! Anh chị không chỉ là người hướng dẫn trong nghiên cứu khoa học, mà còn là người thân như một người anh, một người chị đã tận tình giúp đỡ và chia sẻ vui buồn với em trong công việc và cuộc sống, và đặc biệt khi em bệnh anh chị đã chăm sóc em rất là chu đáo. Những lúc thảo luận nghiên cứu, thí nghiệm chung với nhau, những cuộc trò chuyện vui vẻ và đặc biệt là những lúc phải hết sức là "tình táo" dùng giữa trong những lần tranh luận của anh với chị, những khi niêm này có lẽ sẽ không bao giờ quên. Em thành thật cảm ơn anh chị rất nhiều!

A number of different collaborations provided me with a chance to open up my scientific horizon and learn more collaborations and research techniques. I would like to thank Dr. Christian Moitzi (Anton Paar GmbH, Austria) and Dr. Ilja Voets for SAXS measurement and related discussions, which are described in chapter 3. I would like to thank Jenny Ibach (Max Planck Institute in Dortmund, Germany) for confocal microscopy measurements of membrane proteins which are described in chapter 7.

The members of our group and of the chemical biology laboratory I would like to thank for useful discussions and suggestions at lunch meetings, breakfast meetings and chemical biology meetings. Especially, I would like to thank Michael Sonntag (Micha) for our experimental cooperation. It was great to do research with you; your knowledge and microscopy expertise helped me a lot in my research and also thank you for being paranimf at my defense. I would also like to thank Lech Milroy and Matthew Burton for their English
corrections and useful discussions. Lech, although you didn't have plenty of time, you have put a lot of suggestions and corrections into chapters 1 and 5 of my thesis, I really appreciate that. I would like to thank Melissa Koay and Misha Golynskiy for their friendly help when I first started working in the biological laboratory. I would like to thank Joost van Dongen and Lou Xianwen for SEC-ESI-MS measurement. I learned a lot about LC-MS and related techniques from you. I would like to thank Pauline Neirynck for the chemical synthesis, and Wencke Adriaens for the synthesis of some key peptides. I would like to thank Andrea Schmidt and Annemie Koonen for help with protein crystallization and cell biology experiments.

I would like to thank Joke Rediker for all the official paperwork and for organizing social group and lab activities. I would like to thank Hans Damen and Peggy de Graaf – Heuvelmans for ordering and arranging the stocks at the biolab. Thanks also to Björne Mollet for cell lab management that made it easier to work there.

From my office members, I would like to thank Katja Petkau–Milroy, Dana Uhlenheuer, Maelle Carraz and Marcel Koenigs with whom I shared many scientific and non-scientific experiences, which improved my scientific research and made life more enjoyable. Katja thanks for being a great friend and helping me a lot with literature and research discussions. Dana, we worked on nearly the same topic from two different angles, thus it was really easy for us to discuss and your suggestions were always useful for me.

I want to thank the students who contributed with their research projects to my research. Thanks go to Loes Stevers for working on cucurbit[8]uril-induced surface protein membrane dimerization (chapter 7), Ilknur Aksoy for helping me with a series of proteins with point mutations, Jurgen Schill for work on the cucurbit[8]uril-induced protein tetramers (chapter 4), Yvonne Dorland and Thuur Onzen for working on caspase-8 (chapter 6), Jeroen Briels for working on the generation of proteins bearing an FGG motif in mammalian cells.

Scientific research needs to be accompanied with an enjoyable free time. I want to thank Sascha, Katja & Lech, Tess & Matt for wonderful parties with fantastic dancing. I also want to thank Brian Janssen for being the second paranimf at my defense, the Nolte-Meijer-Cup-players for soccer games on Wednesday evenings, and thank Lou for being my regular tennis opponent in the weekends.

Finally, I would like to show my gratitude to my family and Vietnamese friends. Ba Má kinh mến! Trong gia đình, con là người Ba Má thường yêu nhất và ngược lại Ba Má là người mà con yêu quý nhất. Dù không ở bên cạnh Ba Má, nhưng lòng thường yêu của Ba Má đã giúp con vững bước trên con đường sự nghiệp, và đây cũng chính là phần thưởng quý giá của con dành cho Ba Má. Con kính chúc Ba Má được mắn khỏe sống lâu để có thể tận hưởng cuộc sống vui vẻ hạnh phúc cùng con cháu. Cảm ơn các bạn Eindhoven (đặc biệt Việt & Liên) cho những bữa tiệc cuối tuần đầy ý nghĩa.

Dũng