Chemical Biology Approaches Targeting Cell Membrane Proteins

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 donderdag 27 juni 2013 om 16.00 uur

door

Michael Horst Sonntag

geboren te Berlijn, Duitsland
This research has been financially supported by STW grant 11022-NanoActuate

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-3395-4

Cover Design: M.H. Sonntag
Printed by: Wöhrmann Print Service, Zutphen, the Netherlands
# Table of Contents

## Chapter 1
**Targeting plasma membrane proteins**
1.1 The cell membrane................................................................. 2
1.1.1 Structure and function .......................................................... 2
1.1.2 Membrane proteins – a cell communicator .................................. 3
1.2 The epidermal growth factor receptor (EGFR)................................. 4
1.2.1 Introduction ......................................................................... 4
1.2.2 Structure of the epidermal growth factor receptor (EGFR)............. 5
1.2.3 EGFR activation and internalization ......................................... 6
1.2.4 Ligand independent monomer association ................................... 7
1.2.5 EGFR and cancer ................................................................... 8
1.2.6 The plasma membrane – controlled dynamics ............................ 10
1.2.7 EGFR association in nano-domains and its consequences regarding activity ................................................................................. 10
1.3 Probes to target plasma membrane proteins .................................... 11
1.4 Aim and outline of this thesis ....................................................... 14
1.5 References ............................................................................... 16

## Chapter 2
**Expression of the human Epidermal Growth Factor (hEGF) containing orthogonal ligation sites**
2.1 Introduction .............................................................................. 22
2.1.1 The epidermal growth factor (EGF) ........................................ 22
2.1.2 Biological relevance of EGFR and challenges to address dynamic interactions on the molecular scale .......................................................... 23
2.1.3 Strategies for site-specific labeling of human EGF ....................... 25
2.1.4 Design and expression strategy for human EGF featuring site-specific ligation sites ................................................................. 26
2.2 Results and Discussion ................................................................ 29
2.2.1 Refolding and purification approach 1 ...................................... 29
2.2.2 Refolding and purification approach 2 ...................................... 32
2.2.3 Refolding and purification approach 3 ...................................... 35
2.3 Conclusions .............................................................................. 39
2.4 Experimental ............................................................................ 40
2.5 References ............................................................................... 43

## Chapter 3
**Site-specific labeling of human Epidermal Growth Factor (hEGF)**
3.1 Introduction .............................................................................. 46
3.1.1 Bio-orthogonal ligation strategies ............................................ 46
3.1.2 Cysteines, a natural site-specific ligation site ............................... 49
3.1.3 Cysteines and native chemical ligation (NCL) ......................... 50
3.1.4 The thiazoline-ring, a cysteine protection group ........................... 51
3.2 Results and Discussion ............................................................... 51
3.2.1 N-terminal cysteine protection with thiazolidines ....................... 51
3.2.2 Deprotection of N-terminal cysteines ....................................... 55
3.2.3 Site-specific labeling of hEGF ................................................ 57
3.2.4 Activity and Live Cell Imaging of site-specific labeled hEGF ........... 58
3.2.5 A streptavidin platform generating mono- and multivalent hEGF constructs ......................................................................................... 60
3.3 Conclusions .............................................................................. 64
3.4 Experimental ............................................................................ 65
3.5 References ............................................................................... 68
Chapter 4
Expression and labeling of site-specific modified growth factors and nano-bodies
4.1 Introduction ................................................................................................................. 72
4.1.1 Growth factors and their ligands ........................................................................... 72
4.2 Results and Discussion ............................................................................................. 75
4.2.1 Expression and purification of Cys-Gly-Cys0_hEGF ............................................. 75
4.2.2 Expression and purification of TGFα ................................................................. 77
4.2.3 EGB4_Gly-Cys-Gly, a nano-body featuring a C-terminal orthogonal ligation site ... 80
4.3 Conclusions ............................................................................................................... 84
4.4 Experimental ............................................................................................................ 84
4.5 References .............................................................................................................. 86

Chapter 5
Fluorescent proteins with integrated RGD motifs
5.1 Introduction ............................................................................................................... 90
5.1.1 Fluorescent proteins ........................................................................................... 90
5.1.2 Integrins and the RGD motif .............................................................................. 91
5.2 Results and Discussion ......................................................................................... 93
5.2.1 Strategy for generating fluorescent mCitrine featuring RGD motifs .................. 93
5.2.2 Expression and purification of mCitrine mutants ............................................... 95
5.2.3 Fluorescence of mCitrine mutants containing RGD or cysteine motifs .......... 96
5.2.4 Determining the binding capability of mCitrine-RGD constructs to integrin receptors ... 97
5.3 Conclusions .............................................................................................................. 104
5.4 Experimental .......................................................................................................... 105
5.5 References .............................................................................................................. 108

Chapter 6
Supramolecular polymers applied to cells
6.1 Introduction ............................................................................................................. 112
6.1.1 The supramolecular polymer of choice – a self-assembling fluorescent unit .......... 113
6.1.2 Supramolecular discotics – a dynamic tool for biology ....................................... 114
6.2 Results and Discussion ......................................................................................... 115
6.2.1 Charge mediated cell uptake ............................................................................. 115
6.2.2 Receptor mediated cell uptake .......................................................................... 117
6.2.3 Electroporation mediated cell uptake ............................................................... 122
6.3 Conclusions ............................................................................................................ 124
6.4 Experimental .......................................................................................................... 125
6.5 References ............................................................................................................ 127

Chapter 7
Epilogue
7.1 Overview of the achievements described in this thesis ........................................... 130
7.2 Application of hEGF containing orthogonal ligation sites ....................................... 130
7.3 Application of RGD-mCitrine proteins .................................................................. 132
7.4 Application of supramolecular discotics inside and outside the cell ....................... 132
7.5 References ............................................................................................................ 134

Summary .................................................................................................................... 135
List of Abbreviations .................................................................................................. 138
List of Publications ..................................................................................................... 142
Curriculum Vitae ........................................................................................................ 143
Acknowledgement ...................................................................................................... 144
Targeting plasma membrane proteins

Abstract. The plasma membrane, the dynamic boundary of the cell, is decorated with a diverse range of proteins, which enable the cell to selectively sense and convert signals from its surrounding environment. Signaling molecules are actively targeting specific cell surface receptors, leading to a dynamic and reversible association and dissociation of proteins and other components within the plasma membrane, connecting extracellular signal with the intracellular signaling circuit (network). This responsive nature of the cell membrane is regulated by dynamic and reversible interactions, most of them lacking full understanding. In this chapter an overview over important plasma membrane receptors is given with a particular focus on the epidermal growth factor receptor (EGFR). Also, developed methods to probe membrane proteins site-selectively and to retrieve information about spatio-temporal protein interactions and conformational changes are discussed. These insights then serve to formulate the challenges addressed in this thesis regarding the development of chemical biology approaches to target and probe membrane proteins.
1.1 The cell membrane

1.1.1 Structure and function

The cell membrane is a highly complex and selective barrier surrounding all cells in a multicellular organism. Permeable for hydrophobic substances and small molecules, but impermeable for polar molecules, it enables the cell to separate and control intracellular material from the cell surrounding it. These gatekeeper properties, essential for cell survival and function, are accomplished with membrane proteins such as transporter proteins (channel proteins, carrier proteins, signaling proteins), which provide an indispensable mechanism for the cell to control the exchange with its environment.\(^1\)

Phospholipids are forming the lipid bilayer and are the most abundant compound found in membranes. The insertion of cholesterol into the bilayer makes the membrane more semi-fluidic and combined with different membrane proteins, which attach to the cytoskeleton or the extracellular matrix, giving the cell a stability that the plasma membrane alone would not be able to provide (Figure 1-1).\(^2\) Most phospholipids are equally mixed throughout the membrane, but they are non-equitably distributed between the inner and outer monolayer, a process that is actively maintained by lipid translocators such as P-type ATPases. This phospholipid asymmetry is functional important for recruiting, activating or guiding other molecules as well as for cell communication and signaling. As an example, the anionic phospholipid phosphatidylserine is in normal healthy cells restricted to the inner leaflet of the plasma membrane. However when the cell is getting activated to enter apoptosis (programmed cell death) the lipid asymmetry is disturbed. This results in the loss of the non-equal distribution and exposition of phosphatidylserine at the outer leaflet. The extracellular presentation serves as apoptosis marker within the organism and leads to the fast removal of the apoptotic cell via macrophages.\(^1,3\) The plasma membrane provides the cell therefore not only with structural stability and a gateway to move specific molecules out and into the cell, but it is also used as a dynamic and complex platform for cell communication and signaling.\(^4\)

![Figure 1-1: The eukaryotic plasma membrane. Membrane proteins are inserted into the membrane and connect extracellular and intracellular environment with each other. Adopted from reference.\(^2\)](image-url)
1.1.2 Membrane proteins – a cell communicator

Membrane proteins are located on or within the phospholipid bilayer in various ways and enable the cell to respond adequately to the hundreds of different signals from its surrounding environment. Membrane proteins are non-covalent associated with the cell membrane and act often as signal receptors in cell communication. Transmembrane protein (receptors), extend through the membrane bilayer for example via a β-barrel or via single or multiple α-helices.\(^1\)\(^5\)

The largest family of cell surface receptors in eukaryotic cells involved in signal transmission are G-protein-coupled receptors, (GPCRs).\(^6\) This type of receptor mediates responses to a wide variety of stimuli ranging from fatty acids,\(^7\) neurotransmitters\(^8\) and hormones, such as insulin/glucagon, important regulators for maintaining the blood sugar level in the human body,\(^9\) as well as is responsible for senses such as smell, taste or sight. GPCR associated signaling pathways are playing an important role in a number of normal physiological processes, but are also associated with human malignancies. Over 600 inactivating and about 100 activating mutations in human GPCRs were already linked to several human diseases\(^10\) and to date not even all GPCRs or their corresponding ligands are identified.

Integrin receptors are another important class of cell surface proteins responsible for cell adhesion and migration.\(^11\) As α/β heterodimeric receptor, termed integrin by Hynes et al. in 1986,\(^12\) the transmembrane receptor makes cell-cell and cell-matrix interactions connecting the intracellular cytoskeleton with the extracellular matrix (ECM).\(^13\) To date 18 α subunits and 8 β subunits are identified in humans which combine in different compositions to form 24 different integrin receptors, expressed on virtually every cell surface.\(^14\) Several splicing variants increase the integrin diversity further generating a complex pool of surface receptors essential for cell organization in organs and tissues as well as during cell migration and proliferation.\(^14\)\(^-\)\(^16\) As general integrin binding motif the peptide sequence RGD was identified and several integrin receptors are binding in a RGD depending fashion (for an overview see chapter 5 and Table 5-1). Integrin receptors detect and translate extracellular mechanical forces into intracellular chemical signals.\(^17\) The cell combines these signal with input signals received from other surface receptor such as GPCRs and receptor tyrosine kinases (RTKs) (ex.: epidermal growth factor receptor) in order to stabilize or reorganizing the cytoplasmic skeleton, crucial processes in normal as well as malignant cells and essential events in cell attachment, cell invasion, and metastasis typical for cancer.\(^17\)\(^-\)\(^19\)

Receptor tyrosine kinases (RTKs) are α-helical transmembrane proteins with similar structure, with 58 family members identified in the human genome, such as the epidermal growth factor receptor (EGFR).\(^20\)\(^21\) Many RTK family members are involved in a number of important cellular processes, mostly key events that mediate cell proliferation, differentiation and survival.\(^22\) Tight control of this receptor family and its signal transduction pathways guarantees that cells maintain their characteristics in a multi-cellular environment.\(^23\)\(^24\) Ligand binding mostly dimerizes the receptors and leads to the auto-phosphorylation of cytoplasmic key tyrosine residues, docking sites for the intracellular signal machinery.\(^1\) Mis-regulation of the activation or deactivation
process by disturbing the tightly controlled underlying molecular mechanisms is often associated with cancer, and is the focus of intensive research today.

Figure 1-2: Receptor tyrosine kinase receptor (RTK), G-protein coupled receptor (GPCR) and integrin receptor are cell surface receptors that convert external stimuli into intracellular signals. Reprinted and adopted with permission from reference. Copyright (2009) John Wiley and Sons.

1.2 The epidermal growth factor receptor (EGFR)

1.2.1 Introduction

The epidermal growth factor receptor (EGFR) has been the focus of intensive research. The EGFR also named ErbB1 or Her1, is a member of a tyrosine kinase receptor superfamily, which consists of EGFR (ErbB1), ErbB2, ErbB3 and ErbB4. This class of receptors provides a basis for many different signaling events and is strongly involved in cell development as well as proliferation.

The different ErbB receptor subtypes can form homo- and/or heterodimers and have different ligand specificities. They are expressed in a broad variety of tissues ranging from epithelial, mesenchymal and neuronal origin. Upon activation of the ErbB receptors, the recruitment of a diverse range of signal molecules results in the activation of different signal pathways. The three main pathways are: the MAPK pathway (mitogen-activated protein kinase), PI3K /AKT pathway (phosphatidylinositol 3-kinase) and the STAT pathway (signal transducer and activator of transcription proteins). Other receptor signaling pathway can cross-talk with these pathways and enhance their signaling output, such as the ligand activated integrin receptors or G-protein-coupled receptors.

Within the epidermal growth factor family the EGFR is the most-studied receptor, due to its strong association with human malignancies such as bladder, lung and breast cancers. Misregulation of the tightly controlled EGFR signaling disturbs often key pathways of normal
physiologic regulation and is involved in many pathological processes in particular cancer progression. Mutations in the EGFR often result in over-expressed, hyperactive or constitutively active receptors caused by alteration of the receptor inactivating properties. As crucial player in tumor growth and metastasis, the elucidation of these mutations could provide entry to interfere or treat cancer progression by developing drugs that selectively target or rescue altered receptors.

1.2.2 Structure of the epidermal growth factor receptor (EGFR)

The overall structure of all four members of the ErbB family is very similar, even though the amino acid homology is low. The ErbB receptors consist of one single polypeptide chain, in which the hydrophobic transmembrane spanning domain connects the ligand binding extracellular domain with the intracellular domain containing a tyrosine kinase domain. The N-terminal ligand-binding domain of EGFR is 621 residues long and is divided in four sub-domains I-IV; the leucine rich sub-domains I and III as well as the two-cysteine rich sub-domains II and IV (Figure 1-3). The two sub-domains I and II interact directly with the ligand and the domain II is providing the dimerization arm for receptor homo- or heterodimerization. The receptor transmembrane domain consists mainly of hydrophobic amino acids and anchors the receptor into the plasma membrane. At the C-terminal intracellular end of the protein is situated the juxtamembrane domain that separates the intracellular tyrosine kinase domain from the transmembrane domain as well as the C-terminal receptor tail.

X-ray crystallization studies revealed that each receptor can only interact with one ligand, resulting in a 1:1 stoichiometry between receptor and ligand. However, upon ligand binding, structural re-arrangements take place that lead to the exposition of the dimerization arm of the extracellular sub-domain II, resulting in a back to back dimerization of two receptors (Figure 1-3). In this created 2:2 complex the dimerization arm is crucial to enable receptor-mediated-dimerization, as mutations in the domain II loop demonstrated to prevent ligand induced back to back dimerization.

Figure 1-3: Schematic structure of EGFR. Left: closed or tethered conformation (ligand unbound state) and Right: the extended conformation (ligand bound) dimerized receptor.
1.2.3 **EGFR activation and internalization**

The activation of the EGFR family is initiated by ligand binding. Different EGFR ligands have different specificities and affinities to the individual EGFR members and are classified into four groups, regarding their binding characteristic for particular ErbB receptors (illustrated in chapter 4, Figure 4-3). To date no ligand for ErbB2 is known, but ErbB2 is the “preferred dimerization partner of all other ErbB receptors”. Both facts might be explained by the structure of ErbB2 that revealed a “[fixed] ligand activated state” exposing its dimerization arm continuously. ErbB3 is the only EGF receptor within the ErbB family that contains a defective kinase domain. The lack of activity is due to mutations of essential amino acid residues in the kinase domain. The receptor requires consequently hetero-dimerization with another EGFR family member to obtain an active working kinase domain. It is surprising that the ErbB2/ErbB3 complex is one of “the most potent ErbB signaling complex in terms of *in vitro* growth and transformation”.

Ligand induced EGFR dimerization results in the activation of the intracellular tyrosine kinase domain and binding of ATP, utilized for the phosphorylation of several distinctive tyrosine residues in the C-terminal cytoplasmic tail of the receptor (ErbB1: Tyr 992, Tyr 1045, Tyr 1068, Tyr 1086, Tyr 1148, Tyr 1173). The phosphorylated tyrosine residues serve subsequently as docking site for a broad range of proteins such as those featuring SH2 domains (SH2 = Src homology 2, a cytoplasmic tyrosine kinase), linking this way the receptor to the intracellular signaling pathways. The generated pattern of tyrosine phosphorylation depends on the ligand bound to the receptor and its dimerization partner. Defined by the tyrosine phosphorylation pattern, different subsets of intracellular signaling molecules are bound which in turn decides which signal pathway is activated and switched on. Using different phosphorylation patterns, the cells create an additional level of selectivity/diversity/complexity: a) by the different ligand affinities and selectivity to a given receptor pair and b) by the generated tyrosine phosphorylation pattern.

Following receptor activation EGFR is removed from the plasma membrane mostly by clathrin mediated endocytosis, but also clathrin independent endocytosis of EGFR is reported (Figure 1-4). Ligand activated EGFR is rapidly internalized with about 5 times faster rates than non-stimulated EGFR. The process is called ligand-induced down-regulation and decreases the overall number of EGFR proteins on the cell surface. By accelerating or reducing the rate of endocytosis and regulating the amount of receptor degradation and receptor recycling, the cell can actively tune the total amount of available receptors on the cell surface, which in turn regulates the stimulating strength of the available extracellular ligands. Mutations in the C-terminal tail as well as in the kinase domain of EGFR can slow down the internalization rate and therefore extend the lifetime of EGFR signaling.

Interestingly, the receptor internalization cannot only be triggered by ligands of the ErbB family, but also by other cell surface receptors. It is reported that the stimulation of the G-protein-coupled lysophosphatidic acid receptor or beta2 adrenergic receptor reduces the amount of available ErbB receptors on the cell membrane. On the other hand it has been shown that another class of G-protein coupled receptors mediate the activation of metalloproteases.
Metalloproteases are involved in the release of membrane bound EGF (ectodomain shedding, described in chapter 4, Figure 4.1) resulting in the increase of soluble EGF available in the cellular environment.\textsuperscript{52} Increasing the extracellular EGF concentration leads to rapid EGFR phosphorylation and signaling. The process is termed EGFR transactivation and could be linked to cell proliferation and anti-apoptosis properties of human cancer cells.\textsuperscript{53,54} Metalloproteases seem to control not only the amount of soluble ligand but also the activity of the transmembrane receptor itself, as demonstrated in mice.\textsuperscript{55} The characteristic of tumor cells to display an increased level of metalloprotease activity is currently explored, regarding its potential use as cancer biomarker for tumor imaging and target specific antibody activation.\textsuperscript{56,57}

Once the receptor is internalized following EGF stimulation, the ligand-receptor complex is processed in the endosomal system (Figure 1-4). Here the receptor can get recycled and transported back to the plasma membrane, or degraded in the lysosomes. However growing evidence suggests that receptor signaling is not terminated after endocytic uptake and that receptor accumulation in endosomes is used as another form of receptor signaling platform.\textsuperscript{56-58}

Figure 1-4: The EGFR activation pathway. The ligand binds to the extracellular domain of EGFR, which induces receptor dimerization and intracellular phosphorylation. The phosphorylation sites serve as docking site for intracellular signaling molecules. The receptor is following activation shut down by internalization and gets either degraded or recycled back to the cell surface. Reprinted and adopted with permission from reference.\textsuperscript{59} Copyright (2002) Science & Medicine.

1.2.4 Ligand independent monomer association

The above outlined conventional ligand-induced dimerization model for EGFR activation is based on an EGFR that exists as a monomer in the plasma membrane and is activated upon ligand binding, which results in receptor dimerization and downstream signaling. However this simple classification in non-signaling monomers and signaling dimers is not illustrating the complete picture. More and more studies reveal that the receptor exists partially as a non-active pre-dimer on the plasma membrane.\textsuperscript{60-62} Also oligomerization (clustering) of the receptor is observed before and after ligand stimulation.\textsuperscript{53,64} Clayton \textit{et al.}\textsuperscript{63} propose that EGFR populations
in unstimulated cells contain different mono- and oligomeric states, which are in equilibrium between a tethered and an untethered conformation. Following ligand binding the equilibrium shifts towards multimeric and kinase active conformations. Reports regarding the monomer / pre-dimer / oligomer ratios are varying. Hofman et al.\textsuperscript{65} observed 40% of the total EGFR population as non-active pre-dimer.

EGFR monomers and pre-dimers show different ligand affinities, as quantitative EGF binding studies demonstrated. The pre-dimer has a two orders of magnitude higher EGF binding affinity as receptors in the monomeric conformation.\textsuperscript{66,67} Displaying a receptor on the cell surface with two different types of affinities can be beneficial, because it widens the dynamic range of the receptor; the high affinity receptor variant allows a rapid cellular response also at low ligand concentrations whereby the low affinity receptor is active at high EGF concentrations as well as long-term stimulations.\textsuperscript{66,68}

1.2.5 EGFR and cancer

The epidermal growth factor family of tyrosine kinase receptors is involved in a broad range of fundamental cell processes such as cell proliferation and differentiation. Tight control of this receptor family is essential to maintain normal cell characteristics. Dys-regulation, mutations or over-expression of the ErbB receptor family is associated with many human diseases such as cancer (Table 1-1).

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>EGFR related</th>
<th>ErbB member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioblastoma</td>
<td>40-60</td>
<td>ErbB1 (ErbBvIII)</td>
</tr>
<tr>
<td>Head and neck cancer</td>
<td>90</td>
<td>ErbB1</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>30</td>
<td>ErbB2</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>40-80\textsuperscript{70}</td>
<td>ErbB1 + ErbB2</td>
</tr>
</tbody>
</table>

Only ErbB4 seems not to play a role in cancer progression.\textsuperscript{71} However most cancer types, which are EGFR mediated, have a poor clinical outcome. Especially in head and neck cancer as well as in glioblastoma cancers in which the ligand independent ErbB1 receptor variant ErbBvIII is often over-expressed, making this type of cancer very aggressive and therapy resistant; survival rate of patients diagnosed with glioblastoma is 15 months.\textsuperscript{69,72} The truncated receptor ErbBvIII is missing a huge part in the extracellular ligand binding and dimerization domain (amino acid deletion 6-273).\textsuperscript{73} This impairs the receptor to bind its natural ligands and to dimerize. However even without binding a ligand, ErbBvIII remains partially activated and can despite truncated dimerization domain form heterodimers with other family members,\textsuperscript{74} which results in continuous signaling.

Due to the involvement of EGFR in many cancer types, the receptor is a validated drug target and several therapeutical anti-cancer strategies have been developed; a) antibodies against the
extracellular domains, b) tyrosine kinase inhibitors and c) DNA / RNA targeting agents such as antisense RNA (Figure 1-5). Clinical most relevant are the first two approaches. Examples of monoclonal antibodies targeting the extracellular domain of EGFR include cetuximab (erbitux) and panitumumab (vectibix); monoclonal antibodies targeting ErbB2 at different extracellular domains are trastuzumab (herceptin) and pertuzumab. The problem of the outlined antibody therapies is their high rate of non-responding patients, 30% of breast cancer patients over-expressing ErbB2 do not respond to trastuzumab treatment, clear reasons for that are still not know. The epitope region targeted by the antibody or signal cascade cross-talks might be potential routes for cancer cells to escape this type of anti-cancer therapy. To improve target specificity, ErbBvIII could be the target of choice because it exist exclusively on tumor cells. Other emerging strategies are the combination of known therapeutics, the use of reversible blocked antibodies which are getting activated by tumor co-expressed metalloproteases, EGFR targeting nano-bodies or small synthetic antibodies such as SAB-Y1.

The tyrosine kinase domain of the EGF receptor is also an attractive drug target, because when blocked, intracellular downstream signaling is prevented. As such tyrosine kinase inhibitors (TKI) have been developed which compete with the ATP recruitment into the kinase binding pocket. Examples in clinical use today include gefitinib (iressa), erlotinib (tarceva), and lapatinib (tykerb).

Figure 1-5: Possibilities to inhibit EGFR signaling on different levels. The first two approaches are clinical used strategies today. 1) monoclonal antibodies against EGFR, 2) tyrosine kinase inhibitors, 3) antisense RNA 4) ribozymes, 5) bispecific antibodies targeting EGFR and a native immune cell to trigger immune mediated cytotoxicity 6) immunization against EGFRvIII, 7) anti-EGFRvIII monoclonal antibody coupled to toxin, 8) EGFRvIII ligand coupled to toxin. Reprinted and adopted with permission from reference. Copyright (2002) Science & Medicine.
Chapter 1

1.2.6 The plasma membrane – controlled dynamics

The spontaneous self-assembling of phospholipids into a lipid bilayer in water is the basis for the formation of membranes. The two-dimensional fluid structure/matrix forms closed compartments, essential for cell survival and function. These supramolecular structures are composed of a diverse mixture of self-organizing lipids and proteins, which define a dynamic cellular boundary equipped with unique and fundamental functionalities. The non-covalent interactions between lipids and proteins guarantee and maintain the dynamic of this inhomogeneous mixture. However the organization of the molecules within this dynamic landscape is not based on random lateral diffusion as initially proposed by Singer and Nicolson in 1972. This concept was drastically revised in the last decade as it became evident that organization principles in the membrane exist, which restrict the free diffusion of molecules. By quantifying the mobility of individual proteins using single-molecule tracking methods, confined zones on the nanometer level were revealed in which proteins can stay together for a certain time. In these distinctive domains, termed lipid rafts, protein islands or actin corrals, random diffusion rates were determined similar as found in artificial membranes and which are decreasing upon protein clustering.

Based on these observations, the “picket and fence model” was proposed connecting the intracellular actin network with the extracellular mobility restrictions found in the plasma membrane. The fences in the model are displayed by the actin-based membrane skeleton, the pickets are the actin bound transmembrane proteins. The transmembrane proteins are remaining in one confined domain for a certain time period before they are hopping to the next adjacent domain because the proteins are restricted a) by the cytoplasmic fences created by the actin filaments outlining the domain and b) many transmembrane proteins are coupled / linked to the actin network, which creates a fence along the actin filament. The so called hop diffusion described by the model could be shown for several proteins such as E-cadherin, G-protein coupled mu-opioid receptor and the epidermal growth factor receptor. Today it is thought that nature uses these dynamic platforms of 30-300 nm in size to partition the cell membrane into dynamic compartments, to temporal facilitate or prevent specific protein-protein or protein-lipid interactions.

1.2.7 EGFR association in nano-domains and its consequences regarding activity

The cell membrane organization is thought to play a crucial role in signal transduction and cell communication. Nature is using the nano-domains such as lipid rafts for trapping a certain subset of proteins and lipids into a defined region, to concentrate them and to enable or facilitate their biological function. Lipid rafts are small, “heterogeneous, highly dynamic, sterol- and sphingolipid-enriched domains” and several transmembrane proteins such as the epidermal growth factor receptor (EGFR) are predominately localized in them. However the molecular composition of lipid rafts and their role in EGFR accumulation and activation is still elusive. Described are the co-localization of EGFRs in gangliosides (GM) 1 enriched vesicles and its
uptake via clathrin-dependent and independent endocytosis. GM1 is reported to increase the auto-phosphorylation of the EGFR tyrosine kinase domain after EGF stimulation. On the other hand it is known that GM3 is inhibiting EGFR. Hofman et al. propose therefore a modulating effect of both gangliosides on the EGFR activity, depending on their presence in the lipid raft. Cholesterol, an abundant plasma membrane molecule, seems to influence this balance. It is further known to increase the EGFR phosphorylation in general, as several cholesterol depletion experiment demonstrated.

Whether EGFR is only accumulating in lipid rafts or also in other nano-domains such as caveolae is not clear. On the one hand Ringerike et al. showed EGFR enrichment in lipid rafts, but no increased EGFR concentration could be detected in caveolae. On the other hand Lajoie et al. report EGFR enrichment in caveolin-1 microdomains, which are reducing the EGFR activity. Independent whether EGFR signaling is enhanced or reduced in lipid rafts, caveolae or both, the studies demonstrate the dynamic influence of the plasma membrane on proteins such as EGFR. Because these dynamic interactions occur on a nano-meter scale, difficult to reveal with standard methods, highly sophistical techniques are being developed with spatiotemporal resolution to observe the true character of the plasma membrane and its impact on signal transduction.

1.3 Probes to target plasma membrane proteins

The structure and function of plasma membrane proteins, diverse and dynamic in nature, is still poorly understood. This is amongst others due to the challenge to label proteins selectively under physiological conditions, orthogonal to other plasma membrane proteins. Isolation of plasma membrane proteins by detergents and crystallization provides an alternative to protein labeling, but no spatial and temporal information can be obtained.

To accomplish protein labeling in live cells for visualization and quantification, different strategies have been developed. The simplest labeling approach is the use of membrane permeable and/or impermeable cysteine reactive probes that are coupled to a detection groups such as fluorescent dyes. Performing sequential digestion steps on the cell, information about the protein location can be gained; plasma membrane, periplasm or intracellular sub-compartment. However, despite the relative low abundance of cysteine residues in natural proteins, cysteines are not a protein specific labeling site due to their presence in many cellular proteins. Site-specific protein labeling is challenging but four strong methods have emerged: a) immunolabeling, b) fusion tags, c) unnatural amino acids and d) protein-ligand interactions.

Immunolabeling of plasma membrane proteins with antibodies revolutionized our understanding about proteins and is the most widespread technique today due to its simple use, high specificity in combination with low cross reactivity, availability for many targets as well as simple conjugation of fluorophores via conventional conjugation strategies. However antibodies suffer from their big size (about 150-200 kDa) and their restriction to proteins in the outer leaflet. Significant size reduction of the fluorophore-antibody complex can be achieved by
replacing the antibody by a nano-body (about 13 kDa).\textsuperscript{82,105} If no direct antibody / nano-body is available for the protein of interest, epitope tagging is an alternative.

Fusion tags: entire proteins, protein domains or short amino acid sequences, are normally introduced in the protein of interest at the N- or C-terminus.\textsuperscript{102} The fusion of fluorescent proteins (FPs) to the protein of interest is one of the most popular approaches because it results in the fluorescence labeling of the protein that can be analyzed by live cell fluorescent microscopy, regarding location, expression and dynamic.\textsuperscript{106} There exist a diverse range of FPs covering the whole visible spectral range, starting from CFP (cyan fluorescent protein) (blue), GFP (green fluorescent protein) (green), mCitrine (yellow) to DsRed (red).\textsuperscript{107} Additionally, FPs allow the investigation of dynamic protein-protein interactions or conformational changes upon stimulation with spatio-temporal resolution.\textsuperscript{108} Coupling two FPs together, separated by the protein of interest, conformational changes can be investigated via monitoring the FRET ratio (fluorescent resonance energy transfer), which depends on the distance as well as the orientation of the two FPs to each other. Changes in protein conformation shift the orientation or the distance of the two attached FPs, which results in a different FRET ratio. Dynamic protein-protein interactions can also be investigated with fluorescent proteins in live cell.\textsuperscript{109} For example, by coupling a cyan and yellow fluorescent proteins to the α- and β-subunit of the G-protein complex, the dynamic dissociation upon GPCR stimulation could be observed and its re-association after stimulus removal.\textsuperscript{110}

Another commonly used fusion tag is the SNAP tag, developed by K. Johnsson & co-workers. The SNAP tag is based on the modified enzyme hAGT (human O6-alkylguanine-DNA alkyltransferase), which transfers irreversible and selectively the alkyl group from benzylguanine (BG) to one of the cysteine residues in the SNAP-tag.\textsuperscript{111,112} Genetically fused to the protein of interest, the fusion tag provides entry for selective and fast protein labeling, intra- as well as extracellular, to any labeling reagent linked to BG.\textsuperscript{113} This was demonstrated at various targets such as the GPCR NK1.\textsuperscript{114} Despite the advantages of FPs and SNAP / Halo-tag, their large size (FP about 30 kDa, SNAP about 21 kDa) can be a limiting factor which influences the functional properties of the investigated protein. To avoid the mentioned problems and to be, in case of FPs not restricted to the visible spectrum,\textsuperscript{114} smaller fusion tags have been developed, such as the tetracysteine tag, or natural enzymes were adapted for site-specific conjugation, such as the biotin ligase BirA.

In contrast to the large FPs and enzymatic SNAP tags, the genetically encoded tetracysteine-biarsenical system consists only of a 12 amino acids long peptide sequence (LNCCPGCCMEP and HRWCCPGCCCKTF) which contains 4 cysteines for the binding of the biarsenical dyes, mostly FIaSh (green) and ReAsH (red).\textsuperscript{109,115} The biarsenical fluorophores bind via two arsenic atoms to the thiol groups of the cysteines residues in the recognition sequence with very high affinity, about 10 P.M.\textsuperscript{102} Additionally the peptide tag CCKAEACC has been developed, recognized by Cy3-based biarsenical dyes and complementary to the tetracysteine tag.\textsuperscript{116} The advantage of both tags is their size, demonstrated by FRET measurements on several proteins such as the human adenosine A(2A) receptor. When coupled to FPs, receptor downstream signaling is inhibited, but
Targeting plasma membrane proteins

native signaling occurs when using FIAsh labeled tetracysteine tag.\textsuperscript{117} The tetracysteine system allows further sequential pulse labeling because the used dyes are non-fluorescent; however upon binding the fluorescence increase strongly. This was advantageously used to label old and new imported connexin in gap junctions with different colors.\textsuperscript{118} The major disadvantages of both tags is their toxicity, due to the arsenic molecules necessary for binding, and their high background staining due to similar peptide motifs available in natural proteins, that however can be reduced with high concentrations of thiols.\textsuperscript{115,119} Despite these drawbacks, an Src kinase sensor was developed\textsuperscript{120} and tetracysteine tags are getting further explored to probe protein conformations changes (Figure 1-6).\textsuperscript{119}

![Figure 1-6](image_url)

Figure 1-6: Schematic representation of the Src kinase sensor. Conformational change is initiated by phosphorylation and results in the generation of the tetracysteine binding motif, which can be detected with ReAsH. Reprinted with permission from reference.\textsuperscript{119} Copyright (2011) American Chemical Society.

An alternative approach to achieve site specific and rapid labeling is the use of natural enzymes. The most common and best investigated enzyme is BirA, which in \textit{E.coli} biotinylates the biotin carboxy carrier protein (BCCP), but recognizes also a much shorter peptide sequence (15 amino acids long, also known as Avi-tag) as acceptor.\textsuperscript{121} Due to its unique motif, not recognized by eukaryotic ligases,\textsuperscript{102} it is an excellent peptide tag if fused to proteins and is opening up the use of the streptavidin biotin technology for downstream applications. Streptavidin, 60 kDa in size, binds biotin with very high affinity via its four biotin binding sites, but can also introduce protein cross-linking. Biotinylation of proteins by BirA could be shown \textit{in vitro} as well as \textit{in vivo}\textsuperscript{122} as demonstrated for the cell surface protein EGFR\textsuperscript{123} or in mice.\textsuperscript{124} Genetically engineered fluorophore ligases such as modified lipoic acid are able to couple 7-hydroxycoumarin to the lipoic acid ligase acceptor peptide (13 amino acids long, also called LAP).\textsuperscript{125} By exploring biotin ligases of different organism, Ting & co-workers discovered two ligases (\textit{Saccharomyces cerevisiae} and \textit{Pyrococcus horikoshii}) with a broader substrate tolerance, capable to couple biotin-azide conjugates or phosphine probe site-specifically to proteins.\textsuperscript{126} The introduction of a non-native chemical tag is very interesting, because it enables bio-orthogonal chemical reactions for the addition of probes via Staudinger ligation or Click chemistry.
Another approach to generate novel chemically reactive handles in proteins is via the site-directed incorporating of unnatural amino acids. This can be achieved by using the substrate tolerance of the methionine E.coli tRNA-synthetase, which accepts unnatural methionine derivates with azide functionalities such as azidoalanine or azidohomoalanine.\textsuperscript{127,128} Alternatively, Schultz & co-workers developed E.coli strains encoding proteins with 21 amino acids, 20 natural as well as one unnatural amino acid.\textsuperscript{129} This was achieved by using an unique tRNA molecule with an anti-codon that does not encode for a natural amino acid (stop codon) and a corresponding designed tRNA-synthetase, which recognize and charge the tRNA molecule but no other. Using both biomolecules in one microorganism, proteins can be generated featuring an unnatural amino acid at one given position. The amber tRNA suppression approach was demonstrated with several proteins such as insertion of p-aminophenylalanine in myoglobin\textsuperscript{130} or phosphotyrosine in STAT1.\textsuperscript{131} Despite the possibility to insert many different unnatural amino acids into any position of a given protein, general problems encountered are low yields, rejection of amino acids at certain positions and read through.\textsuperscript{132}

An alternative to achieve site-selective targeting of cell surface receptors is to use the ligand-receptor interaction. Ligands are frequently small and display a high specificity to their receptor (target), especially when used in a (hetero)-multivalent fashion.\textsuperscript{133,134} Ligand mediated targeting is used in drug delivery such as VEGF(121)/rGel, a fusion protein containing VEGF conjugated to gelonin,\textsuperscript{135} or in research for local EGFR stimulation such as EGF coated beads.\textsuperscript{136} Despite the successful use of these approaches, control over ligand density, orientation and geometry is limited. Natural ligands do not automatically display orthogonal handles, which would allow defined binding and activation studies such as for the comparison between low and high affinity ligands or stimulations with high and low local concentrations. Ligands featuring these properties are urgently needed because in comparison to global approaches, they enable controlled stimulation and manipulation of proteins, indispensable to elucidate the exact nature of cell communication and signaling.

1.4 Aim and outline of this thesis

Cell development and growth are both highly dependent on the fundamental ability of cells to accurately respond to changes in their environment.\textsuperscript{137} The predominant region in the cell where the conversion of extracellular to intracellular signals occurs is within the plasma membrane. The spatial organization of the proteins responsible for converting such external signals (e.g. the epidermal growth factor receptor (EGFR)) is highly dynamic and involves complex molecular ordering on the nano-scale level after external stimulation.\textsuperscript{138,139} Nevertheless, the mechanisms of receptor assemblies and receptor-receptor communication are still not well understood.\textsuperscript{140} On the other hand mis-regulations of plasma membrane receptors such as EGFR are often associated with human malignancies such as bladder, lung and breast cancers.\textsuperscript{25} Despite the efficiency of new anti-cancer drugs targeting EGFR, only a fraction of patients respond to this type of
treatment and the risk of recurrence is high. Consequently, intensive research is performed to improve the understanding between EGFR and its influence on cancer development.

The organization of plasma membrane receptors such as EGFR with other membrane-resident proteins into dynamic protein clusters or lipid rafts (compartmentalized membrane microdomains) has been suggested to modulate the signal transduction process. However, progress to reveal the composition of these supramolecular protein assemblies has been hampered by the fact that biochemical extraction technique used for isolation, influencing the clusters composition. An emerging view to reveal their composition and gain simultaneously insights in the dynamic assembly process is by super resolution microscopy. These techniques are getting more and more available and approaching resolutions on the nanometer scale, promising to understand the molecular mechanisms behind plasma membrane receptor clustering. However, their success strongly depends on the ability to manipulate these protein-protein interactions on the required nano-scale level without inducing artifacts due to the probe itself. Therefore it exists an urgent need for molecular probes to facilitate these types of nano-scale measurements, such as well-defined bioactive ligands for plasma membrane receptors containing bio-orthogonal ligation handles. The development of such ligands would enable the generation of probes with controlled ligand orientation and density, a pre-request to study the dynamics of receptor-ligand interactions on a single molecule level. Once the ligand density and orientation is controlled, defined actuation experiments on a nanometer / single molecule scale are possible, envisaged to provide new insights into cell signaling via the plasma membrane such as the impact of local versus global ligand stimulation on receptor dynamic and clustering. Controlling the extracellular ligand presentation is therefore highly important and strongly coupled to gain a detailed understanding of the dynamic, stability, clustering and signaling mechanisms of plasma membrane proteins.

The aim of this thesis is to develop probes featuring rationally designed properties, which enable the controlled stimulation, manipulation and observation of plasma membrane proteins. To address this objective, an efficient expression and purification protocol for the human epidermal growth factor (hEGF) featuring orthogonal ligation sites is developed, which might emerge as a versatile tool for the synthesis of a wide range of cysteine rich proteins, featuring site-selective modifications. By inserting RGD motifs at various positions into fluorescent proteins, the possibility is explored if protein scaffolds can serve as a platform to induce RGD conformations with different integrin affinities and specificities. Focusing on synthetic supramolecular scaffolds, self-assembling amphiphiles are investigated regarding their properties for cellular labeling, uptake and sub-compartment targeting.

Chapter 2 describes the development of an expression, refolding and purification method for hEGF bearing one or two orthogonal functionalities. The used intein technology offers thereby an intelligent platform for the removal of the purification tag by controlled auto-excision. The biological activity of the modified hEGF proteins is quantified using an EGFR phosphorylation assay.
In chapter 3 the site-selective deprotection and functionalization of the expressed hEGF constructs with fluorescent dyes or biotin is explored. Especially the deprotection and labeling procedure of the N-terminal ligation site is optimized. Covalent single or stepwise double labeling using maleimide chemistry is investigated and evaluated by mass spectrometry and live cell imaging.

The knowledge about refolding, purification and labeling obtained for hEGF in the previous chapters is applied to other cysteine rich proteins in chapter 4. Additionally a N-terminal double cysteine containing fusion tag is explored regarding its influence on hEGF folding and site-selective labeling.

Chapter 5 deals with genetically encoded and fluorescent RGD motifs based on mCitrine. To evaluate their affinity and specificity against different integrins, a micro-titer plate integrin assay is established. Screening hits are spectroscopically analyzed and combined to optimize their integrin affinity and specificity.

A supramolecular self-assembling system is investigated as dynamic synthetic platform for extra- and intracellular protein interactions in chapter 6. Charge mediated cellular uptake is analyzed in dependence on the peripheral amine density and explored regarding its carrier potential for non-cell permeable discotics. Additionally receptor mediated uptake is investigated using the SNAP-tag technology to allow cell specific targeting and uptake. To study the behavior of the supramolecular system in the cytoplasm of cells, initial electroporation experiments are performed.

1.5 References

(15) de Melker, A. A.; Sonnenberg, A. Bioessays 1999, 21, 499-509.
(16) Tarone, G.; Hirsch, E.; Brancaccio, M.; De Acetis, M.; Barberis, L.; Balzac, F.; Retta, S.
(22) Lemmon, M. A.; Schlessinger, J. Cell 2010, 141, 1117-34.
(23) Greco, H. E.; Schmick, M.; Bastiaens, P. I. Cell 2011, 144, 897-909.
(38) Sanderson, M. P.; Dempsey, P. J.; Dunbar, A. J. Growth Factors 2006, 24, 121-36.
(50) Sorkin, A.; Di Fiore, P. P.; Carpenter, G. Oncogene 1993, 8, 3021-8.
(57) Murphy, J. E.; Padilla, B. E.; Hasdemir, B.; Cottrell, G. S.; Bunnett, N. W. Proc Natl Acad Sci
Chapter 1

(69) Trejo, J. Signal Transduction: Pathways, Mechanism, and Diseases, 2010.
(84) Sako, Y.; Kusumi, A. J Cell Biol 1994, 125, 1251-64.
(112) Hinner, M. J.; Johnsson, K. *Curr Opin Biotechnol* 2010, 21, 766-76.
Expression of the human Epidermal Growth Factor (hEGF) containing orthogonal ligation sites

Abstract. This chapter investigates the recombinant expression of the human Epidermal Growth Factor (hEGF) featuring thiazolidine protected and unprotected orthogonal ligation sites. The cysteine rich protein hEGF was expressed containing additional N- or C-terminal protected or unprotected cysteines using the intein technology allowing for non-enzymatic removal of the affinity tag after purification. Three different purification and refolding approaches were explored to tackle problems regarding protein expression, purification, folding and intein cleavage. The establishment of highly efficient conditions allowed the purification of inclusion bodies expressed hEGF under denaturing conditions as well as for controlled intein cleavage, to yield a library of three differently modified hEGF constructs on a multi-milligram scale: Cys0_hEGF, hEGF_Cys54 and Cys0_hEGF_Cys54, featuring one or two additional cysteines N- or C-terminally incorporated. The biological activity of the newly generated hEGF constructs with additional ligation handles was tested on A431 cells, confirming their equivalent activity in comparison to unmodified natural hEGF. The newly developed protein engineering approach for the genetic introduction of selectively addressable cysteines in a cysteine rich and disulfide folded protein such as hEGF provides a strong molecular biology approach to generate site-specific modified proteins, ready to use for orthogonal labeling and immobilization experiments.
Chapter 2

2.1 Introduction

2.1.1 The epidermal growth factor (EGF)

The epidermal growth factor (EGF) was first discovered and isolated from sub-maxillary glands by Stanley Cohen in 1962. Injecting the crude protein extract in mice, Cohen observed an earlier induction of the eye opening and teeth growing. Motivated by these results, Cohen et al. isolated the protein responsible for it, EGF that set the stage for the identification and characterization of a number of additional growth factors. The isolation of human EGF (hEGF) from urine was simultaneously reported by Cohen and Gregory. Based on its ability to inhibit pepsin- and gastric acid secretion as well as its stimulating effect on gastrointestinal cells, hEGF was called “urogastrone”. Few years later, in 1977 Gregory & Preston could show that both molecules, urogastrone and hEGF, are identical.

In the 1970s, Cohen & Co-workers were able to purify hEGF and determine its primary amino acid sequence. The sequence of mature hEGF is composed of 53 amino acids featuring three intra-molecular disulfide bonds, which are crucial to reach its proper structure (Figure 2-1). The 6.2 kDa long polypeptide is non-glycosylated and very stable. It is generated in mammalian cells from a 1207 amino acid long precursor that is proteolytically processed by metallo-proteases. Interestingly, the precursor molecule contains a hydrophobic domain, assumed to be necessary for anchoring the molecule into the membrane. By regulating the amount of hEGF processing enzymes, cells are able to tune the concentration of cell associated and soluble growth factors (Figure 4.1, chapter 4). The effect of enzymatic solubilized hEGF molecules is mostly paracrine, leading to the activation of EGF receptors (EGFR) from adjacent cells. However also the non-processed pre-EGF forms are potent growth factors.

![Figure 2-1: Cartoon illustrating the primary amino acid sequence and structure of human epidermal growth factor (hEGF). Especially the formation of three disulfide bridges is essential to reach the bio-active conformation of hEGF. Identical amino acids are colored equal. Disulfide bonds between the cysteine are depicted in orange.](image)

An economical interest in EGF was already found in the early 1980s, when Moore et al. reported the rapid shedding of the fleece of Merino sheep, when treated with EGF. The Merino
wool is one of the finest and softest wool of any sheep and influenced by the commercial interest, intensive research was done to achieve high yielding EGF productions. However the traditional isolation of EGF from urine as described by Gregory and Willshire could not satisfy the enormous industrial requests. This resulted early in the exploration of recombinant DNA technology for the production of EGF. Several protocols for the expression of recombinant EGF are reported in literature. The most used expression host in this respect is Escherichia coli (E.coli), “one of the most extensively used prokaryotic organism [...] for the industrial production of proteins of therapeutic or other commercial interest”. E.coli offers, amongst other known and established expression systems, several advantages: it is genetically well characterized because of its long usage as model organism, commercial tools are available to facilitate gene cloning and expression of heterogenous proteins and simple process scale-up routes are established. To achieve high levels of heterogenous proteins in its correct and bioactive conformation, EGF can be expressed into the cytosol, in form of inclusion bodies, or soluble into the periplasm. However also expression systems in Pichia pastoris, Hansemmla polymorpha or Bacillus brevis are reported. The expression of cysteine rich proteins such as EGF in the cytosol leads to the formation of mis-folded and insoluble protein aggregates (inclusion bodies), because prokaryotic organisms keep their cystosol reduced, hindering disulfide formation in this cellular compartment. Subsequently, the expression of EGF in inclusion bodies will result in mis-folded and bio-inactive conformations of EGF. However if a suitable refolding protocol can be developed, the formation of the natural bio-active structure can be regained. Furthermore by over-expressing EGF in insoluble inclusion bodies the isolation and recovery of the expressed protein can be facilitated.

2.1.2 Biological relevance of EGFR and challenges to address dynamic interactions on the molecular scale

High yielding EGF production is of prime importance in basic and clinical research because tumor growth in many different species, including human, is highly dependent on the activation through growth factors, such as EGF. The epidermal growth factor EGF plays a potent role in the cancer progression in many different types of tumors and led to intensive research in the last decades to reveal its exact mechanism of action. EGF acts as a ligand on the trans-membrane receptor EGFR by activating the STAT, MAPK or PI3K/Akt pathways, promoting cell growth and motility. Due to the important role of EGF in cell proliferation, differentiation and apoptosis, “mis-regulation of the EGFR, over-expression or mutation [...] can cause malignant transformation of the cell”. EGF and its receptor have therefore become a main target for anticancer therapy as well as for basic research to reveal its exact mechanism of activation and dynamic behavior of EGF activated EGFR membrane clustering and signaling. However, the complexity of the activation process, which depends on several cellular components including “lipids, membrane proteins, protein-lipid assemblies/scaffolds and cytoskeletal elements” is hampering the ultimate understanding of the receptor activation process. The higher level of
organization and compartmentalization, reflected in protein clustering or membrane associated scaffolds (such as lipid rafts), is inseparable from its biological response and is thought to play a crucial role in the receptor activation process. To tackle this problem and get insights in the dynamic activation process, molecular and physical techniques are evolving to provide nanometer resolution in spatial and temporal space. New and improved visualization techniques such as super resolution microscopy on cells (Stimulated Emission Depletion (STED) microscopy, Stochastic Optical Reconstruction Microscopy (STORM), Photoactivated Localization Microscopy (PALM) or Atomic Force Microscopy (AFM)), allowing a snapshot on the dynamic behavior of the EGFR in living cells. These sophisticated techniques have become commercially available during the last 5-10 years and have been subsequently used by the scientific community to retrieve information on plasma membrane proteins to close the gap between the concept of lipid rafts and their existences in cells. However the techniques raise the need for well-defined suitable probes, because these techniques are reaching a limit in which the measured result can be either the consequence of the given stimulus or are induced artificially due to the used probe.

As such it is possible to observe and manipulate protein interactions on the required nano-scale level today, however precise physical or chemical protein activation or disturbances are only difficult to be accomplished with the available probe techniques. Therefore there is an urgent need for novel, molecular well-defined probes to facilitate nano-scale measurement. Important classes of such probes constitute native ligands, like EGF, featuring incorporated molecular tags via site-selectively positioned bio-orthogonal ligation sites (Figure 2-2).

![Figure 2-2](image-url): Concept and selective examples of bioactive ligand functionalization, based on a ligand featuring a single newly incorporated cysteine as site-selective orthogonal ligation site.

Inspired by the above described challenges, the development of an EGFR probe was envisaged, fulfilling the formulated nano-scale probe requests: a) natural ligand of a plasma membrane receptor, b) full control over density, orientation and geometry of the ligand functionalization and c) coupling methods that use facile and robust ligation chemistry. To achieve these requests, the expression of hEGFs featuring bio-orthogonal ligation sites on a milligram scale was envisaged.
2.1.3 Strategies for site-specific labeling of human EGF

Several strategies for the immobilization and labeling of proteins have been evolved in the last decades. The most common and easy to use are N-hydroxy succinimide (NHS) ester or maleimide based methods, targeting the reactive side chain of lysine or cysteine residues, respectively. However the nature of hEGF, featuring 2 lysines and 6 cysteines forming 3 disulfide bridges, limited to date the success for site-selective bio-orthogonal labeling. Therefore random labeling of hEGF via NHS chemistry is commonly used today.\textsuperscript{33,34} Although this approach is simple to perform, control over the labeling regarding position and number is highly limited.

A small number of attempts for site-specific labeling of hEGF is reported in literature; total chemical synthesis of hEGF,\textsuperscript{35} expression of lysine deficient hEGF variants,\textsuperscript{36} N-terminal PEGylation via reductive amination,\textsuperscript{37} and expression with cysteine containing tags.\textsuperscript{26} The chemical synthesis of the 53 amino acid long hEGF protein is a very attractive, because it opens up an efficient route for the incorporation of orthogonal ligation points such as maleimides or azides or other unnatural amino acids on a mg scale. Such a synthesis approach is reported by Shin et al.\textsuperscript{35} as well as of the EGF-like module of human complement protease C1r by Hernandez.\textsuperscript{38} The hEGF synthesis was realized by Shin et al.\textsuperscript{35} in very good yields, but the amount of correctly folded and bio-active hEGF remained low (\textasciitilde 3 mg). Due to the low yield, the synthetic approach was not further explored. An alternative to the total synthesis of hEGF, the expression of lysine deficient hEGF, is reported by Bach et al.\textsuperscript{36} (Figure 2-3). The approach allows the use of the well-established and commercially available NHS-chemistry to label a single available lysine. However the expression yield of lysine deficient hEGF variants is very low (0.3-0.6 mg) and ligands featuring non-native proteins sequences are normally avoided in biological studies due to their potential to create artifacts. Both facts are limiting their wider application for site-directed immobilization on surfaces and carrier systems. To circumvent this bottle neck, while still making use of the popular NHS-chemistry, scientist often replace hEGF with mouse EGF (mEGF),\textsuperscript{39} which is naturally lacking lysine residues (Figure 2-3). However, for fundamental studies of the human EGFR the use of hEGF would be preferential over mEGF. Additionally, in therapeutic applications the use of hEGF would advantageously avoid the induction of a human anti-mouse immune response.

| Human | NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGIGERCQYRDLKWWELR |
| Pig   | NSYSECPSHDGICLHGGVCMYIEAVDSYACNCVFGYVGERCQHRDLKWELR |
| Dog   | NGYRECPSYYDGYCLNYNCVCMYIEAVDRYACNCVFGYVGERCQHRDLKWE LR |
| Cat   | NSYQECPSYDGYCLNYNCVCMYIEAVDRYACNCVFGYVGERCQHRDLKWE LR |
| Rat   | NSNTGCPSPYDGYCLNGGVCYMIEVSDRYVCNICVGMVIGGERCQHRDLRRWKL |
| Mouse | NSYPCPSYDGYCLNGGVCMLHIESLDSYTCVCVIGYSGDRCQTRGLRWVELR |

Figure 2-3: EGF amino acid sequence of various mammalian species. The lysine residues are highlighted.\textsuperscript{36}

An alternative reported approach for site-specific coupling via an amine functionality is N-terminal PEGylation via a reductive amination reaction. The method is applied directly on
expressed and folded natural hEGF, avoiding problems regarding refolding of synthesized or genetically modified hEGF. The method itself makes use of differences in the pKₐ value of the amino groups in hEGF protein: the pKₐ value of the N-terminal primary α-amine is 7.8 in comparison to 10.1 for the ε-amino group in the lysine side chains. Lee et al. applied this approach to hEGF, however the labeling efficiency remained very low (~10% labeling of hEGF), making it practically and financially not attractive for a broad application.

![Figure 2-4: Reaction scheme for N-terminal PEGylation. PEG-propionaldehyde reacts with the N-terminal amino group at pH 5-8 and forms a stable amine bond followed by the reduction by sodium borohydride.](image)

2.1.4 Design and expression strategy for human EGF featuring site-specific ligation sites

Human EGF (hEGF) was expressed in Escherichia coli as fusion protein, featuring N-terminal a chitin binding domain (CBD) for purification as well as an intein domain for traceless removal of the purification tag (Figure 2-5).

![Figure 2-5: Schematic representation of the envisioned fusion protein CBD-Intein-hEGF. The expressed CBD-Intein-hEGF will be purified via chitin affinity chromatography. hEGF will be liberated from the resin via intein mediated auto-cleavage. CBD = chitin binding domain.](image)

The designed fusion construct (CBD-Intein-hEGF) was expressed in form of inclusion bodies as described by Esipov et al. The expression of hEGF in inclusion bodies would be advantageously, because it facilitates the isolation and recovery of the expressed protein and combined with the intein technology, provides a promising approach for the production of hEGF featuring orthogonal ligation sites. The introduction of the bio-orthogonal ligation sites should be achieved by: a) one N-terminal thiazolidine-ring protected cysteine (Table 2-1) and b) one C-terminal unprotected cysteine as a second ligation site (Table 2-1). We envisioned that this would be advantageously and lead to the generation of bio-active hEGF, featuring bio-orthogonal
ligation sites by providing full control over the density and location of the applied label via standard thiol chemistry.

Using the inducible and auto-catalytic self-cleavage activity of protein splicing elements such as inteins, any stage of conventional enzymatic cleavage of the fusion construct, for separating the purification tag from hEGF, is avoided. Unlike other purification systems, the intein technology generates a hEGF molecule possessing an N-terminal free amino acid. Having all these benefits in mind, different fusion constructs were designed, to produce hEGF with N- or C-terminal cysteines as orthogonal ligation sites (Table 2-1).

Table 2-1: Overview of envisaged modified hEGF constructs. Cysteine residues, thiazolidine-rings and disulfide bridges are highlighted in orange.

<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys0_hEGF</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>hEGF_Cys54</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>Cys0_hEGF_Cys54</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

The solubilization of the hEGF inclusion bodies was carried out in high concentration of chaotropic reagents such as urea. It was hypothesized that aldehyde impurities in urea itself or the addition of aldehydes should result in the reversible protection of the N-terminal cysteine as thiazolidine-ring (Figure 2-6). The use of thiazolidine-ring protected cysteines is a common approach in peptide synthesis and can be in combination with native chemical ligation a very powerful tool, as demonstrated by Kent et al. with the one-pot synthesis of crambin, IGF and VEGF. However the application of the thiazolidine protected cysteine in expressed proteins has not been applied to our knowledge to date.
Figure 2-6: Reversible protection and deprotection of the introduced N-terminal cysteine on human EGF. The deprotection is induced at acidic pH in the presence of O-methylhydroxylamine.

The protection of the newly introduced N-terminal cysteine did prevent interference of the otherwise freely available sulfhydryl group with the six natural cysteines of hEGF. This selective protection method avoids potential problems with refolding of the native protein, by disulfide shuffling, cysteine oxidation or hEGF dimerization. Additionally, the reversible protection of the N-terminal cysteine could allow the introduction of a second cysteine at the hEGF C-terminus, providing a second ligation site. As result, with all natural occurring cysteines in hEGF forming disulfide bridges, one thiazolidine-ring protected and one unprotected cysteine would be available as orthogonal ligation site, selectively and step-wise accessible (Figure 2-7). Due to the protection of the first cysteine by thiazolidine-ring formation, the second unprotected cysteine is available for selectively labeling via thiol reactive compounds. After the first labeling reaction is completed, the N-terminal thiazolidine-ring protected cysteine can be selectively deprotected and subsequently labeled, thus allowing the site-selective introduction of two different labels (Figure 2-7). This was envisioned to be advantageous for the molecular control over hEGF modifications and should allow the sequential orthogonal double labeling of hEGF with different probe elements. The detailed procedure of the labeling protocol is described in chapter 3.

Figure 2-7: Cartoon illustrating the double labeling strategy of human epidermal growth factor (hEGF) featuring N- and C-terminal site-selective ligation sites.
2.2 Results and Discussion

Inspired by Esipov et al., illustrating the production of unmodified epidermal growth factor in *E.coli* using the intein technology, three different fusion proteins were envisaged: CBD-Intein-Cys0_hEGF, CBD-Intein-hEGF_Cys54, and CBD-Intein-Cys0_hEGF_Cys54 (Table 2-1). All three fusion constructs contain an *N*-terminal chitin binding domain (CBD) for affinity purification and an intein domain for auto-catalytic and traceless removal of this purification tag. The intein domain, an auto-catalytic self-cleavable protein element, generates a free amino acid at the *N*-terminus of hEGF, planned to be used as bio-orthogonal ligation point (Figure 2-7).

The expression of small proteins and peptides such as hEGF is often difficult to achieve in *E.coli*, because they are not folding correctly or are subject to proteolytic degradation. The fusion of hEGF to the large CBD-Intein domain and its expression as fusion protein was therefore followed to stabilize the hEGF produced in the *E.coli*. Inclusion bodies are typically not prone to proteolytic degradation and therefore the over-expression of the fusion protein construct in these insoluble protein aggregates was thought to favor the overall yield.

2.2.1 Refolding and purification approach 1

![Cartoon illustrating the expression and purification approach 1.](image)

The fusion construct CBD-Intein-Cys0_hEGF was generated by cloning the Cys0_hEGF gene into the expression vector pTWIN1. This resulted in a DNA construct containing a chitin binding domain (CBD) as well as an intein domain fused in frame *N*-terminal to the Cys0-hEGF. The
expression of the fusion construct CBD-Intein-Cys0_hEGF in inclusion bodies was successfully achieved in E.coli BL21 following the expression protocol from Esipov et al. After extraction of the fusion protein from E.coli, the protein was refolded at air using the condition described by Esipov et al. Subsequently CBD-Intein-Cys0_hEGF was purified on chitin beads according to Esipov et al., and on-column intein cleavage was induced by lowering the pH to 6.0 and increasing the temperature to room temperature. However, on column intein cleavage was not efficient, resulting in low amounts of released Cys0_hEGF (100-200 µg per liter LB media) in comparison to milligrams of uncleaved product (Figure 2-9).

Figure 2-9: SDS-PAGE demonstrating the low intein cleavage efficiency. CBD-Intein-Cys0_hEGF was purified on chitin beads and intein cleavage was induced on column by increasing the temperature to 37°C and lowering the pH to 6.0. Lane 1 and 2: concentrated elution fractions, top band CBD-Intein purification tag, lower band Cys0_hEGF. Lane 3: chitin resin after intein cleavage. The resin fraction illustrated the high amount of uncleaved fusion protein that is still available on the column.

To improve the low cleavage efficiency responsible for the insufficient yield, the temperature during intein cleavage was increased to 37°C, reported to facilitate Ssp dnaB intein mediated protein cleavage. However no improvement of the intein cleavage efficiency was observed (Figure 2-9). To further investigate if the amino acid composition adjacent to the intein cleavage site had an unfavorable influence, the amino acids at the cleavage site were exchanged to well-established cleaving pairs, such as cysteine-arginine or serine-arginine, or cleaving pairs such as glycine-cysteine, cysteine-glycine, cysteine-aspartic acid-glycine, serine-aspartic acid (Figure 2-10). However no improvement regarding the on-column cleavage efficiency could be obtained after inducing intein mediated protein cleavage: temperature increased to 37°C, pH lowered to 6.0, confirmed by the strong upper protein band in the SDS gel (uncleaved fusion protein: CBD-Intein-Cys0_hEGF, Figure 2-10).
Expression of the human Epidermal Growth Factor (hEGF) containing orthogonal ligation sites

Figure 2-10: Different N-terminal amino acids used in CBD-Intein-Cys0_hEGF fusion protein to investigate the intein mediated protein cleavage efficiency. The splicing efficiency was analyzed by SDS-PAGE. **Left:** Amino acid composition tested in the intein cleavage site, **Right:** SDS gel; Thick protein band: CBD-Intein-Cys0_hEGF with different amino acids in the intein cleavage site, weaker lower protein band: CBD-Intein purification tag.

Intein cleavage following inclusion body refolding resulted in low amounts of liberated Cys0_hEGF, as outlined above. To increase the intein cleavage efficiency and subsequently total yield, protein expression in different cellular compartments (periplasm and cytosol) were explored in E.coli strain BL21. The over-expression of CBD-Intein-Cys0_hEGF in the periplasm failed. However expression of CBD-Intein-Cys0_hEGF in the soluble fraction of the cytosol was accomplished, but was accompanied with high levels of intracellular intein auto-cleavage, of about 50%, as SDS-PAGE analysis revealed (Figure 2-11, pH 7.5 fraction). This result can be rationalized since the Ssp dnaB intein is engineered to undergo pH and temperature dependent cleavage to liberate its target protein.\(^45\) Using standard E.coli expression conditions at 37°C, optimal conditions for intein cleavage were available.

To investigate if pH and temperature induced intein cleavage can be achieved, the soluble expressed CBD-Intein-Cys0_hEGF fusion protein was incubated at 37°C overnight within a pH range of 6.0 to 7.5 (Figure 2-11). The electrophoretic analysis revealed pH depending intein cleavage, observed by the decreasing intensity of the upper protein band (uncleaved fusion protein CBD-Intein-Cys0_hEGF) in an increasing acidic environment (Figure 2-11). In order to evaluate whether chitin binding domain (CBD) in the fusion construct allows the purification via affinity chromatography, the protein was incubated with chitin beads. However no binding of the CBD to the chitin resin was observed. With no second purification tag in the fusion construct available, the purification in high yield and purity could not be achieved. Re-considering the results, it was concluded that not the amino acids in the intein cleavage sites are the limiting factor for high cleavage efficiency, but the refolding of the whole fusion protein itself. Taking this into account, the focus was put back on refolding the CBD-Intein-Cys0_hEGF inclusion bodies. The refocusing should not only facilitate the expression and purification, but also the binding of
the chitin-binding domain (CBD) to the chitin resin. Therefore several refolding conditions were investigated in more detail in Refolding and purification approach 2.

Figure 2-11: SDS-PAGE demonstrating the pH dependency of intein mediated protein cleavage. CBD-Intein-Cys0_hEGF is expressed in soluble form in E.coli BL21 and incubated at 37°C under different pH conditions.

2.2.2 Refolding and purification approach 2

Figure 2-12: Cartoon illustrating the expression and purification approach 2.

The expression of the fusion construct CBD-Intein-Cys0_hEGF in form of insoluble inclusion bodies allowed the simple pre-purification and isolation of the fusion construct from the crude
cell extract. The pre-purification was achieved via repeated intensive washing and centrifugation of the cell pellet, which resulted in the removal of most soluble substances. The remaining aggregates in the cell pellet were insoluble, containing the recombinant fusion construct CBD-Intein-Cys0_hEGF in high yield. The inclusion bodies were solubilized in high concentrations of chaotropic reagents (Urea) and subsequently further purified by size-exclusion chromatography. Having the pure fusion construct in hand, different refolding and cleavage conditions were investigated.

Figure 2-13: S SDS-PAGE demonstrating the enrichment of the CBD-Intein-Cys0_hEGF fusion construct in inclusion bodies after washing and its high purity following size-exclusion chromatography. Lane 1: supernatant, Lane 2: washed inclusion bodies, Lane 3: size-exclusion purified inclusion bodies.

The intein cleavage efficiency was first investigated on chitin beads, since a simultaneous purification and cleavage strategy would be beneficial for the yield of the purification process; following the intein cleavage, the hEGF is released and eluted from the chitin resin while the CBD-intein tag remains bound on the chitin beads via the chitin binding domain (CBD). Refolding of the fusion protein CBD-Intein-Cys0_hEGF according to Esipov\textsuperscript{41} and incubation with equilibrated chitin beads resulted in the binding of the fusion construct to the chitin resin. However, no intein cleavage was observed upon cleavage induction, using a temperature increase to 37°C and pH 6.0. Besides pH and temperature, the influence of Tris-(2-carboxyethyl)-phosphine (TCEP) on the intein cleavage efficiency was investigated by equilibrated the chitin resin in presence of small amounts of TCEP (0.1-0.5 mM TCEP) prior induction of the intein cleavage. However no improvement regarding the cleavage efficiency was observed. Speculating that optimal intein cleavage might be hindered sterically or by protein aggregation on the chitin resin, solution induced intein cleavage was explored. The fusion construct CBD-Intein-Cys0_hEGF was refolded in different refolding buffers (Table 2-2). Subsequently the intein cleavage was induced by lowering the pH to 6.0 and increasing the temperature to 37°C. As
depicted in Table 2-2 / Figure 2-14, the intein cleavage could be induced in fusion constructs refolded in buffers containing a redox pair (cysteine – cystine) or when the fusion construct was treated prior to refolding with small amount of TCEP. These results confirmed that intein cleavage is possible in the tested fusion construct, but that refolding of the fusion construct as well as cleavage environment play a key role for high intein cleavage efficiency.

Table 2-2: Composition of tested refolding buffers.

<table>
<thead>
<tr>
<th>Refolding buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>50 mM Tris (pH: 6.0), 200 mM NaCl, 10 mM EDTA</td>
</tr>
<tr>
<td>2)</td>
<td>50 mM Tris (pH: 10.0), 200 mM NaCl, 10 mM EDTA</td>
</tr>
<tr>
<td>3)</td>
<td>1 M GndHCl, 0.1 M Tris, 12 mM Cysteine, 2 mM Cystine</td>
</tr>
<tr>
<td>4) 2) + TCEP</td>
<td>Denatured inclusion bodies treated with TCEP prior refolding</td>
</tr>
<tr>
<td>5) 3) + TCEP</td>
<td>Denatured inclusion bodies treated with TCEP prior refolding</td>
</tr>
</tbody>
</table>

Figure 2-14: SDS-PAGE illustrating cleavage efficiency of CBD-Intein-Cys0_hEGF after refolding and cleavage in different refolding buffers with and without TCEP incubation prior to refolding. Intein cleavage could be induced when the fusion construct was treated with TCEP prior refolding or was refolded in buffer containing a redox pair.

Establishing conditions for successful refolding and intein cleavage in solution, high yielding expression of Cys0_hEGF, hEGF_Cys54 and Cys0_hEGF_Cys54 should be possible. However, affinity purification of the CBD-Intein-hEGF variants via the chitin tag is only possible under native conditions. Consequently, in urea dissolved inclusion bodies can not be purified via the chitin affinity tag. Therefore the purification of the inclusion bodies was carried out by intensive washing and size-exclusion chromatography (Figure 2-12). After refolding and intein cleavage in solution, a second size-exclusion chromatography step was performed to separate the CBD-intein domain from the hEGF constructs (Figure 2-12). Using this expression and purification method, all desired constructs (Cys0)_hEGF_(Cys54) could be obtained in very pure form, illustrated by the representative LC-MS spectrum of hEGF_Cys54 in Figure 2-15. However, the yields were low, about 1 mg per liter LB media for Cys0_hEGF, most probably due to the non-optimal purification
strategy. Furthermore the advantage of the intein technology, purification and intein cleavage on a solid support for intelligent removal of the purification tag could not be realized using this strategy. Considering these disadvantages and aiming to use the intein technology more efficiently the focus was put on another purification tag, the His-tag.

Figure 2-15: LC-MS traces of purified hEGF_Cys54. **Left:** MS Total Ion Count, **Right:** mass to charge ratio, Inset deconvoluted mass spectrum; expected mass hEGF_Cys54 $3x$ S-S: 6319.1, observed mass 6320.1.

2.2.3 Refolding and purification approach 3

Figure 2-16: Cartoon illustration the expression and purification approach 3.
The His-tag is one of most frequently used affinity tags for protein purification and is normally composed of 4-8 histidines localized at the N- or C-terminus of the recombinant protein of interest. The popularity of the His-tag for purification is mainly caused by its remarkable high affinity and selectivity to nickel-nitrilotriacetic acid (Ni-NTA) matrices, but also other ions on metal-chelating surfaces can be used such as cobalt or copper.\textsuperscript{46-48} The His-tag is small, not interfering with the folding of the protein and allows the purification of the recombinant protein under native and denaturing conditions. Especially the purification under denaturing conditions is attractive for inclusion body purification, because affinity chromatography via other affinity tag such as chitin-tags is only possible under native conditions. Due to this restriction of the chitin-tag, the affinity purification of all three fusion constructs CBD-Intein-(Cys0)_hEGF_(Cys54) were only possible after refolding, relying strongly on the correct formation of tertiary structure of the chitin domain. As described in refolding and purification approach 1, the correct conformation of the chitin binding domain could not always be achieved, hampering the overall purification process. To exclude these problems and to obtain pure protein prior to refolding, the purification under denaturing conditions was envisaged. Therefore the in Table 2-3 illustrated fusion constructs were designed containing 9 histidine residues at the N-terminus for affinity purification.

Table 2-3: Overview of designed His-tag fusion constructs.

<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-CBD-Intein-Cys0_hEGF</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>His-CBD-Intein-hEGF_Cys54</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>His-CBD-Intein-Cys0_hEGF_Cys54</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Having the respective constructs in hand, the expression and purification using the His-tag strategy under denaturing conditions was successfully carried out (Figure 2-17). However, when inducing the intein cleavage in solution, as described in refolding and purification approach 2, a reduction of the cleavage efficiency of about 50% was observed (Figure 2-17). The position of the His-tag is localized far away from the intein cleavage site and should typically not have an influence on the cleavage efficiency.

Figure 2-17: Size-exclusion trace at 215 nm of His-CBD-Intein-Cys0_hEGF after refolding and cleavage induction with imidazole (150 mM) (left) and without imidazole (right). The cleavage efficiency was reduced by 50% in the presence of imidazole, indicated by the two peaks at the retention time 160 min (left). 100% intein cleavage efficiency was obtained under imidazole free conditions (right).

To investigate the reduced intein cleavage efficiency in more detail, the focus was put on imidazole, a competitive agent used in the purification process. The imidazole ring is part of the histidine structure that binds to the nickel ions immobilized on the Ni-NTA resin. The surface exposed histidine residues in the His-tagged fusion protein bind specifically to the purification media. Untagged host cell proteins do not bind to the Ni-NTA resin. By washing the purification media several times a separation between untagged protein and the target protein is achieved. To elute the pure target protein from the resin, free imidazole was added, a His-tag competitive agent disrupting the binding between the histidine residues and Ni-NTA resin that led to the elution of the His-tagged protein. Consequently, the eluted fusion constructs contained high concentrations of imidazole. To investigate the influence of imidazole on the intein cleavage, imidazole was removed prior to refolding and intein cleavage induction by dialysis. By carrying
out refolding and intein cleavage under imidazole free conditions, full intein cleavage efficiency could be regained (Figure 2-17).

Using the above described expression and purification conditions: a) recombinant expression of His-CBD-Intein-hEGF fusion protein in inclusion bodies, b) inclusion body purification under denaturing condition via the His-tag, c) refolding in solution in the presence of small amounts of TCEP, d) cleavage induction under imidazole free conditions at 37°C, pH 6.0 and e) removal of the purification tag by size-exclusion; all three desired constructs (Cys0)_hEGF_(Cys54) could be obtained in very pure form on a milligram scale (Figure 2-18). The total yield was enhanced strongly to up to 3.5 mg per liter LB-media for Cys0_hEGF following purification and refolding approach 3. In comparison, refolding and purification approach 1 resulted in 200 µg Cys0_hEGF, refolding and purification approach 2 resulted in 1 mg Cys0_hEGF. The expression yield could be further raised to up to 10 mg per liter expression media when using a nutrition rich auto-induction media, such as MagicMedia (Table 2-4). The yield for hEGF_Cys54 and Cys0_hEGF_Cys54 was slightly lower with 2.7 mg and 1.5 mg per liter LB media when using refolding and purification approach 3 (Table 2-4). All final products were analyzed by mass spectrometry, confirming the results of the SDS gel in respect to purity and mass, representative LC-MS spectrum of Cys0_hEGF can be found in Figure 2-18.

Table 2-4: Overview of obtained yields of Cys0_hEGF, hEGF_Cys54 and Cys0_hEGF_Cys54 in LB and MagicMedia using refolding and purification approach 3.

<table>
<thead>
<tr>
<th></th>
<th>LB media</th>
<th>MagicMedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys0_hEGF</td>
<td>3.5 mg / l</td>
<td>10 mg / l</td>
</tr>
<tr>
<td>hEGF_Cys54</td>
<td>2.7 mg / l</td>
<td>_</td>
</tr>
<tr>
<td>Cys0_hEGF_Cys54</td>
<td>1.5 mg / l</td>
<td>_</td>
</tr>
</tbody>
</table>

Figure 2-18: LC-MS traces of purified Cys0_hEGF. Left: MS Total Ion Count, Right: mass to charge ratio, Inset deconvoluted mass spectrum; expected mass Cys0_hEGF 3x S-S: 6319.1, expected mass Thiazolidine: 6331.1, observed mass: 6331.5.
To investigate the biological activity of the recombinantly expressed hEGF variants, the new constructs were tested in an EGFR activity assay. The assay determines the capability of the recombinant expressed hEGF to activate the epidermal growth factor receptor (EGFR) on A431 cells (an EGFR over-expressing cell line) and allows the comparison with the natural unmodified hEGF. By incubating the generated hEGF constructs Cys0_hEGF and hEGF_Cys54 on A431 cells, full biologically activity was detected (Figure 2-19). The activity was even slightly higher in comparison to commercial unmodified hEGF, probably an effect caused by the small number of measured cell samples or due to the possible higher purity in comparison to commercial hEGF. The genetic introduction of additional and selectively addressable cysteines, protected and unprotected, in the natural cysteine rich hEGF protein did not disturb its native disulfide bonding. Full biological activity equivalent to unmodified hEGF was confirmed indicating the native conformation of the generated hEGF constructs.

Figure 2-19: Activity of Cys0_hEGF and hEGF_Cys54 on A431 cells in comparison to natural unmodified hEGF.

2.3 Conclusions

The development of an efficient expression protocol for the production of the cysteine rich protein human epidermal growth factor (hEGF) containing N- and C-terminal orthogonal ligation sites, in the form of additionally incorporated cysteines, has been described. By introducing thiazolidine protected and unprotected cysteines at the N- or C-terminus of hEGF respectively, the following constructs: Cys0_hEGF, hEGF_Cys54 and Cys0_hEGF_Cys54 were successfully expressed in E.coli on a multi-milligram scale. By using the intein technology, intelligent removal of the purification tag via pH and temperature shift was achieved.

In a systematic approach three different expression and purification protocols were developed and investigated regarding yield, intein cleavage efficiency and purification. In the purification and refolding approach 1, hEGF over-expression in insoluble inclusion bodies was successfully achieved, but on-column intein cleavage efficiency was low. Exploring the over-expression of hEGF in soluble form, excellent intein cleavage properties were obtained, but affinity purification
via the chitin tag failed. Faced by this problem, purification and refolding approach 2 refocused on the over-expression of hEGF in inclusion bodies and refolding them under different conditions. Refolding hEGF in the presence of the redox pair cysteine / cystine or treatment of the denatured hEGF with TCEP prior to refolding resulted in very good intein cleavage characteristics in solution. However the complicated as well as time-consuming inclusion body purification process reduced the total yield to 1 mg per liter LB media. Purification and refolding approach 3 investigated possibilities to improve the total yield. By introducing a His-tag, inclusion body purification was strongly facilitated and resulted in final yields of up to 3.5 mg per liter LB media for Cys0_hEGF. The amount of expressed and purified Cys0_hEGF was further increased to 10 mg per liter expression media, using MagicMedia. The biological activity of the N- and C-terminal modified hEGF constructs were proven on A431 cells, showing equivalent activities to natural unmodified hEGF.

The generation of hEGF probes on a good milligram scale featuring selectively addressable handles, enables bio-orthogonal hEGF labeling and immobilizations experiments with full control over density and orientation, probes urgently needed in field of bio-nano-science. As versatile ligand for functionalization and activity studies, initial bio-orthogonal labeling experiments were carried to illustrate the site-specific control gained over hEGF. These experiments are described in chapter 3.

2.4 Experimental

Materials and methods

All chemicals and compounds have been used without prior purification, if not otherwise stated. The restriction enzymes, T4 ligase, Fast AP thermosensitive alkaline phosphatase, adenosine triphosphate (ATP) and isopropyl-beta-D-thiogalactopyranoside (IPTG) have been obtained from Fermentas. DNA polymerase Platinum-Taq, TOPO-TA cloning kit, MagicMedia, RPMI1640 media, penicillin streptomycin (Pen/Strep), fetal bovine serum (FBS) and SeeBlue® Plus2 pre-stained standard were purchased from Invitrogen. All oligonucleotides were ordered from eurofinsgenomic. Bio-safe Coomassie Brilliant Blue, Econo-Pac chromatography columns, 30% acrylamide mix (30% Acrylamide/Bis Solution, 29:1 (3.3%)), brom phenol blue, 6x DNA loading dye and N,N,N,N′,N′-tetramethyl-ethane-1,2-diamine (TEMED) was purchased from BioRad. Sodium dodecyl sulfate (SDS), yeast extract, sodium chloride, imidazole, and ammonium persulfate (APS) were ordered from Merck. Glycerol, glycine, guanidine hydrochloride (GdnHCl) and ethylenediaminetetraacetic acid (EDTA) were obtained from Calbiochem. Chitin beads and pTWIN1 vector were bought from New England Biolabs (NEB). All competent E.coli cells, benzonase, phosphate buffered saline (PBS) tablets and Ni-NTA resin was obtained from Novagen. Peptone was ordered from Fluka. Blue-white select screening reagent, β-mercaptoethanol, ampicillin, urea, egg lysozyme, nickel(II)sulfate hexahydrate (NiSO4), trypsin, sodium hydroxide (NaOH), tris (2-carboxyethyl) phosphine (TCEP), cysteine, cystine, triton-x 100, diithiothreitol (DTT), and tris-base were obtained from Sigma. Agar was purchased from Difco. B-Per and Inclusion Body Solubilization Solution was obtained from Thermo Scientific. Amicon dialysis and concentration devices were obtained from Millipore. Filters (0.2 µm, 0.45 µm) were obtained from Pall. Hydrochloric acid (HCl) has been purchased from VWR. A431 cells (ACC: 91) have been obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). NucleoSpin Gel and PCR Clean-up kit was ordered from
Expression of the human Epidermal Growth Factor (hEGF) containing orthogonal ligation sites

Macherey-Nagel, QIAprep Spin Miniprep kit from Qiagen and PathScan Phospho-EGF Receptor (Tyr1173) Sandwich ELISA kit from Cell Signaling.

*hEGF Plasmid Construction:*
The hEGF gene was PCR amplified from pEZZ-hEGF (gift from Bas van der Woning, Radboud University Nijmegen, The Netherlands) using Platinum Tag DNA Polymerase and the forward primer hEGF-SapI-F (5’-ttgggtgtctctccaacagcactgtaagc-3’) and the reverse primer hEGF-BamHI-R (5’-gtggtgatctctgagcatgcc-3’). To generate Cys0_hEGF, the forward primer hEGF-SapI-F was replaced by the primer hEGF_Cys0-SapI-F (5’-ttgggtgtctctccaacagcactgtaagc-3’) and for hEGF_Cys54 by the reverse primer hEGF-BamHI-R (5’-gtggtgatctctgagcatgcc-3’). To create Cys0_hEGF_Cys54, the forward primer hEGF_Cys0-SapI-F and the reverse primer hEGF_Cys54-BamHI-R was used. The amplified PCR fragments were gel purified and digestion with SapI and BamHI for 2 hours at 37°C in the presence of Fast AP thermosensitive alkaline phosphatase. The digestion was stopped by an incubation step at 80°C for 20 min. The digested DNA fragments were purified, ligated into the SapI/BamHI linarized expression vector pTWIN1 using T4 ligase and transformed in competent E.coli (Novablaue, XL1 or XL10-Gold).

The sequence of the resulting constructs [CBD-Intein-(Cys0)]_hEGF_(Cys54)] were confirmed by DNA sequencing. To add one N-terminal His-tag to the CBD-Intein-constructs, all constructs were amplified using Platinum Tag DNA Polymerase using the forward primer Ndel-His-FM (5’-gtgatatacgacactcatcataaattgaaggaaggtctgaata-3’) and the reverse primer hEGF-BamHI-R or hEGF-Cys54-BamHI-R, respectively. The amplified PCR fragments were TOPO-TA cloned into pCR2.1 vector and transformed in competent E.coli cells (Novablaue, XL1 or XL10-Gold). The transformants were sequenced and TOPO-TA clones containing the correct sequence were purified and digested with Ndel and BamHI for 2 hours at 37°C in the presence of Fast AP thermosensitive alkaline phosphatase. The digestion was stopped by incubating the reaction mix at 80°C for 20 min. The digested DNA fragments were purified and ligated into the Ndel / BamHI linarized expression vector pTWIN1 using T4 ligase. The sequence of the resulting plasmids were confirmed by DNA sequencing.

Amino acid sequence of [HIS]-CBD-Intein-(Cys0) - hEGF (Cys54):
- Chitin binding domain (CBD): underlined sequence
- HIS-tag: italic and underlined sequence in brackets
- Intein domain: italic sequence
- hEGF: bold sequence
- in brackets: N- or C-terminal mutations

```
M(HHHHHHHHHH)KIEEGKLTNPGVSAWQVNTAYTAGOLVTVNGKTYKCLOPHT
SLAGWEPSONVPALWQLQNNNGNGLELRESGAISGDLSILSALTGKRVISIKDLLE
KDFEIWAINEQTMKLESASAKSVRVFCXKLYILVTRLGRITKATANHRFLTIDGWK
RLDELSLKEHALPRKLESSMSLQISPEIKLSQSDIYWSVISITETGVEEVFDLTVP
PHNFVANDIIIVHN[C]NSDSECPLSHDBGCLHDEGMYCMEALDKYACNCVGVYI
GERCQYRDLKWWELR[C]
```

Expression:
The constructed pTWIN1 hEGF constructs have been transformed in chemical competent E.coli BL21 (DE3) cells. Clones containing the respective plasmid were inoculated in 20 ml Luria Bertani (LB) culture containing 50 µg / ml ampicillin. The culture have been grown in a shaking incubator at 37°C overnight. On the next day, 10 ml of the overnight culture were used to inoculate 1 liter LB media containing 50 µg / ml ampicillin. The culture was placed on a shaking incubator at 37°C for about 2.5 hours. At an optical density of A600 = 0.8, the expression was
induced with IPTG, final concentration 0.5 mM. The culture was placed back into the shaking incubator and incubated at 37°C for 5 more hours. Subsequently, the culture was harvested and pelleted by centrifugation at 10000 g, 10 min, at 4°C. The cell pellet was stored at -80°C.

Refolding and Purification approach 1:
Cells were lysed and purified as described by Esipov et al.\textsuperscript{41}

Refolding and Purification approach 2:
Cells were lysed in B-Per containing 1 µl benzonase / ml for 30-45 min on a rocker at room temperature. The lysate was centrifuged at 17600 g, 45-60 min, at 4°C and the supernatant discarded. The remaining pellet was resuspended in 20 ml B-PER containing 10 mg lysozyme from chicken egg white. The suspension was incubated for 30-45 min on a rocker at room temperature and centrifuged at 17600 g, 45-60 min, at 4°C. The remaining inclusion body pellet was dissolved in 3-10 ml Inclusion Body Solubilization Reagent overnight at 4°C and on the next day clarified by centrifugation at 17600 g, 30 min, at 4°C as well as filtration (0.45 µm). The dissolved inclusion bodies were applied onto a size exclusion column (GE Healthcare HiLoad 26/60 Superdex 75 prep grade, equilibrated with 4 M urea, flow-rate 1 ml / min). The peak corresponding to the fusion construct [CBD-Intein-(Cys0)-hEGF-(Cys54)] were collected and concentrated in Amicon filters, cut-off 10 kDa.
To refold all fusion constructs, the concentrated fractions were treated with 0.2 mM TCEP for 0.5 hours on ice and added drop wise into ice-cold refolding buffer R (50 mM Tris-HCl, 200 mM NaCl, 10 mM EDTA, pH 10) under strong stirring. The folding was performed 48 hours at 4°C at air. After the oxidative folding (to form the 3 disulfide bridges in hEGF), the intein cleavage was induced in solution by reducing the pH to 6.0 and increasing the temperature to 37°C overnight. On the next day, occurring white precipitates have been removed by centrifugation at 17600 g, 10 min, 4°C and filtration (0.45 µm). The clarified solution was concentrated via Amicon filters, cut-off 3 kDa prior to removal of the His-CBD-Intein purification tag by size exclusion (GE Healthcare HiLoad 26/60 Superdex 75 prep grade, equilibrated with PBS pH 6.0, flow-rate 1 ml / min). The fraction containing the hEGF was concentrated via Amicon filters, cut-off 3 kDa and analyzed by LC-MS.

Refolding and Purification approach 3:
Cells were lysed in B-Per containing 1 µl benzonase / ml for 30-45 min on a rocker at room temperature. The lysate was centrifuged at 17600 g, 45-60 min, at 4°C and the supernatant discarded. The remaining pellet was re-suspended in 20 ml B-PER containing 10 mg lysozyme from chicken egg white. The suspension was incubated for 30-45 min on a rocker at room temperature and centrifuged at 17600 g, 45-60 min, at 4°C. The remaining inclusion body pellet was dissolved in 3-10 ml Inclusion Body Solubilization Reagent overnight at 4°C and on the next day clarified by centrifugation at 17600 g, 30 min, at 4°C. The clarified supernatant was diluted in 12-40 ml binding Buffer B (8 M urea; 0.1 M NaH₂PO₄; 0.01 M Tris-HCl, pH 8.0) and loaded onto 5 ml Ni-NTA resin equilibrated with Buffer B. The resin was washed with 5 column volumes Buffer B, 5 column volumes Buffer B containing 10 mM Imidazole and eluted in 5 column volumes Buffer B containing 300 mM imidazole. The eluent was dialyzed 4 times against 4 M Urea and concentrated via Amicon filters, cut-off 10 kDa to a final volume of 5-10 ml. To refold all fusion constructs, the eluent was treated with 0.2 mM TCEP for 0.5 hours on ice and added drop wise into ice-cold refolding buffer R (50 mM Tris-HCl, 200 mM NaCl, 10 mM EDTA, pH 10) under strong stirring. The folding was performed 48 hours at 4°C at air. After the oxidative folding (to form the 3 disulfide bridges in hEGF), the intein cleavage was induced in solution by reducing the pH to 6.0 and increasing the temperature to 37°C overnight. On the next day, occurring white precipitates have been removed by centrifugation at 17600 g, 10 min, 4°C and filtration (0.45 µm). The clarified solution was concentrated via Amicon filters, cut-off 3 kDa prior to removal of the His-CBD-Intein purification tag by size exclusion (GE Healthcare HiLoad 26/60 Superdex 75 prep grade, equilibrated with PBS pH 6.0, flow-rate 1 ml / min). The fraction containing the hEGF was concentrated via Amicon filters, cut-off 3 kDa and analyzed by LC-MS.
Expression of the human Epidermal Growth Factor (hEGF) containing orthogonal ligation sites

**SDS-PAGE:**
SDS-PAGE electrophoresis was performed on a Mini-PROTEAN 3 electrophoresis system (Biorad, Hercules, California). The gel consisted of a 5% stacking gel and a 12% or 15% separating gel, respectively. The running buffer contained 25 mM Tris-HCl, 250 mM glycine, and 0.1% (w/v) SDS in H2O. Prior loading onto the SDS-gel, all protein samples have been a) mixed with sample buffer (100 mM Tris-HCl, 20% (v/v) glycerol, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, pH 6.8 in Millipore H2O) contained 25 mM Tris, 200 mM glycine and 0.1% (w/v) SDS, b) reduced by addition of β-mercaptoethanol and c) denatured by heating up the sample to 95°C for 10 min. The electrophoresis was run at room temperature at 80 V for 20 min and then at 140-160 V for 60 min. The protein bands were stained with Coomassie Brilliant Blue.

**EGF activity assay:**
To determine the activity of hEGF and its conjugates, the endogenous level of phospho-EGF Receptor (Tyr1173) in A431 cells has been determined, using the PathScan Phospho-EGF Receptor (Tyr1173) Sandwich ELISA kit. Cell lyses, sample preparation and ELISA have been carried out as described by the manufacture. In brief, A431 cells were seeded in 24 well plates (10⁵ cells / well) and cultured in RPMI1640 media supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin / streptomycin at 37°C, 5% CO₂ overnight. At the next day, the culture medium was replaced by RPMI1640 medium without FBS. Following the 12-24 hours starvation, cells were stimulated with 1 ml Cys0-hEGF or its conjugates (100 ng / ml) for 5-10 min at 37°C and then lysed.

2.5 References

(2) Garfield, E. Essays of an Information Scientist 1987, 10, 106.
Abstract. There is a strong need for novel, efficient and site-selective biomolecule functionalization strategies for generating fundamental insights into the functioning of a wide variety of proteins in- and outside the cell, for normal as well as malignant processes. Especially for more complex proteins, such as those folded via multiple disulfide bridges, these biomolecule functionalization strategies require urgent addressing. Here the expression and site-selective and step-wise labeling of the cysteine rich protein, human epidermal growth factor (hEGF), featuring one or two orthogonal ligation sites is described. Insertion of an N-terminal thiazolidine protected cysteine in hEGF, allows successful protein expression and folding without interference of the N-terminal cysteine with the disulfide formation of the native six cysteines. This enabled the insertion and selective modification of a second unprotected cysteine as an orthogonal ligation site at the C-terminus of hEGF. Addressing the N- and C-terminus in a stepwise fashion, sequential single or double labeling of hEGF could be achieved. Full functionality of the protein conjugates was illustrated by EGFR activation experiments, confirming its bio-active conformation. Initial experiments with streptavidin featuring different numbers of biotin binding sites demonstrate the potential use of hEGF displaying orthogonal ligation sites.
3.1 Introduction

3.1.1 Bio-orthogonal ligation strategies

Covalent modification of bio-molecules is a very powerful tool to gain inside into the biological function of the respective proteins, by monitoring or manipulating cellular events in their native setting.\(^1\) In nature, post-translational modifications provide a wide range of molecular changes, that add additional layers of functional properties to the proteome and in turn mediate protein activity.\(^2\) Covalent modifications on proteins are therefore highly attractive\(^3\) and have permitted the development of a number of assays for protein analysis \textit{in vitro} as well as \textit{in vivo}. Sharing a common interest to investigate biological processes using chemical methods (commonly referred to as chemical biology), chemists and biologists provide chemical and biological access to a wide variety of covalently modified proteins, to reveal their precise role in their natural environment. In combination with innovative techniques such as genetically encoded fluorescent proteins or monoclonal antibodies,\(^3,4\) protein analyses and tracking experiments could be performed, data otherwise difficult to obtain. However, large fluorescent proteins often introduce major structural disturbances, causing loss of the biological activity and antibodies are restricted to the extracellular space.\(^4\) Limited by these approaches, biochemists explored orthogonal functionalities not found in nature which not interfere with the biological function of proteins and that react spontaneously, rapidly and site selectively under physiological conditions.\(^5\) These so-called “bio-orthogonal reactions” are an ongoing challenge for scientists because they have to be performed under conditions preventing protein denaturation: a) in aqueous solutions at biocompatible pH (6 to 8), b) in the presence of salt and/or surfactants, c) at low non-toxic concentration of reactant and d) under low to ambient temperature.\(^2,6\) To find reactions that satisfy all these criteria is difficult and has limit the number of available bio-orthogonal ligation methods.\(^7\) In spite of the described difficulties, different methods have been developed such as: the Diels-Alder reaction, Staudinger ligation, the Huisgen cycloaddition, the Cu-free click reaction (Table 3-1),\(^3,8,9\) but also tagging methods such as SNAP or tetracysteine tags.\(^10,11\)
Table 3-1: Selected bio-orthogonal reactions.\textsuperscript{3,8,9}

<table>
<thead>
<tr>
<th>Reaction Type</th>
<th>Reaction Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diels-Alder reaction</td>
<td><img src="image" alt="Diels-Alder Reaction" /></td>
</tr>
<tr>
<td>Staudinger ligation</td>
<td><img src="image" alt="Staudinger Ligation" /></td>
</tr>
<tr>
<td>Staudinger ligation, traceless</td>
<td><img src="image" alt="Traceless Staudinger Ligation" /></td>
</tr>
<tr>
<td>Click, Cu-catalyzed</td>
<td><img src="image" alt="Click, Cu-catalyzed" /></td>
</tr>
<tr>
<td>Click, Strain promoted</td>
<td><img src="image" alt="Click, Strain promoted" /></td>
</tr>
<tr>
<td>Tetrazine ligation</td>
<td><img src="image" alt="Tetrazine Ligation" /></td>
</tr>
<tr>
<td>Oxime ligation</td>
<td><img src="image" alt="Oxime Ligation" /></td>
</tr>
</tbody>
</table>

In the classical Staudinger reaction, the ligation occurs in two steps between an azide and a phosphine, yielding first anaza-ylide intermediate followed by hydrolysis in water, producing an
primary amine and a phosphine oxide. The modified Staudinger ligation developed by Saxon & Bertozzi in 2000, traps the unstable aza-ylide intermediate by an adjacent carbonyl group, leading to an amine bond formation in water. In consequence of its bio-orthogonality under aqueous mild conditions, the modified Staudinger ligation has been intensively used for labeling or immobilization of proteins.

The copper catalyzed azide-alkyne cycloaddition (Huisgen cycloaddition), broadly defined as Click reaction, was re-invented by Tornoe and Meldal in 2001, when they recognized that copper catalyses the reaction strongly and introduces regio-selectivity. The reaction is very robust, quantitative, orthogonal and general, but needs exogenous copper, resulting in cytotoxic effects, limiting its use in living organism. Because of its problems regarding bio-compatibility, the copper free click reaction was developed by Bertozzi.

The Diels-Alder reaction ligates an electron rich diene and an electron deficient dienophile into an cyclohexen. The reaction kinetics and selectivity are thereby significantly enhanced in aqueous solutions, compared to organic solvents. The capability of the chemo-selective Diels-Alder reaction has been demonstrated for proteins on surfaces, for labeling or chemically activated proteins.

These approaches energized the scientific community to explore more and new reaction methodologies but reduced their attention regarding their limitations. The major restriction of all bio-orthogonal reactions is the requirement of azide or alkyne functionalities. To incorporate these artificial cross linkers, frequently traditional labeling chemistry is used, such as succinimidyl ester or thiol maleimide chemistry, techniques that do not necessarily allow site-specific labeling. Alternatively unnatural amino acids can be integrated into the protein, featuring the respective functionalities. Although these sophisticated techniques allow the bio-orthogonal labeling and immobilization of proteins, they were mainly applied in vitro or in- and outside cells. Only the radioactive labeling of tetrazine labeled antibodies was illustrated in mice by inversed Diels-Alder reaction. The generation of the tetrazine moiety on the antibodies was done by in vitro NHS-chemistry.

Protein fusion tags are an alternative to the chemical modification of proteins and several of these unique sequences have been developed, such as SNAP or tetracysteine tags. The SNAP-tag developed by Johnsson and co-workers is a modified version of the human O-alkylguanine-DNA alkytransferase (hAGT) that reacts rapidly with an O-Benzyguanine (BG) derivative, often equipped with optical or chemical functionalities. The SNAP-tag can be genetically engineered to the N- or C-terminus of proteins or covalently fused to other genetically specified motifs like tetracysteine tags. The tetracysteine tag consists of a 12 amino acid recognitions motif containing 4 cysteines. It binds biarsenical dyes with high affinity and must not be located at protein extremities, because for efficient binding only the two cysteine pairs have to be arranged in an optimal spacing and orientation. Therefore tetracysteine tags can be used as an alternative to terminal located big protein tags such as the described SNAP-tag or fluorescence proteins.

As outlined above, current bio-orthogonal methods have expanded substantially the range and the flexibility of protein labeling methods, but rely still on classical in vitro ligation techniques to
endogenous lysine or cysteine residues. Surface exposed nucleophilic amino acids such as lysines are popular targets for protein labeling because an enormous number of conjugation techniques are available, such as N-hydroxysuccinimidyl esters (NHS-esters), sulfonyl chlorides, isocyanates, isothiocyanates or reductive amination with aldehydes. The broad abundance of lysine residues in native proteins leads to high conjugation efficiencies, often beneficial for imaging or immobilization experiments. Among the available methods to cross link synthetic compounds to biological amino groups, NHS chemistry is the most popular and common used technique today. NHS ester activated cross-linkers react with amines under physiological conditions (aqueous environments, pH 7.0 – 9.0) to form stable and covalent amide bonds. The reaction rates are fast but reactions in water have to compete with NHS-ester hydrolysis. The inactivation by hydrolysis depends on temperature and pH; with a half life of 4-5 hours at pH 7.0 and 0°C, and a half life of 10 min at pH 8.6 and 4°C. Other versatile cross-linking techniques are using endogenous cysteines residues as ligation sites. As a relatively rare amino acid, cysteines can often be used for more site-selective labeling.

3.1.2 Cysteines, a natural site-specific ligation site

Cysteines, being less prevalent than lysines in proteins, are often used for site-specific labeling. Cysteines feature in their reduced state a free thiol group, the most reactive functional group in proteins, which enables orthogonal cross linking under physiological conditions to proteins via, for example, haloacetyl compounds, disulfides or maleimides. The most popular reaction is the Michael addition on maleimides, to form a stable thioether bond. Maleimides can also react towards deprotonated amines, but the reaction of maleimides with thiols is 1000 times faster at neutral pH compared to amines. Therefore orthogonal labeling of cysteines is possible in the presence of unprotected amines. In combination with its low abundance in proteins cysteines are an excellent target for site-selective labeling.

Cysteine oxidation competes with the labeling reaction by forming reversible disulfide bonds or irreversible sulfonic acid. Free thiol groups can be regenerated from disulfides by UV-light or more commonly by treatment with disulfide reducing agents such as dithiothreitol (DTT) or Tris-(2-carboxyethyl)-phosphine (TCEP). Because these are global approaches, influences on the biological activity can not be excluded. Treatment of proteins that do not contain any natural cysteines is unproblematic and by genetically engineering sulfhydryl groups into the protein, distal from the active site, site-specific ligation points can be generated.

To limit or avoid potential interference with the biological activity, the introduction of unique cysteines (by site-directed mutagenesis) or tags are mostly planned at the N- or C-terminus. Nevertheless, proteins modified at their extremities must be tested regarding their identical biological properties compared to the unmodified protein because changes in the polypeptide sequence can alter the protein conformation, activity or intracellular localization.
3.1.3 Cysteines and native chemical ligation (NCL)

N-terminal cysteines in a polypeptide feature a thiol as well as an amine functionality and can be used for native chemical ligation (NCL),\textsuperscript{2} a method established by Kent and Tam.\textsuperscript{43,44} NCL relies on the free accessibility of an N-terminal cysteine, because the reaction is only applicable between a thioester and a N-terminal cysteine, yielding a native peptide bond. The huge strength of NCL is its possibility to combine chemical peptide synthesis with protein expression, giving access to semi-synthetic proteins featuring any desired synthetic modification.\textsuperscript{45,46} The semi-synthesis of proteins, via so-called expressed protein ligation (EPL) can be applied to a broad class of proteins such as ion channels, polymerases, kinases, etc., while keeping their natural biological characteristics.\textsuperscript{47-51} The combined use of EPL and inteins, a protein self-cleavable element, eliminates the challenge to generate a free N-terminal cysteine, making conventional cleavage via endogenous methionyl aminopeptidases or exogenous proteases such as factor Xa or TEV unnecessary.\textsuperscript{46,52,54}

Table 3-2: Classical bio-conjugation methods to modify cysteine residues.\textsuperscript{3} If the N-terminal cysteine is protected as thiazolidine-ring, deprotection must occur prior to labeling / conjugation.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disulfide exchange</td>
<td><img src="image" alt="Disulfide Exchange" /></td>
</tr>
<tr>
<td>Alkylation</td>
<td><img src="image" alt="Alkylation" /></td>
</tr>
<tr>
<td>Michael addition</td>
<td><img src="image" alt="Michael Addition" /></td>
</tr>
<tr>
<td>Dehydroalanine + Michael</td>
<td><img src="image" alt="Dehydroalanine + Michael Addition" /></td>
</tr>
<tr>
<td>addition</td>
<td></td>
</tr>
<tr>
<td>Thiol-ene reaction</td>
<td><img src="image" alt="Thiol-ene Reaction" /></td>
</tr>
<tr>
<td>Native chemical ligation</td>
<td><img src="image" alt="Native Chemical Ligation" /></td>
</tr>
</tbody>
</table>
3.1.4 The thiazoline-ring, a cysteine protection group

Thiazolidine-ring protected N-terminal cysteines are recently more used in peptide synthesis which can be deprotected under mild acidic condition in the presence of O-methylhydroxylamine.\textsuperscript{59} The strategy combined with NCL was introduced by Kent and co-worker in 2004, illustrated by the one-pot synthesis of crambin,\textsuperscript{56} followed by the total synthesis of insulin-like growth factor (IGF)\textsuperscript{57} and human vascular endothelial growth factor (VEGF).\textsuperscript{58} The approach was facilitated with the commercial introduction of Boc-protected 1,3-thiazolidine-4-R-carboxylic acid, a building block easy to insert in the last step of peptide synthesis.\textsuperscript{55,56} The use of thiazolidines on N-terminal cysteines is not restricted to total peptide synthesis. To immobilize biomolecules on microarrays, thiazolidines were used to minimize orientation and interacting problems.\textsuperscript{59,60} On proteins, thiazolidines can be actively used for purification\textsuperscript{61} or labeling reactions.\textsuperscript{52}

The reversible protection of cysteines via thiazolidines offers an attractive possibility to introduce additional cysteine residues in proteins that feature already a natural high number of native cysteines, such as human EGF containing six cysteines folded via disulfide bridges. The reversible protection avoids any potential type of interference of the artificially introduced cysteine with the six native cysteines in hEGF such as disulfide shuffling, oxidation or dimerization, enabling the protein to establish its natural bio-active conformation. If the correct tertiary structure is accomplished, each thiol group is occupied by its natural partner. Liberation of the thiazolidine-ring then generates one unprotected thiol group per protein molecule, readily accessible to thiol reactive conjugation reagents. By using this concept an orthogonal ligation point is created at the N-terminus of the protein that enables site-specific bio-conjugation reactions by conventional thiol-maleimide chemistry or other reactive functionalities. Additionally the reversible protection of the N-terminal cysteine might enable the introduction of a second ligation site in the form of another unprotected cysteine at the C-terminus that can be selectively addressed. This approach was envisioned to be advantageous and to allow sequential and orthogonal double labeling of hEGF.

3.2 Results and Discussion

The different modified human epidermal growth factor (hEGF) constructs were expressed in \textit{E.coli} featuring N- or C-terminal cysteines as orthogonal ligation sites as described in the previous chapter. All constructs were obtained on a milligram scale and in excellent purity.

3.2.1 N-terminal cysteine protection with thiazolidines

N-terminal cysteine residues have been exploited in several highly successful protein modification methods and are a recurring phenomenon\textsuperscript{63} in the field of chemical biology. While optimizing the culture and refolding conditions of N- and C-terminal modified hEGF (chapter 2), a deprotection strategy of the N-terminal thiazolidine protected cysteine was investigated in more
detail. By providing hEGF with a N-terminal thiazolidine protecting group, which effectively prevents any unfavorable reaction with internal native hEGF cysteines, an opportunity was created to introduce another cysteine, unprotected at the C-terminus. If successful, it enables the sequential and site-selective double labeling of hEGF.

The analyses of all N-terminal cysteine containing hEGF constructs by mass spectrometry showed a mass addition of 12 Dalton; Cys0_hEGF mass expected: 6319.1, mass obtained: 6331.5; Cys0_hEGF_Cys54 mass expected: 6422.3, mass obtained 6435.4 (Figure 3-1). The mass addition of 12 Dalton corresponds to the protection of the N-terminal cysteine via thiazolidine-ring formation.

Figure 3-1: LC-MS traces of Cys0_hEGF and Cys0_hEGF_Cys54 featuring a protected N-terminal cysteine. **Top**: Cys0-hEGF, **Bottom**: Cys0_hEGF_Cys54. **Left**: MS Total Ion Count, **Right**: mass to charge ratio, top corners deconvoluted mass spectra. Expected mass protected / deprotected Cys0_hEGF 3x S-S: 6331.1 / 6319.1, observed mass: 6331.5; expected mass protected / deprotected Cys0_hEGF_Cys54 3x S-S: 6434.2 / 6422.2, observed mass: 6435.4.

To demonstrate the protection of the N-terminal cysteine in Cys0_hEGF, the protein was incubated with Alexa647-maleimide. If the N-terminal cysteine is unprotected, labeling of Cys0_hEGF would take place, if it is protected, no labeling would occur. The incubation with Alexa647 resulted in no labeling of Cys0_hEGF. To confirm N-terminal cysteine protection via
thiazolidine-ring formation, Cys0_hEGF was reduced, alkylated and digested with trypsin. Analysis by mass spectrometry revealed a mass addition of +12 Dalton in the N-terminal fragment (Figure 3-2), corresponding to a thiazolidine-ring. To deprotect the N-terminal cysteine in Cys0_hEGF, the protein was incubated with O-methylhydroxylamine, a thiazolidine-ring opening reagent. Analysis by mass spectrometry confirmed the removal of the protecting group, corresponding to the loss of 12 Dalton; Cys0_hEGF mass before deprotection: 6331.5, mass obtained after deprotection: 6319.5 (Figure 3-4).

<table>
<thead>
<tr>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys0_hEGF</td>
</tr>
<tr>
<td>Fragment 1</td>
</tr>
<tr>
<td>Fragment 2</td>
</tr>
<tr>
<td>Fragment 3</td>
</tr>
<tr>
<td>Fragment 4</td>
</tr>
<tr>
<td>Fragment 5</td>
</tr>
</tbody>
</table>

Figure 3-2: Trypsin digestion of Cys0-hEGF. **Top:** table with expected protein fragments after trypsin digestion. Cysteines are highlighted in bold. **Bottom:** LC-MS trace of alkylated and trypsinated Cys0_hEGF. **Left:** MS Total Ion Count, **Right:** mass to charge ratio, top corner deconvoluted mass spectrum. The N-terminal fragment is marked with an arrow. Expected mass of the 3x alkylated N-terminal Cys0_hEGF fragment: 3401.5; observed mass 3400.8.

The formation of N-terminal thiazolidine-rings on polypeptides results from the condensation reaction of the N-terminal cysteine with ketones or aldehydes. This can be occur intracellularly through metabolites such as pyruvate, or extracellularly by additives such as urea, which is prone to be contaminated with traces of aldehydes. Based on the expression of hEGF as an intein fusion construct in which the N-terminal cysteine residue of Cys0_hEGF is covalently linked to the intein via a native amide bond, intracellular thiazolidine protection can be excluded. Only upon intein cleavage Cys0_hEGF is released from the fusion construct and the N-terminal
cysteine is available for condensation reactions with aldehydes. Considering thiazolidine-ring formation after intein splicing, additives used shortly before and after intein cleavage were the most probable source of aldehyde. Suspecting urea, which was used in high concentrations for inclusion body solubilization, Enhanced Cyan Fluorescent Protein (ECFP) containing one free N-terminal cysteine (Cys0_ECFP) was incubated with urea in a test experiment. Analysis by mass spectrometry of the urea treated Cys0_ECFP revealed thiazolidine-ring protection of the N-terminal cysteine; expected mass thiazolidine ring protected 28455, observed mass 28455 (Figure 3-3). This result verified that small traces of impurities in urea, most probably formaldehyde, are causing the thiazolidine protection of Cys0_hEGF. The mass of the non-protected Cys0_ECFP was measured one Dalton higher as calculated, expected mass 28443, observed mass 28444 (Figure 3-3). This is a small inaccuracy of the measurement.

Figure 3-3: LC-MS-traces of Cys0_ECFP (top) and urea treated Cys0_ECFP (bottom). Left: MS Total Ion Count, Right: mass to charge ratio, top corners deconvoluted mass spectra. Expected mass Cys0_ECFP: 28443, observed mass: 28444; expected mass thiazolidine-ring protected Cys0_ECFP: 28455, observed mass 28455.
3.2.2 Deprotection of N-terminal cysteines

The formation of thiazolidines through ketones/aldehydes at the amino terminus of cysteine occurs in aqueous or semi-organic environments and can be actively used for total protein synthesis, purification or labeling reactions on proteins. The reaction itself is chemically reversed by applying O-methylhydroxylamine, an aminoxy containing compound, which reacts rapidly with ketones or aldehydes to form an oxime. The use of O-methylhydroxylamine as a competitor for aldehydes regenerates a free cysteine moiety, thus creating a reactive sulfhydryl group feasible for N-terminal site-specific labeling of a protein.

The incubation of Cys0_hEGF with O-methylhydroxylamine resulted in the removal of the thiazolidine group from the N-terminal cysteine, demonstrated by the loss of 12 Dalton; Cys0_hEGF mass protected: 6331.5, mass obtained after deprotection: 6319.5 (Figure 3-4). However, the deprotection of the N-terminal cysteines led to isomerization of Cys0_hEGF, illustrated by the two peaks in the LC trace with the same mass 6319.5 (Figure 3-4), which corresponds to the deprotected Cys0_hEGF. The isomerization was most probably due to intra-molecular disulfide shuffling, a thiol-disulfide interchange reaction caused by the liberation of the N-terminal sulfhydryl group. Furthermore, deprotection under aerobic conditions led to partial oxidation of the unprotected cysteine into a sulfinic- or sulfonic acid, indicated by the mass addition of +32 or +48 Dalton. The oxidation of the N-terminal cysteine should be avoided because it blocks irreversible the thiol functionality. Deprotection of the N-terminal cysteine under anaerobic conditions eliminated this problem.

Isomerization due to disulfide shuffling should be excluded too, because it results in a heterogeneous molecular product. Vijver & co-workers overcame this problem using a simultaneous deprotection and labeling strategy. Simultaneous deprotection and labeling blocked thereby the N-terminal cysteine thiol before it could undergo intra-molecular disulfide shuffling reactions. Inspired by this concept, Cyso_hEGF was deprotected and labeled with Alexa647 simultaneously. The approach eliminated the above described isomerization problem, illustrated by one single peak of Alexa647_Cys0_hEGF in the mass spectrum that corresponded to the mass of monovalent labeled Cys0_hEGF; mass expected: 7299.5, mass obtained 7300.5 (Figure 3-4).

No isomerization was observed for hEGF_Cys54. It was concluded that the second introduced unprotected cysteine at the C-terminus does not take part in disulfide exchange reactions with the internal hEGF cysteines. However, inter-molecular disulfide mediated dimerization of hEGF_Cys54 via the newly incorporated cysteine was observed, a tendency that could be suppressed under slightly acidic conditions.
Figure 3-4: LC-MS traces of Cys0_hEGF featuring one protected N-terminal cysteine. **Top:** N-terminal protected Cys0-hEGF, **Middle:** O-methylhydroxylamine deprotected Cys0_hEGF, **Bottom:** simultaneous deprotection and Alexa647 labeling of Cys0_hEGF. **Left:** MS Total Ion Count, **Right:** mass to charge ratio, top corners deconvoluted mass spectra. Expected mass of protected / deprotected Cys0_hEGF 3x S-S: 6331.1 / 6319.1, observed mass after deprotection for LC-trace peak 1 and 2: 6319.5; expected mass of N-terminal labeled Alexa647-Cys0_hEGF 3x S-S: 7299.5, observed mass: 7300.5.
3.2.3 Site-specific labeling of hEGF

The above deprotection and conjugation protocol was consequently applied to site-specifically label the N- and C- terminus of Cys0_hEGF or hEGF_Cys54, respectively. In case of Cys0_hEGF_Cys54 both protein termini were sequentially addressed that enabled controlled two-step conjugation of synthetic dyes to hEGF; following C-terminal labeling, the N-terminus was deprotected and labeled. All cross-linking reactions were performed on genetically introduced sulfhydryl groups in hEGF using standard thiol maleimide coupling reactions. Because the formation of a stable thioether bond between maleimide and sulfhydryl groups can be hindered through air oxidation, all reactions were carried out in degassed buffers and under argon atmosphere.

The N-terminus of Cys0_hEGF was labeled with DyLight549- or Alexa647-maleimide, respectively (Figure 3-5). To avoid disulfide shuffling of the disulfide bonds after deprotection, the removal of the N-terminal thiazolidine protecting group and the maleimide conjugation were performed simultaneously. Due to the acidic conditions necessary for thiazolidine deprotection, but unfavorable for maleimide coupling, the fluorophore attachment was carried out in excess of maleimide activated labeling reagents. Analyses by mass spectrometry revealed site-specific labeling of the N-terminus of Cys0_hEGF with Alexa647, indicated by the mass addition of 981.5 after labeling; mass Cys0_hEGF deprotected: 6319.5, mass observed 7300.5 (Figure 3-5, left). No unspecific side-reactions were observed by mass spectrometry, illustrating on one hand the selectivity of this labeling approach and on the other hand the correct folding of the six native cysteines of hEGF into disulfide bonds.

The C-terminus of hEGF_Cys54 containing an unprotected cysteine was labeled with Alexa647-maleimide. To create optimal conditions for maleimide activated labeling, the pH was adjusted into the neutral range (pH 7.5). Analyses by mass spectrometry revealed site-specific and monovalent labeling of the C-terminus of hEGF_Cys54 with Alexa647, indicated by the mass addition of 981.5 after labeling; mass hEGF_Cys54: 6319.5, mass observed 7300.5 (Figure 3-5, center). No multi-labeling of hEGF_Cys54 was observed, illustrating the selectivity of this labeling approach as well as the correct conformation of the hEGF_Cys54 and its six natural cysteines.

The stepwise double labeling of hEGF was performed on Cys0_hEGF_Cys54 containing a N-terminal thiazolidine protected cysteine and an unprotected C-terminal cysteine. In the first step the C-terminus of the protein was modified with Alexa647-maleimide. Following the removal of excess of Alexa647 dye by dialysis, the N-terminus was ligated to biotin-maleimide using the deprotection and labeling procedure outlined before for Cys0_hEGF. Analyses by mass spectrometry revealed a mass addition of 1432 to Cys0_hEGF_Cys54, corresponding to the two-fold labeling of Cys0_hEGF_Cys54 with one biotin and one Alexa647 (Figure 3-5, right). The C-terminus was site-specific labeled with Alexa647, the N-terminus of the protein was site-specific coupled to biotin.

The genetic introduction of one N-terminal thiazolidine protected cysteine as well as one C-
terminal unprotected cysteine in hEGF did not interfere with the six internal cysteines of the protein. The illustrated site-selective labeling of the genetically introduced cysteines makes this approach highly attractive for the production of cysteine rich proteins featuring one or two orthogonal ligation sites. The coupling via sulfhydryl groups is thereby highly favorable because of the enormous number of commercially available thiol reactive cross-linking reagents.

![Diagram](image)

Figure 3-5: LC-MS-traces of Alexa647-Cys0_hEGF (left), hEGF_Cys54-Alexa647 (middle) and Cys0_hEGF_Cys54 labeled N-terminal with Biotin and C-terminal with Alexa647 (right). Top: MS Total Ion Count, Bottom: UV-trace at 647nm. Expected mass Alexa647-Cys0_hEGF $3\times S-S$: 7299.5, observed mass 7300.5; expected mass hEGF_Cys54-Alexa647 $3\times S-S$: 7299.5, observed mass 7300.5; expected mass Biotin-Cys0_hEGF_Cys54-Alexa647 $3\times S-S$: 7854.7, observed mass 7855.1. All labeling reactions resulted in a dye / protein ratio of 1:1. No side reactions were observed.

### 3.2.4 Activity and Live Cell Imaging of site-specific labeled hEGF

To show the full activity of site-specifically labeled hEGF, its EGFR activation potential was measured on A431 cells, an EGFR over-expressing cell line. EGFR activation followed by hEGF binding results in “EGFR dimerization [and] activation of its intrinsic tyrosine kinase [that is] followed by rapid auto-phosphorylation of multiple tyrosine residues in the cytoplasmic EGFR domain”. Tyrosine phosphorylation leads, on the one hand, to a complex pattern of biochemical
messages which controls cell proliferation, differentiation etc., but can, on the other hand, be used to quantify the level of EGFR activation. Targeting the phosphorylation site of tyrosine 1173 with an anti-phospho-tyrosine antibody using a commercially available ELISA (see experimental), the phosphorylation status of EGFR following hEGF stimulation could be determined. By comparing the phosphorylation level of EGFR following binding of unmodified or site-specific labeled hEGF, conclusions can be drawn regarding its biological activity.

Incubation of Alexa647-Cys0_hEGF, hEGF_Cys54-Alexa647, Biotin-Cys0_hEGF_Cys54-Alexa647 and DyLight549-Cys0_hEGF with A431 cells revealed an EGFR phosphorylation pattern comparable to natural unmodified hEGF, except for Alexa647-Cys0_hEGF that showed a about 50% higher signal (Figure 3-6). However this might be an effect due to the small number of tested cell samples. The activity of all site-specifically modified hEGF proteins, whether labeled at the N- or C-terminus or on both termini, was equal to natural unmodified hEGF (Figure 3-6), confirming full biological activity of site-specifically labeled hEGF.

![Graph showing EGFR Activity in Percent](image)

Figure 3-6: Activity of Alexa647-Cys0_hEGF, hEGF_Cys54-Alexa647, Biotin-Cys0_hEGF_Cys54-Alexa647 and DyLight549-Cys0_hEGF in comparison to commercial unlabeled hEGF measured on A431 cells.

The activity of all site-specifically labeled hEGF proteins was additionally proven by confocal live cell microscopy. HeLa cells were transfected with EGFR-mCitrine or EGFR-mCherry, respectively. Following incubation with DyLight549-Cys0_hEGF, Alexa647-Cys0_hEGF, hEGF_Cys54-Alexa647 or Biotin-Cys0_hEGF_Cys54-Alexa647, live cell confocal microscopy was performed. After stimulating HeLa cells transfected with EGFR-mCitrine for 15 min, DyLight549-Cys0_hEGF was detected on the plasma membrane and in intracellular vesicles (Figure 3-7). When the mCitrine and DyLight549 image were merged into a single image, co-localization of ligand and receptor was revealed (Figure 3-7). This result confirmed binding of DyLight549-Cys0_hEGF to EGFR-mCitrine that triggered active internalization of the EGF-EGFR complex into the cells by endocytosis.

Incubating Alexa647-Cys0_hEGF, hEGF_Cys54-Alexa647 and Biotin-Cys0_hEGF_Cys54-Alexa647 with EGFR-mCherry transfected HeLa cells for 5 min, resulted in the detection of all
modified hEGF variants mainly on the plasma membrane (Figure 3-7). But also slight internalization was observed, but not as strongly as for DyLight549-Cys0_hEGF due to the shorter stimulation with hEGF. The result confirmed the binding of all three modified hEGF proteins to EGFR and reconfirmed the results of the EGFR activity assay from 3.2.4.

Figure 3-7: A) HeLa cells transfected with EGFR-mCitrine and incubated with DyLight549_Cys0_hEGF for 15 min, Left: mCitrine channel; Middle: DyLight549 channel; Right: overlay. B) HeLa cells transfected with EGFR-mCherry and incubated for 5 min with Alexa647-Cys0_hEGF (left); hEGF_Cys54-Alexa647 (middle), Biotin-Cys0_hEGF_Cys54-Alexa647 (right).

3.2.5 A streptavidin platform generating mono- and multivalent hEGF constructs

Streptavidin is a tetrameric protein of 56 kDa that binds biotin with femtomolar affinity (Kd $10^{-15}$ to $10^{-14}$ M), an affinity higher than any other known non-covalent interaction in
nature. This unique property of tight and specific binding in combination with its high stability regarding temperature, denaturing reagents and organic solvents made streptavidin a versatile and powerful tool in life science. The wide use of streptavidin in patterning, labeling and purification resulted in the development of even stronger and more stable streptavidin variants such as trapavidin. However the creation of streptavidin mutants containing a different number of biotin binding sites was hampered by the fact, that each of the four biotin binding sites in wild-type streptavidin are located in the interface between two subunits, in which Trp-120 of one subunit makes contact with the biotin bound by an neighboring subunit. Due to the contribution of one subunit to the biotin binding site of the adjacent subunit, the disruption of the tetrameric structure resulted in a strong decrease in biotin affinity (to $10^{-8}$ or $10^{-9}$ M for avidin). In 2006 Howarth et al. first reported the development of monovalent streptavidin with femtomolar biotin binding affinity. The engineered streptavidin tetramer contained only one biotin binding subunit, while the characteristics of the natural streptavidin remained. The approach introduced 3 mutations in the biotin binding site of a streptavidin monomer (N23A, S27D, S45A) (named Dead subunit), reducing its biotin affinity to $10^{-3}$ M, while the tetrameric structure was kept. By mixing wild-type subunits (named Alive subunit) with Dead subunits in different ratios, monovalent, divalent and trivalent tetrameric streptavidin was obtained with one, two or three active biotin binding sites. The different streptavidin compositions were purified from each other using the His-tag at the C-terminus of the wild-type subunit (Alive) and nickel affinity chromatography. The different number of affinity His-tags per tetramer allowed thereby the separation of the different streptavidin variants (Figure 3-8).

![Streptavidin structure and expression purification route](image)

**Figure 3-8:** Left: Streptavidin structure and expression / purification route. Wild-type streptavidin is a tetrameric molecule featuring four Alive subunits (blue boxes) for binding four biotin molecules. Monovalent streptavidin contains three Dead subunits and one Alive subunit for biotin binding. Divalent streptavidin contains two Dead subunits and two Alive subunits for biotin binding. Trivalent streptavidin has a ratio of 3:1 between Alive and Dead subunits. Right: Semi-native SDS-PAGE analysis illustrating the different streptavidin tetramers; lane 1: dead streptavidin, lane 2: monovalent streptavidin, lane 3: divalent streptavidin, lane 4: trivalent streptavidin. The size difference between the Dead and Alive monomer is 854 Da. The streptavidin tetramers are purified by nickel affinity column.
The expression and purification of the different streptavidin variants on a multi-milligram scale was carried out according to Howarth et al.\textsuperscript{79} The purity of all expressed constructs (mono- and divalent streptavidin) was confirmed by semi-native SDS-PAGE (Figure 3-8). To test the number of biotin binding sites per streptavidin molecule, mono- and divalent streptavidin were incubated with a Biotin functionalized peptide (Biotin-LTERHKILHRLQEGSPD-NH\textsubscript{2}, size 2.455 kDa). The SDS-PAGE analyses revealed a single or double gel shift of mono- and divalent streptavidin, respectively (Figure 3-9). The reversed gel shift direction observed resulted most probably from the higher negative charge of the peptide-streptavidin complex in comparison to the unoccupied streptavidin, since semi-native SDS-gel electrophoresis separates proteins regarding both size and charge.

To further verify the number of biotin binding sites per streptavidin variant, mono-, di- and trivalent streptavidin were incubated with and without biotin and analyzed by mass spectrometry. Each biotin molecule bound to streptavidin increases the mass of the protein complex by 244 Dalton. Analyses of the mass spectrometry data from mono-, di- and trivalent streptavidin incubated with biotin revealed a mass addition of one, two or three biotin molecules per mono-, di- and trivalent streptavidin molecule, respectively (Table 3-3). These results confirmed the mono-, di- and trivalence of each streptavidin construct and were in line with the results of the semi-native SDS-PAGE analysis (Figure 3-9) as well as from Howarth et al.\textsuperscript{79}

![Figure 3-9: Semi-native SDS-PAGE of mono- and divalent streptavidin incubated with and without Biotin-LTERHKILHRLQEGSPD-NH\textsubscript{2}. The gel shift is not upwards as expected, but downwards, most probably due to the higher negative charge of the overall complex.](image)

Table 3-3: Predicted and measured mass of different streptavidin tetramers with and without biotin. The mass of one biotin molecule is 244 Dalton. The results confirmed the mono-, di- and trivalence of each streptavidin construct.

<table>
<thead>
<tr>
<th>Streptavidin tetramer</th>
<th>Predicted mass</th>
<th>Observed mass (without biotin)</th>
<th>Observed mass (with biotin)</th>
<th>Change in mass</th>
<th>Number of bound biotins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monovalent</td>
<td>53816</td>
<td>53814</td>
<td>54060</td>
<td>246</td>
<td>1</td>
</tr>
<tr>
<td>Divalent</td>
<td>54669</td>
<td>54669</td>
<td>55157</td>
<td>488</td>
<td>2</td>
</tr>
<tr>
<td>Trivalent</td>
<td>55523</td>
<td>55523</td>
<td>56257</td>
<td>734</td>
<td>3</td>
</tr>
</tbody>
</table>
Protein assemblies such as the dimerization of trans-membrane receptor EGFR are playing a key role in biological processes and can change their signaling function. Cross-linking of membrane receptors influences their dynamic behavior within the plasma membrane and receptor clustering can facilitate the auto-phosphorylation of their cytoplasmic tyrosine domains. Using the above described streptavidin variants (mono-, di-, and trivalent) for the presentation of a distinctive number of hEGF ligands, a platform is created providing simple entry for EGF receptor cross-linking and co-localization experiments (Figure 3-10). Having both building blocks in hand, streptavidin with different biotin valencies and commercial hEGF containing one N-terminal biotin moiety, the generation of streptavidin molecules was possible featuring one, two, three or four hEGF ligands per streptavidin molecule (Figure 3-10).

Consequently, the four different streptavidin proteins (mono-, di-, tri- and tetravalent) were mixed with commercial biotin-hEGF, to create streptavidin constructs featuring one, two, three or four hEGF ligands per molecule. Following removal of the excess of biotin-hEGF by dialysis, all four streptavidin / hEGF constructs were incubated on A431 cells, an EGF over-expressing cell line. The incubation of the different streptavidin / hEGF constructs revealed no difference in EGFR activation pattern (Figure 3-11). Furthermore, it was noticed that an EGFR activation signal could be only obtained when using high concentrations of streptavidin / hEGF; 25 times higher for streptavidin / hEGF and 10 times higher for free Biotin-hEGF in comparison to natural unmodified hEGF. This was surprising because expressed hEGF showed an identical EGFR activation behavior as natural unmodified hEGF. Considering these results, the EGFR activation itself or the readout of the ELISA assay must be hampered when using the streptavidin / hEGF constructs. Focusing on commercial biotin-hEGF, used for synthesis of all streptavidin / hEGF constructs, it was discovered that it is stabilized with BSA. Bovine serum albumin (BSA) is a well-known blocking reagent widely used in ELISA assays. The blocking of the readout signal by BSA is very likely and would explain the high concentration of streptavidin / hEGF necessary to obtain a readout signal. To test this assumption, BSA should be removed from biotin-hEGF prior streptavidin / hEGF coupling by size-exclusion chromatography (SEC) or should be generated via the protocol developed in chapters 2 and 3.

![Figure 3-10: Schematic representation of EGFR cross-linking via different streptavidin / hEGF constructs. Tetravalent wild-type streptavidin contains four hEGF molecules resulting in EGFR cross-linking. Monovalent streptavidin contains one hEGF molecule and results in no EGFR cross-linking.](image)
Chapter 3

Figure 3-11: EGF receptor phosphorylation induced by mono (MS)-, di (DS)-, tri (TriS)- and tetravalent (TS) streptavidin / hEGF constructs measured on A431 cells. The concentrations applied were 10-25 times higher in comparison to unmodified hEGF. The phosphorylation efficiency was independent from the applied streptavidin / hEGF constructs.

3.3 Conclusions

In conclusion, a small library of cysteine rich hEGF constructs containing orthogonal ligation sites were expressed, and following refolding in its bio-active conformation, site-selectively modified with different functionalities, such as, fluorescent dyes and biotin. The insertion of one N-terminal thiazolidine protected cysteine in hEGF enabled the integration of another additional reactive cysteine in hEGF C-terminal, consequently used for stepwise conjugation of synthetic molecules to the protein. Both insertions resulted in fully bio-active hEGF indicating that the cysteine insertion approach did not interfere with the formation of the native disulfide bonds in hEGF. Deprotection of the N-terminal thiazolidine protected cysteine resulted in hEGF isomerization due to intra-molecular disulfide shuffling, prevented by simultaneous deprotection and labeling of the N-terminal cysteine. Performing all bio-conjugation reactions under anaerobic conditions, irreversible oxidation of the unprotected cysteines to sulfenic- or sulfonic acid was avoided. EGFR activation experiments, EGFR phosphorylation assay and live cell EGFR activation, confirmed the functional activity of the generated hEGF conjugates.

Using commercial N-terminal biotinylated hEGF and streptavidin molecules featuring a different number of biotin binding sites, initial EGFR cross-linking and co-localization experiments were performed. Although the readout was probably hampered by additives in the applied commercial biotin-hEGF, an alternative is provided through the expression and labeling protocol here developed. This example demonstrates the potential use of site-specific conjugation strategies and illustrates possible applicability of the thiazolidine protected cysteine insertion approach for generating homologous products containing hEGF ligands with controlled geometry and density.
3.4 Experimental

Materials and methods
All materials and compounds not mentioned below have been purchased as described in chapter 2. Plasmids for the ‘dead’- and ‘alive’-subunit were obtained from A. Y. Ting (MIT) and monovalent, divalent and trivalent streptavidin was expressed according to a published protocol by Alexander Colditz. Tetravalent streptavidin, biotin, biotin-maleimide, O-methylhydroxylamine, trypsin, dithiothreitol (DTT), iodoacetamide and sodium acetate (NaCH₃COOH) were obtained from Sigma. DyLight546-maleimide was purchased from Thermo Scientific. AlexaFluor647®C2-maleimide and BiotinXX-EGF conjugate were obtained from Invitrogen. Formaldehyde and dimethylsulfoxide (DMSO) were obtained from Acros. Dimethylformamide (DMF) was purchased from Biosolve B.V.. Commercial hEGF was obtained from Jena Bioscience. Wencke Adriaens synthesized the peptide Biotin-LTERHKILHRLLQEGSPSD-NH₂.

Trypsin Digestion:
About 30 µg Cys0_hEGF has been reduced with 10 mM DTT for 1 hour at RT in the presence of 1 M GndHCl. The denatured and reduced Cys0_hEGF was alkylated with 50 mM iodoacetamide for 15 min at room temperature and consequently digested with 1.5 ng trypsin. The resulting peptide fragments were analyzed by LC-MS.

Deprotection and Coupling:
N-terminal cysteine:
The deprotection and labeling of the N-terminal cysteine was carried out simultaneously. About 200 µg Cys0_hEGF or other N-terminal conjugates have been mixed with 1 M degassed sodium acetate (pH 4.5) containing 250 mM O-methyl-hydroxylamine and 5 µl Biotin-maleimide (10 mg / ml) (DyLight546 maleimide) dissolved in DMSO. The reaction mixture (final pH 4.6) was incubated under argon for 4 hours at 37°C. The excess of Biotin-maleimide was removed through dialysis. The final product was analyzed by mass spectrometry.

C-terminal cysteine:
To label the unprotected C-terminal cysteine, a 5 times excess of AlexaFluor647®C2 maleimide dissolved in DMF was mixed with hEGF_Cys54 or other C-terminal conjugates. All reactions were performed in degassed aqueous solutions (PBS, pH 7.4) under Argon at 4°C in the dark for 2 hours or overnight. The excess of dye was removed via dialysis and the product was analyzed by mass spectrometry.

EGF activity assay:
The EGF activity assay was performed as described in chapter 2.

Transfection and confocal imaging:
The activity of labeled hEGF was additionally proven by live cell fluorescent microscopy. HeLa cells (gift from Moniek de Liefde, Technical University Eindhoven, The Netherlands) were seeded in 8-well glass bottom dishes (LabTek) (media: DMEN + 10% FBS + 1% P/S, Invitrogen) and transfected with EGFR-mCitrine according to manufacture protocol with Fugene 6 (Roche). 24 hours after transfection cells were serum starved overnight. The next day cells were imaged on a Leica TCS SP5 AOBS equipped with an HCX PL APO ×60/1.4 NA oil immersion lens and a temperature-controlled incubation chamber maintained at
Chapter 3

37°C and 5% CO₂. mCitrine was excited using the 514 nm Ar laser line; DyLight549 was excited using the 561 nm DSSP laser line. Fluorophore emission bands were detected in the following range: mCitrine, 520–560 nm; DyLight549, 580–650 nm. DyLight549-Cys0[hEGF] was added at 100 ng / ml in imaging medium and cells were observed 15 minutes (image taken every 15 seconds). Transfection and imaging was carried out by Jenny Ibach.

SNAP-EGFR-mCherry and Alexa647-hEGF constructs were imaged on a Leica TCS SP5 AOB5 equipped with an HCX PL APO CS x63/1.2 NA water immersion lens and a temperature-controlled incubation chamber maintained at 37°C. Both fluorophors, mCherry and Alexa647, were excited with a white light laser; mCherry at 585 nm, Alexa647 at 647 nm. Fluorophore emission bands were detected in the following ranges: mCherry, 600–635 nm, Alexa647, 670–730 nm.

Construction of SNAP-ErbB1-mCherry:
The plasmid SNAP-ErbB1-mCherry was provided by Jenny Ibach.

Amino acid sequence of SNAP-ErbB1-mCherry:
-SNAP-tag: underlined sequence
-ErbB1: bold sequence
-mCherry: italic sequence

MDKDCEMKRTTLDSPGLKELSGCEQGLHEIKLLGKTSADAADVVEPAPAAVLGGPEPMLMQATAWLNAYFHQPEAEIEFPYPALHHPVFQOQESFTQVLWKL
LKVVKFEGVISYQQALALAGNPAATAAVKTALSGNVPIPLPCHRVSSSGAVGGYEGGLAVKEWLALLAHEGHRGLKPGLGARGGNGLEEEKVQGTSNKL
TQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVRNYDLSFLKTIQEVAGY
VLALNVTIREPLENLQIRGNMYENSYALAVLSNYDANKTGLKELPMRN
LQELILHGAVRFSPNAPCNVESIQWRDIVSDFLSNMSMDFQNLHSG
QKCDPSCPNGSCWGAGEEENCQKLTIICAQQCSGRGKRKGSQDCCHNQC
AAGCTGPLRESDCLVCRKRDEATCKDTCPLMLYNPTTYQMDVNEPQKYSFGATBTCVKKPRNYVTDHGSVCRACGADSYMEDEGRVKCKCEGPC
RKVCNNGIGIGEFKDLSINATNIHKFKNCTSISGDNLHILPVAFRGDSFTHT
PPLDPQELDIILKTVEITGFLLIQAWPENRTDLHAFNELIIRGRTKQHGQ
FSLAVVSLNITSGLRLKLEISDGDIISGNKNCYANTINWKLFGTSQ
KTIIIISNRSGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRGEC
VDCCCKALLEPREFVENSECICQECHPECLPQAMNTCTGRGPDNCIQCAHY
IDGPHCVKTCAGVMGENNTLWVKYADAGHVCHLHPNTYCSTGPQG
EGCPTNPGPKIPSATGMVGALLLLLVALGIGLMFMMRRRHIVRKRTRLRL
QRELVEPLTSGAPQNALLRIKETFKKIKVLGSGAFGTVYKGLWIPE
GEKVKIPVAIKELREATSPDKHEDEAYVMSAVNHPVCRLLGICLST
VQLITQLMPFGCLLYDLYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRLLV
HRDLAARNVLVKTPQHVKITDFGLAKLLGGAEKEYHAEGGKVPKWMAL
ESILHRITHQDSVWYSVTWELMTFSKPYDGIPASEISILEKGERLP
QPPICITIDYVMIMVKCWMDADSRPKFRELIIIEFSKMARDPQRYLVIGD
ERMHLPSPTDNSYRALMEDDEMDDDVVDADEYLIPQGGFSSPSTRTPL
LSSLCSATSNSTVACIDNRGLQSCPIKESFLQRYSDPTGALTEDSIDDT

66
Site-specific labeling of human Epidermal Growth Factor (hEGF)

FLPVPYEYNQSVPKRPAGSVQPVPNHYNHQQLNPAPSRDPHPYSTAVGNPEYLTNVTQPTCVNSTFDSPAHWDAQKGSQISLDPDYQDDFFPKEAKPNFIGKGSTAEANAYLVRAPQSEFIGAAGGGGGGTMVSKGEDNMAIKEFMRFKVMHMEGSVNGHEEEIEEGEGRYPTQTKLKVTKGGPLFWDILSPQFMYGSKAYVKHPADIPYLKLSPFEGFKEWQMVNFEDGGVVTQDSSLQDGFIYKVKLRGTNPSDGPVMQKKTMGWEASSERMYPDEGALKGEIQRKLKDGGHYDAEVTTTYKAKKPVQLPGAYNVN1KLDITSNEDYTYEQQYERAEGRHSTGGMDELYK

Construction of CBD-Intein-Cys0_ECFP:
The plasmid was obtained from Hoang D. Nguyen.

Amino acid sequence of CBD-Intein-Cys0_ECFP:
- Chitin binding domain (CBD): underlined sequence
- Intein domain: italic sequence
- ECFP: bold sequence
- in brackets: N-terminal cysteine

MKIEEGKLTNPQSVASWQVNTAYTQGNYKTYKCLQPHTSLAGWEP SNVPALWQLQNGNNGLELRESGAISSGDSLISLASTGKRVSIKDLLDKEKFEIWAINEQTMKLESAKVRFCPTKLLVYLKTRLGRITKATANHRELTDGKWRLDLSLKEIALPRKLESSLQLSPEIEKLSQSDIYWDISIVSITETGVEEVFDLTVPGPHNFVANDIIHNC[G]GGASHHHHHHHHSMVSKGEELTGVPVILVEDGDNVGHKFSVSGEGGDATYGKLTLKFICTTGKLPVPWTLLVTTLTWGVQCFSRYPDHMKQHDFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGettyLvniELKIDFKEDGNIKGKLEYNYISHNXYITADKQKNGIANKFIRHNIEDGVSVLADHYQQNTPIGDGPVLPDNPYNLTQSLSKDNPENKRDMVLLLEFVTAAGITLGMDELYK

Expression of Cys0_ECFP:
Protein Expression and Purification of Cys0-ECFP: The expression and purification was carried out as described by Nguyen.85

Incubation of Cys0_ECFP with urea:
Enhanced Cyan Fluorescent Protein (ECFP) containing one N-terminal cysteine (Cys0_ECFP) were incubated in the presence of 3.6 M urea for 96 hours at 4°C. In the control experiment Cys0_ECFP were treated with 0.5% formaldehyde for 30 hours at room temperature.

Construction of Dead and Alive streptavidin:
The plasmids Dead and Alive encoding for streptavidin were obtain from Howarth and Ting (MIT).79

Amino acid sequence of Dead and Alive streptavidin monomers:
- underlined sequence : His-tag
- bold residues: mutated amino acids in the Dead subunit to disturbed Biotin binding
**Alive monomer:**
MAEAGITGWYNQFGSTFIVTAGADGALTGYEASAVGNAERYSVTGRDPSAPADGSATLGWTVAWKNNYRNASATTWSQYVGGAEARINTQWLLTSGTTEANAWKSTLGVHDFTFKVPSAASHHHHHH

**Dead monomer:**
MAEAGITGWYALQGDSTFIVTAGADGALTGYEAVGNAERYSVTGRDPSAPADGSATLGWTVAWKNNYRNASATTWSQYVGGAEARINTQWLLTSGTTEANAWKSTLGVHDFTFKVPSAAS

**Expression of mono-, di- and trivalent streptavidin:**
The expression of all streptavidin variants was carried out according to the protocol from Howarth and Ting,\(^7\) by Alexander Colditz.

**Streptavidin / hEGF constructs:**
To generate the different streptavidin / hEGF constructs, commercial BiotinXX- EGF conjugate was mixed with mono-, di-, tri- and tetravalent (wild-type) streptavidin in a molar ratio 1:1:1 and incubated for 2 hours on ice. The excess of BiotinXX- EGF conjugate was removed via dialysis.

**SDS-PAGE:**
SDS-PAGE gel electrophoresis was performed as described in chapter 2.

**Semi-native SDS-PAGE:**
Semi-native SDS-PAGE electrophoresis was performed on a Mini-PROTEAN 3 electrophoresis system (Biorad, Hercules, California). The gel consisted of a 5% stacking gel and 8% separating gel, respectively. The running buffer contained 25 mM Tris-HCl, 250 mM glycine, and 0.1% (w/v) SDS in H\(_2\)O. Prior loading onto the SDS-gel, all protein samples have been mixed with sample buffer (100 mM Tris-HCl, 20% (v/v) glycerol, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, pH 6.8 in Millipore H\(_2\)O) in a ratio 1:3. To avoid denaturing of the proteins, all samples were not exposed to reducing agents (β-mercaptoethanol) or heating up. The electrophoresis was run at room temperature at 80 V for 20 min and then at 140-160 V for 60 min. The protein bands were stained with Coomassie Brilliant Blue.

### 3.5 References

(20) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. Angew Chem Int Ed Engl 2002, 41, 2596-9.
Expression and labeling of site-specific modified growth factors and nano-bodies

Abstract. In the previous chapters, the development of an expression and purification protocol was established for the site-specific labeling of the cysteine rich protein, hEGF. Here, the boundaries of this established approach were tested by expressing a double cysteine containing fusion tag at the N-terminus of hEGF. The expression resulted in good yields, but the modified protein was unprotected and therefore unfeasible for orthogonal labeling reaction. The previously developed method was further applied to several proteins with different numbers of native cysteines to investigate the global applicability of the approach. TGFα, a member of the epidermal growth factor family, was expressed with additional N- or C-terminal ligation points, but EGFR binding activity was not observed. The expression of the EGFR binding but not activating nano-body EGb4 was achieved in good yields. Site-specific C-terminal labeling resulted in monovalent-labeled protein with equivalent bio-functional characteristics to unmodified nano-bodies.
4.1 Introduction

4.1.1 Growth factors and their ligands

The protocol developed in the previous chapter for the expression of hEGF featuring orthogonal ligation sites could be an excellent platform for the production of a variety of modified growth factors or cysteine rich proteins displaying site-specific ligation points. The proposed strategy would avoid the generation of heterogeneously labeled “products with unknown distribution and functional characteristics”\(^1\) by simultaneously keeping their biological activity. Expected that the developed protocol will provide entry in a standardized approach to couple imaging or therapeutic agents to various proteins, the expression and labeling strategy was applied to further ligands of the epidermal growth factor (EGF) family: the transforming growth factor α (TGFα) as well as the anti-epidermal growth factor receptor (EGFR) nano-body EGb4. Furthermore the developed protocol was reviewed on hEGF molecules containing two extra cysteines at the N-terminus. Inserting such a double cysteine containing fusion tag at the N-terminus of hEGF, in which one cysteine is protected as a thiazolidine-ring and one displaying its natural thiol functionality, two orthogonal ligation sites at the N-terminus of the protein are created, allowing sequential and site-specific double labeling without affecting the protein biofunctionality.

In mammalian organisms, the activation of the four closely related tyrosine kinase receptors [ErbB1 (EGFR/HER1), ErbB2 (HER2/Neu), ErbB3 (HER3) and ErbB4 (HER4)]\(^2\) is regulated by at least 7 distinct ligands: epidermal growth factor (EGF), transforming growth factor α (TGFα), heparin-binding EGF-like growth factor (HBEGF), amphiregulin (AREG), epieregulin (EREG), epigen (EPGN), betacellulin (BTC)\(^3\) and several neuregulin isotypes.\(^4\) The EGFR ligands themselves are synthesized as cell surface pre-cursors (pro-form).\(^5\) The pro-form encodes for a hydrophilic N-terminal extension, an EGF domain and a C-terminal cytoplasmic region, also known as cytoplasmic tail (Figure 4-1).\(^6,7\) The EGF pre-cursor is thereby unique, featuring 9 EGF motifs, but only the EGF module adjacent to the plasma membrane is functional as EGFR binder (Figure 4-1).\(^8\) Via proteolytic cleavage by metallo-proteases, the mature soluble growth factor is released, a process called ectodomain shedding. This essential step enables cells to control the amount of free ligand available for EGFR activation and gene transcription.\(^9,10\) The shedding process therefore plays an important biological role in normal and malignant cells.\(^11,12\) However, various studies demonstrated that the remaining cytoplasmic tail (Figure 4-1) has a potential role in gene silencing. For example, the cytoplasmic pro-HBEGF domain was identified in the inner nuclear membrane, a region in which silent genes were identified.\(^12,13\) A potential interaction partner for pro-HBEGF is PLZF that has been proven to decreased transcriptional activity of their target genes.\(^12,14\) These results indicate that growth factor ligands have a dual modulation function on the gene expression, gene activation and gene silencing.
Ligands that bind to the extracellular domain of the ErbB receptors initiate conformational rearrangements, leading to the exposition of the dimerization interface and subsequently dimer formation with other receptors (Figure 4-2). The dimerization can occur between two different ErbB receptors (heterodimers) or between two receptors of the same type (homodimers).\textsuperscript{15,16} The EGF-like growth factors that bind to the ErbB receptors are classified into four groups, regarding their binding characteristic for particular ErbB receptors (Figure 4-3). The first group consists of EGF, TGF\textalpha\ and AREG binding to EGFR selectively, the second group is composed of BTC, HBEGF and EREG, which have dual specificity to EGFR, and ErbB4 and the third and fourth group is containing the neuregulins. The neuregulins are divided into two subgroups, classified based on their ability to bind ErbB3 and ErbB4 (NRG1 and 2) or only ErbB4 (NRG3 to 5). No direct ligand for ErbB2 is known to date, but ErbB2 is the “preferred dimerization partner of all other ErbB receptors” (Figure 4-3).\textsuperscript{4,10}

![Diagram of EGFR activation and dimerization](image)

Figure 4-2: Mode of EGFR activation and dimerization. The tethered EGFR structure with four extracellular domains, I to IV. It is proposed that EGF binds to domain I and III, which adopts to the conformation of domain II, exposing the dimerization arm and promoting receptor dimerization.\textsuperscript{10,17}
Figure 4-3: ErbB receptor dimer pairs with distinct ligand binding groups. Ligand binding induces homo- or hetero-dimerization the EGFR receptor.\textsuperscript{4,10} The numbers refer to the respective EGFR receptors ErbB1 – 4.

The amino acid identity between all EGFR ligands is very low (about 25%),\textsuperscript{3} but the central structural element responsible for EGFR binding and activation is conserved. All members feature one or several units of a 35-40 amino acid sequence motif containing six cysteines, forming three intra-molecular disulfide pairs.\textsuperscript{6,18} The conserved structure creates three loops for high affinity binding to the ErbB receptors (Figure 4-4).\textsuperscript{3,19} Due to the structural relationship of all ligands, the developed expression platform for hEGF should be highly applicable for other ligands of the ErbB receptor family. Subsequently the concept was investigated for TGFα and the nano-body EGb4 (Table 4-1). Also further modifications on hEGF were explored regarding a double cysteine fusion tag (Table 4-1).

Figure 4-4: The secondary structure of the human EGF (hEGF). The cysteines are forming three disulfide bridges (orange) generating the characteristic 3 EGF loops.
Table 4-1: Overview of modified proteins explored in this chapter. Cysteine residues, thiazolidines and disulfide bridges are highlighted in orange.

<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys-Gly-Cys0_hEGF</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>Cys0_TGFα</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>TGFα_Cys51</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

4.2 Results and Discussion

4.2.1 Expression and purification of Cys-Gly-Cys0_hEGF

The strategy to produce hEGF featuring an N-terminal double cysteine fusion tag was identical to the expression and purification approach 2, described in chapter 2. The ligation side, in the form of two cysteines separated by one glycine, was introduced in the CBD-Intein-Cys0_hEGF construct via site-directed mutagenesis, resulting in the fusion construct CBD-Intein-Cys-Gly-Cys0_hEGF. Following intein cleavage, Cys-Gly-Cys0_hEGF is available for sequential N-terminal double labeling (Figure 4-5).

![Image](image4.png)

Figure 4-5: Cys-Gly-Cys0_hEGF featuring a N-terminal double cysteine fusion tag for site-specific two-step labeling. The first N-terminal cysteine is protected as thiazolidine-ring, the second cysteine is unprotected. After direct site-specific labeling of the unprotected cysteine, orthogonal labeling of the N-terminal cysteine is envisaged after deprotection.
Having the fusion construct in hand, the successful expression and purification under denaturing conditions was carried out, yielding pure Cys-Gly-Cys0_hEGF (about 1 mg per liter LB-media) as confirmed by SDS-gel electrophoresis (elution fraction Figure 4-6). The product was further analyzed by mass spectrometry, reconfirming the results of the SDS gel in respect to purity (Figure 4-6). However, the deconvoluted mass did not correspond with the calculated mass of Cys-Gly-Cys0_hEGF, containing one cysteine with an unprotected thiol functionality and one thiazolidine-ring protected cysteine at the N-terminus. The measured mass 6479.2 (Figure 4-6) was 12 Dalton smaller than the expected molecular weight, reflecting a Cys-Gly-Cys0_hEGF molecule with 6 oxidized cysteines, one reduced cysteine and no thiazolidine-ring (mass 12) at the N-terminus. To analyze if the N-terminal cysteine could be still protected to a thiazolidine-ring, Cys-Gly-Cys0_hEGF was incubated in the presence of formaldehyde. However, no mass addition by +12 was detected. This result was surprising but suggested either the inaccessibility of the N-terminal cysteine caused by incorrect folding of the protein or mis-directed intra-molecular disulfide bonding blocking the N-terminal thiol functionality. Both assumptions indicated unfavorable refolding conditions or interference of the double cysteine containing fusion tag with the native disulfide formation. To express recombinant Cys-Gly-Cys0_hEGF protein in its functionally active conformation, optimization of the protein folding conditions or the linker length of the double cysteine containing fusion tag is recommended.

Refolding of denatured proteins depend mostly on the maintenance of a “reasonable pace of disulfide bond shuffling, so that the native set of disulfide bonds is established”. Conventionally a mixture of reduced and oxidized glutathione is used to allow reversible disulfide formation between glutathione and cysteines, which drives the protein into its thermodynamically favorable and biologically active conformation. Improving the redox-environment by using redox pairs such as reduced and oxidized glutathione or cysteine / cystine could lead to refolding conditions resulting in the formation of correct and biologically active Cys-Gly-Cys0_hEGF.

On the other hand, misfolding of the Cys-Gly-Cys0_hEGF might be caused by the introduction of double cysteine containing fusion tag itself, which probably interferes with the disulfide formation of the native hEGF cysteines. Lacking a natural partner for disulfide formation, the spacer length between the two cysteines within the fusion tag or to the hEGF protein might be crucial. By inserting a glycine-serine spacer (G(S)_3 between hEGF and the double cysteine containing fusion tag, potentially unfavorable interaction could be prevented. By extending the distance between the two cysteines within the double cysteine containing fusion tag might improve the recovery of properly folded and fully bio-functional Cys-Gly-Cys0_hEGF.
Expression and labeling of site-specific modified growth factors and nano-bodies

4.2.2 Expression and purification of TGFα

The TGFα sequence was ordered in the expression vector pTWIN1, resulting in the fusion construct CBD-Intein-TGFα. By introducing additional N- or C-terminal cysteines via site-directed mutagenesis, the desired fusion constructs CBD-Intein-Cys0_TGFα or CBD-Intein-TGFα_Cys51 were generated, respectively. Both fusion constructs were expressed in E.coli and purified under denaturing conditions as described in chapter 2, following the expression and purification approach.
Each purification step of the protein purification process was monitored using SDS gel electrophoresis. After induction of intein splicing, the protein pattern revealed three major bands: a) the expressed protein fusion construct CBD-Intein-TGFα_Cys51, b) the cleaved CBD-Intein purification tag and c) the desired final product TGFα_Cys51 (Figure 4-8, lane 3). These results were equivalent for Cys0_TGFα. The strong appearance of the expressed protein fusion construct CBD-Intein-TGFα_Cys51 in the SDS gel illustrated incomplete intein cleavage. The cleavage efficiency was about 50%, significantly lower as for hEGF, 100% intein cleavage (chapter 2). However, the result demonstrated the importance of the amino acids composition of the expressed protein. It is known that for good intein cleavage the terminal residues adjacent to the intein cleavage site are highly relevant. For the Ssp DnaB intein used in the present CBD-Intein-TGFα_Cys51 fusion construct, good cleavage amino acid pairs are reported to be cysteine-arginine, glycine-arginine or serine-arginine.\(^{23}\) The addition of these amino acids to the N-terminus of the natural TGFα sequence might increase the total cleavage efficiency. Also the exchange of the chosen intein Ssp DnaB to other reported inteins such as Npu DnaE intein\(^{24}\) might be considered and might provide an attractive alternative to reach high intein cleavage activity.

The final yield of the obtained Cys0-TGFα and TGFα_Cys51 was about 0.5 mg per liter LB media. The purity and corresponding mass of both constructs was confirmed by SDS-gel electrophoresis and mass spectrometry, TGFα_Cys51 expected mass: 5649.3, measured mass: 5650.1 (Figure 4-8). Subsequently TGFα_Cys51 was labeled with Alexa647 maleimide and its activity tested on A431 cells, as described in chapter 2 and 3. The incubation of labeled TGFα on EGFR over-expressing A431 cells resulted in no EGFR activation, revealed by undetectable endogenous levels of phospho-EGF receptor (tyrosine 1173) protein (Figure 4-9). The failure to trigger EGFR auto-phosphorylation indicated the biological inactivity of the generated TGFα constructs. This result was confirmed by confocal live cell imaging, in which no binding of TGFα to the EGF receptor on A431 cells was observed. Cys0_TGFα was not tested regarding its biological activity.

---

Figure 4-7: **Left:** N-terminal thiazolidine-ring protected Cys0_TGFα. **Right:** TGFα_Cys51 containing an extra C-terminal cysteine with natural thiol functionality.
Figure 4-8: Top: SDS-gel illustrating the purification and intein splicing process of TGFα_Cys51: lane 1/2: supernatant of the washed CBD-Intein-TGFα_Cys51 inclusion bodies; lane 3: washed CBD-Intein-TGFα_Cys51 inclusion bodies; lane 4: refolded and spliced CBD-Intein-TGFα_Cys51; lane 5: SEC purified TGFα_Cys51. Bottom: LC-MS trace of purified TGFα_Cys51. Left: MS Total Ion Count, Right: mass to charge ratio, Inset deconvoluted mass spectrum; expected mass TGFα_Cys51 3x S-S: 5649.3, observed mass 5650.1.
Chapter 4

Figure 4-9: Biological activity of TGFα_Cys51_Alexa647 regarding EGFR in comparison to commercial hEGF. The activity of TGFα_Cys51_Alexa647 was measured on A431 cells, an EGFR over-expressing cell line, by determining the intrinsic phosphorylation level of EGFR.

4.2.3 \textit{EGb4\_Gly-Cys-Gly, a nano-body featuring a C-terminal orthogonal ligation site}

The EGFR receptors are validated anti-cancer drug targets and to date three different approaches were developed to eliminate EGFR signaling in over-expressing tumor cells: a) development of small inhibitor molecules against the intracellular tyrosine kinase activity such as iressa, PKI-166, traceva or GW-2016; b) intracellular chaperone antagonists such as heat shock protein (HSP90) blockers geldanamycin or radicicol; and c) the isolation of monoclonal antibodies against the extracellular EGFR domain such as cetuximab or trastuzumab.\textsuperscript{25,26} Antibodies are competing with the ligand for receptor binding, inducing receptor dimerization and internalization, but inhibiting ligand depending tyrosine kinase activity.\textsuperscript{25} Next to conventional antibodies (Figure 4-10), a second class of heavy chain antibodies (HcAbs) were discovered in the family of camelidae (i.e. Dromedary, Camel and Lama).\textsuperscript{27} These HcAbs are heavy-chain dimers lacking the light chains of classical IgGs. The heavy chain has no constant domain (CH1), structural replaced by an extended hinge region (Figure 4-10). Using small fragments of the variable heavy chain of heavy chain antibodies (VHH), small antibodies were developed called nano-bodies.\textsuperscript{28} The advantages of nano-bodies compared to classical antibodies are their small size (about 15 kDa) with high affinity. The combination of simple production and engineering properties with good solubility and high stability under harsh conditions,\textsuperscript{29,30} making nano-bodies ideal for research, industry and clinic.\textsuperscript{31,32} First commercial applications can be already found in shampoos to prevent dandruff,\textsuperscript{33} for protein purification as capturing device,\textsuperscript{34} in biosensors\textsuperscript{35} or in oncology research as possible therapeutic reagent.\textsuperscript{31,36}
Figure 4-10: Composition of conventional antibodies versus heavy chain antibodies and nano-bodies. Conventional IgG antibody contains of a heavy and a light chain, both responsible for antigen binding. Heavy chain only antibodies missing the light chain but full antigen binding are achieved. Nano-bodies only consist of the VHH domain. The difference between VH and VHH is the length of the complementary determining regions (CDR) of CDR1 and CDR3. Furthermore, in nano-bodies the CDR1 and CDR3 region is often connected via internal disulfide bonding. [C = constant, V = variable, H = heavy, L = light].

Roover et al. developed several Llama glama nano-bodies in the field of oncology. The designed nano-bodies block the EGF receptor catalytic function by competing with its natural ligand EGF. However screening against EGFR revealed also one non-EGF competing anti-EGFR nano-body, called EGb4. The nano-body binds to the extracellular domain I of ErbB1 and does not block EGF mediated signaling and cell proliferation. EGb4 is about 13 kDa in size, featuring two cysteines forming one disulfide bond. By introducing an additional cysteine at the C-terminus via site-directed mutagenesis, the construct EGb4_Gly-Cys-Gly (EGb4_GC) was created containing an orthogonal ligation site for site-specific C-terminal labeling (Figure 4-11).

Figure 4-11: EGb4_GC containing one extra C-terminal cysteine with its natural thiol functionality for site-selective labeling.

To express the nano-body EGb4_GC featuring one C-terminal site-specific ligation site, the fusion construct CBD-Intein-EGb4 was created, by cloning the EGb4 gene into the pTWIN-1 vector. Using site-directed mutagenesis, the C-terminal GCG tag was introduced leading to the final fusion construct CBD-Intein-EGb4_GC. After expression in E.coli, the fusion construct was purified as described in chapter 2, approach 2. The intein cleavage efficiency was with 100%,
comparable with the results obtained for the hEGF purification (chapter 2). The high purity of the recombinant expressed nano-body EGB4_GCG was confirmed by SDS-gel electrophoresis as well as mass spectrometry. However the measured mass of EGB4_GCG was 13594.7, two Daltons higher as expected (Figure 4-13). This mass difference either indicated the lack of disulfide bonding of the two internal EGB4 cysteines or a small measuring inaccuracy of the mass spectrometer (Figure 4-13). To investigate, if only one free thiol group is available for site-selective ligation, the nano-body was incubated with Alexa647 maleimide. If more than one thiol functionality would be available, multiple coupling of the Alexa647 dye to the expressed nano-body EGB4_GCG would occur. However, if correct internal disulfide formation took place during refolding, only one C-terminal cysteine is available for thiol-maleimide coupling. The incubation of the EGB4_GCG with Alexa647 maleimide showed a mass increase from 13594.7 to 14577, illustrating a 1:1 ratio of dye molecules to protein (Figure 4-13). Although complete C-terminal labeling of the nano-body could not be obtained (Figure 4-12), the result confirmed the correct folding of the EGB4_GCG nano-body. It demonstrated further the free accessibility of the C-terminal introduced thiol functionality for site-specific labeling.

To investigate the EGFR non-activating characteristics of the EGB4_GCG nano-body, the intracellular tyrosine phosphorylation levels of EGFR were measured, in nano-body treated and non-treated A431 cells (cell line over-expressing EGFR). The incubation of EGB4_GCG did not result in an increase of endogenous levels of phospho-EGF receptor (tyrosine 1173) protein, confirming the EGFR non-activating properties of EGB4_GCG (Figure 4-14). Subsequently, the site-specifically Alexa647 labeled nano-body EGB4_GCG-Alexa647 was used in live cell confocal imaging. EGB4_GCG-Alexa647 binding to the plasma membrane of A431 cells was observed with very low levels of internalization (Figure 4-15). No binding could be observed for EGB4_GCG-Alexa647 to the plasma membrane of HeLa cells (low expression level of EGFR) (Figure 4-15). Both results, EGFR activity assay and confocal microscopy, demonstrated the non-activating but EGFR selective binding properties of EGB4_GCG-Alexa647 to the EGFR. It further illustrated that the C-terminal labeling does not interfere with the biological characteristics of the nano-body.
Figure 4-13: LC-MS trace of EGB4_GCG. Top: EGB4_GCG, Bottom: EGB4_GCG-Alexa647. Left: MS Total Ion Count, Right: mass to charge ratio, Insets deconvoluted mass spectra; EGB4_GCG: expected mass 1× S-S: 13592.8, observed mass 13595; EGB4_GCG-Alexa647: expected mass 1× S-S: 14572.3, observed mass: 14576.3.

Figure 4-14: Test of the EGFR signaling response of EGB4-GCG in comparison to commercial hEGF. The activity of EGB4-GCG was measured on A431 cells, by determining the intrinsic phosphorylation state of EGFR.
4.3 Conclusions

The broad use of the developed protocol for site-specific labeling of cysteine rich proteins depends highly on the possibility to apply it to a wide range of different proteins. Chapter 2 and 3 describe the development of a protocol for hEGF and its function as a robust platform for the site-directed labeling of the protein on the N- as well as C-terminus. In this chapter, the protocol was applied to another member of the growth factor family TGFα, resulting in the successful expression of the modified protein but in a biological inactive form. Although incorrectly folded, the expression of TGFα provided practical experience regarding the sensitivity of the intein expression system. To test the limits of the developed protocol, hEGF was N-terminal modified with a double cysteine fusion tag leading to unprotected and mis-folded hEGF. However increasing the spacer length between the two cysteines within the fusion tag might circumvent the experienced negative influence on the hEGF structure. The protocol was additionally tested on an EGFR binding but not activating nano-body, Egb4. Introduction of one C-terminal site-specific ligation site generated a modified nano-body, displaying an orthogonal ligation point for direct immobilization or ligand attachment. Labeling of the nano-body with a fluorescent dye resulted in mono-labeled protein, featuring its natural biological characteristics demonstrated by live cell confocal microscopy and EGFR phosphorylation assays.

4.4 Experimental

Materials and methods
All materials and compounds not mentioned below have been purchased as described in chapter 2.

Construction of CBD-Intein-Cys-2-Gly-1-Cys0_hEGF:
To generate Cys-2_Gly-1_Cys0_hEGF, the plasmid CBD-Intein-Cys0-hEGF was modified by point mutation using the Quick change Kit (Stratagene) and the primers CGC-hEGF-F: (5’-
Expression and labeling of site-specific modified growth factors and nano-bodies

gacatgatcaacctgcggtgccacagcactgatcgaa-3’) and the reverse primer CGC-hEGF-R: (5’-ttcagctgctgttgcgccacgctgagctgtttgataatgcct-3’). The DNA sequence was verified by DNA-sequencing.

Amino acid sequence of CBD-Intein-Cys-2-Gly-1-Cys0_hEGF:
-Chitin binding domain (CBD): underlined sequence
-Intein domain: italic sequence
-hEGF: bold sequence
-in brackets: Cys-Gly-Cys sequence

Expression of CBD-Intein-Cys-2-Gly-1-Cys0_hEGF:
Cys-2-Gly-1-Cys0_hEGF was expressed following the expression and purification scheme 2, described in chapter 2.

Construction of CBD-Intein-[(Cys0)_TGFα (Cys51)];
Cys0_TGFα was purchased in pTWIN1 from Invitrogen. To generate TGFα_Cys51, the N-terminal cysteine was removed via site-directed mutagenesis using the primers Del_Cys0_TGF_F: (5’-gaatgacatcattgtacacaacgttgtgagccactttaatgatt-3’) and Del_Cys0_TGF_R: (5’-aattcatattaaagtggctcacaacgttgtgtacaatgatgctattc-3’) and the C-terminal cysteine was introduced using the primer pair, Ins_Cys51_TGF_F: (5’-ccgatctgctggcatgttaaggatccggctg-3’) and Ins_Cys51_TGF_R: (5’-cagccggatctataatgtcagcgctg-3’). The sequence was confirmed by DNA sequencing.

Amino acid sequence of CBD-Intein-[(Cys0)_TGFα (Cys51)];
-Chitin binding domain (CBD): underlined sequence
-Intein domain: italic sequence
-TGFα: bold sequence
-in brackets: N- or C-terminal cysteine

Expression of CBD-Intein-[(Cys0)_TGFα (Cys51)];
[(Cys0)_TGFα, (Cys51)] was expressed following the expression and purification scheme 2, described in chapter 2.
Construction of CBD-Intein-EGb4_GCG:

EGb4 was obtained in pUR5850 from Rob Roovers (University Utrecht) and sub-cloned in pTWIN1 by introducing the C-terminal cysteines using the primer pair: EGb4 SapI F: (5'-ggtggtggtcctcaacaggtgcagtcggag-3') and EGb4 BamHI R GCG: (5'-ggtggtggtcctcaacaggtgcagtcggag-3'), respectively.

Amino acid sequence of CBD-Intein-EGb4_GCG:
- Chitin binding domain (CBD): underlined sequence
- Intein domain: italic sequence
- EGb4: bold sequence
- in brackets: C-terminal cysteines

MKIEEGKLTPGVSASWAQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWOLQONGNNGNLELRESGAISGDLSLISLASTGKRVSISKDLLEDKEFIAWINEQTMKLESAKVSRCFCTKKLVYLKTLRGTIKATANHRFLTDGWRKLDELSKEHIALPRKLESSLQLSPEIEKLSQSIDYWDSISVITGVEEVFDLTVPGPHNFDVANDIIHVNHQVQLQESGGGSVQAGGSLKLSCAASGRSFSTYAMGWFRQAPGDREVFVATISWTDSTYADSVKGRFTISRDNASKNTGYLQMNSLKPEDTAVYYCAADRWAASSRNRVDYDYGWGGTQVTVSS[GC8]

SDS-PAGE:
SDS-PAGE gel electrophoresis was performed as described in chapter 2.

4.5 References

(1) Backer, M. V.; Levashova, Z.; Levenson, R.; Blankenberg, F. G.; Backer, J. M. Cysteine-Containing Fusion Tag for Site-Specific Conjugation of Therapeutic and Imaging Agents to Targeting Proteins; Vol. 494.
Expression and labeling of site-specific modified growth factors and nano-bodies

(23) NEB Manual 2006, 1.4.
Fluorescent proteins with integrated RGD motifs

Abstract. The design, expression and evaluation of fluorescent mCitrine proteins featuring RGD motifs are described. A broad range of mCitrine proteins with RGD motifs incorporated at several positions within the protein main chain were generated and characterized. It is envisaged that the protein scaffold induces a diverse range of RGD conformations, based on the localization of the RGD motif in mCitrine. Depending on the generated conformation different integrin affinities as well as specificities should evolve. On the one hand original wild-type peptide sequences were exchanged into an RGD motif on positions reported not to influence the fluorescent properties. On the other hand the RGD motif was additionally inserted into mCitrine amino acid sequence by site-directed mutagenesis. The evaluation of both approaches revealed excellent fluorescent properties of the RGD containing mCitrine mutants. Screening experiments against four human integrin receptors revealed two strong binders as well as two potential selective binders, confirming the scaffold inducing properties of mCitrine, which can be explored for integrin binding and immobilization assays.
Chapter 5

5.1 Introduction

5.1.1 Fluorescent proteins

Since the discovery of the green fluorescent protein (GFP) in *Aequorea victoria* by Shimomura in the 1960s, a broad range of GFP variants were engineered with fluorescent emission ranging from blue, green to red. The key feature of all fluorescent proteins (FP) is their ability to spontaneously generate the intrinsic chromophore themselves without the need of additional co-factors or enzymes. The self-catalyzed intramolecular re-arrangement and protein folding mechanism is thereby essential for correct chromophore formation and results in the emission of strong fluorescent light, a property highly attractive for the direct visualization of biological functions and dynamic processes in living organisms. Understanding the correlation between structure and optical properties led to an enormous body of fluorescence based methods for researchers. Examples include: monitoring gene activation, labeling and analysis of individual proteins in single cells, cell-compartments, or biosensors for ion flux, pH or redox potential.

Digestion experiments with aequorin and later the crystal structure of GFP revealed the light absorbing peptide sequence in FPs, generated by the cyclization of the three adjacent amino acids Ser65-Tyr66-Gly67. Roger Tsien demonstrated that the alteration of serine 65 simplifies the excitation spectrum to a single peak. The substitution of serine at position 65 for threonine (S65T) also increased the brightness of GFP by a factor of 5 and led to a faster chromophore formation, more useful in cell studies. Mutations in the β-barrel positioned close to the chromophore influence the fluorescent spectrum of FPs and were used in the development of long stoke shift variants such as Saphire (T203I) or the yellow fluorescent protein (YFP), (T203Y). The enhanced YFP (EYFP) is one of the most used FPs in life science, but features a low compatibility to acidic pH and chloride ions. The exchange of glutamine at position 69 for methionine (Q69M) resulted in Citrine, an EYFP type with reduced acid liability and increased chloride stability. By introducing an extra point mutation A206K, the dimerization tendency of Citrine was decreased, leading to monomeric Citrine (mCitrine) (Figure 5-1), which in combination with cerulean is one of the best FRET pairs used today. A characteristic structural element of all FPs is the β-barrel, protecting the inner chromophore from the environment and restricting its flexibility. Therefore, rigorous re-arrangements or the introduction of additional amino acids inside the β-barrel might not be tolerated by FPs, because the structure is indispensable for the correct chromophore formation. However, intensive modification of the β-barrel identified positions in the protein compromising insertion and modifications without leading to loss of the fluorescent activity (Figure 5-1).
5.1.2 Integrins and the RGD motif

The integrin receptor, the major cell surface receptor responsible for cell-cell and cell-matrix interaction, is a heteromeric glycoprotein in which each subunit contributes to the ligand binding site. To date, 24 different mammalian integrins are known, each dimeric member generated by the combination of one of the 18 α- and 8 β-subunits. The role of each integrin type is highly specific, demonstrated by the gene knock out of certain integrins in mice. The ligand recognition occurs thereby often via a key tripeptide sequence, argine-glycine-asparagine called RGD motif, not only utilized by fibronectin, but many other proteins such as laminin or Von Willebrand Factor. The interplay between RGD containing proteins with their distinctive integrin receptor regulates a diverse array of fundamental cellular functions, ranging from cell-adhesion, -migration, -differentiation and -signaling. Mis-regulations of these processes are involved in osteoporosis, inflammation and several types of cancer, making integrins an appealing target for anti-cancer therapy. This has resulted in an intense and ongoing research to design RGD mimics as basis for integrin antagonists. To increase the low affinity of the linear RGD peptide, cyclic peptides were synthesized featuring higher integrin affinity as well as specificity combined with increased protease stability. The use of unnatural D-amino acids or β-amino acids as building blocks allowed peptide conformations with further increasing integrin binding activity, with at the same time reduced protease sensitivity. The integration of β-amino acids in cyclic RGD peptides gives the opportunity to fine tune the ring size in C1 steps. They can further stabilize the whole peptide secondary structure.

About half of the 24 integrins known, which are responsible for cell surface interaction, bind in a RGD dependent manner (Table 5-1). The combination of synthetic materials used for medical application such as prostheses, implants, tissue engineering etc. with RGD peptides is promising to solve inadequate interactions between artificial materials and cells. To date, the RGD
sequence is the most employed peptide sequence on synthetic surfaces to stimulate cell adhesion, because of its unique recognition motif regarding cellular distribution and use. However, despite the broad affinity of the integrin proteins to the RGD motif, a “single integrin [receptor] is able to discriminate between the RGD sequences of different proteins”, mainly due to the conformation and orientation of the RGD containing loop. Also the flanking amino acids influence the integrin affinity, such as the fourth amino acid serine in fibronectin (RGDS). The influence of the micro-environment on the integrin affinity to the RGD motif is probably beneficial when integrated into a protein scaffold. To investigate this in more detail, the chapter explores the insertion of RGD sequences into several different positions of mCitrine, which might lead to fluorescent RGD proteins featuring different integrin affinities and specificities depending on the position inserted or integrated. The mCitrine structure could thereby stabilize or induce bioactive conformations favorable for high integrin affinity. Structural investigation, such as crystallization, could correlate integrin binding activity to RGD geometry. This might not only reveal conformations similar to known good integrin binders, but also evolve new geometries for high affinity ligands. Having new structures in hand, extended computer based modeling could guide the scientific community to new and unique receptor targeting motifs. By screening a broad range of integrin receptors regarding RGD binding, conformations might be revealed displaying diverse integrin specificities. Developing fluorescent protein based integrin markers selective for only one or a restricted number of integrin receptors would be beneficial and a powerful alternative to integrin specific antibodies. The intrinsic fluorescence of the RGD-mCitrine conjugates would thereby be an advantage, making labeling with fluorescent dyes unnecessary. Further applications could also be envisaged in the generation of bioactive surfaces. Immobilization of RGD-mCitrine proteins on artificial surfaces will potentially enhance the biocompatibility. By combining the high affinity properties of the RGD motif with the intrinsic fluorescence of mCitrine, the detection of bio-activated patterns on synthetic surfaces can be simply accomplished by fluorescence microscopy. This advantage will especially play off against synthetic RGD peptides. Although simple to produce, RGD peptides featuring fluorescence properties can only by synthesized by using expensive fluorescent amino acids or via post-synthesis functionalization with a fluorescent dye. The bulky structure of mCitrine facilitates additionally the detection via atomic force microscopy (AFM), a characteristic not available for artificial RGD-peptides due to their small size.
Fluorescent proteins with integrated RGD motifs

Table 5-1: Selected integrins and their recognition motif. Binding that is weak or only seen under special conditions is indicated by parenthesis.32

<table>
<thead>
<tr>
<th>KQAGDV</th>
<th>LDV/IDS</th>
<th>RLD/KRLDGS</th>
<th>RGD</th>
<th>L/IET</th>
<th>R…D</th>
<th>YYGDLR/FYFDLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>αIIbβ3</td>
<td></td>
<td></td>
<td>αIIbβ3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>αVβ1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>αVβ3</td>
<td></td>
<td></td>
<td>αVβ3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>αVβ5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>αVβ6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>αVβ8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(α1β1)</td>
<td></td>
<td></td>
<td>(α2β1)</td>
<td></td>
<td></td>
<td>α2β1</td>
</tr>
<tr>
<td>(α3β1)</td>
<td></td>
<td></td>
<td>(α4β1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(α5β1)</td>
<td>(α5β1)</td>
<td>(α5β1)</td>
<td>(α7β1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>αMβ2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>αLβ2</td>
</tr>
</tbody>
</table>

5.2 Results and Discussion

5.2.1 Strategy for generating fluorescent mCitrine featuring RGD motifs

Fluorescence proteins (FP), such as the yellow fluorescent protein variant mCitrine, are highly stable molecules, often used as reporter systems to determine protein localization and function ex vivo as well as in vivo (transgenic animals).47,48 The wide use of FPs in molecular biology as a transcriptional reporter, fusion tag, biosensor, or partner for fluorescence resonance energy transfer (FRET) is based on the easy and spontaneous formation of a fluorescent unit, only quenched by acid pH or denaturation.49

To combine the intrinsic fluorescent of mCitrine with the integrin binding properties of the RGF motif, two strategies were followed:

a) insert / add one RGD motifs into the mCitrine sequence and
b) exchange the wild-type amino acid sequence of mCitrine into a RGD motif at given positions

The first approach aimed to integrate the RGD motif directly into the amino acid sequence of mCitrine. The cylindrical 11-stranded β-barrel (Figure 5-1) that serves as a protecting shell around the chromophore does not look like to tolerate major changes in its amino acid composition. However, permutations and insertion experiments demonstrated the possibility to insert foreign protein or peptide sequences at certain positions into the β-barrel without losing the fluorescent property of the FP.28,30 By using the reported integration sites for the new integration of the RGD motif, the risk of losing the fluorescence properties of mCitrine was minimized. Furthermore, the addition of extra amino acids in the mCitrine structure may result in a better surface exposition of
the RGD motif.

In the second approach the wild-type amino acid sequence of mCitrine was exchanged into a RGD motif at given positions (Table 5-2). By using this approach the loop size or the β-barrel length of the FP will be maintained. This should minimize major structural re-arrangement in mCitrine and lead to correct chromophore positioning and maturation, necessary for the protein to display its fluorescent properties.

Both strategies, insertion and exchange, are envisioned to result in fluorescent mCitrine featuring surface exposed RGD motifs available for integrin binding. The structure of mCitrine could thereby serve as a scaffold, inducing different conformations of the RGD motif, leading to genetically encoded fluorescent RGD proteins with different affinities and specificities within the wide range of RGD binding integrins.

The N- and C-termini of mCitrine are also attractive fusion sites because they are exposed outside of the rigid β-barrel and can be used for peptides or proteins coupling without influencing the structural integrity of the FP itself. To utilize the N- or C-terminus, additional RGD motifs were introduced at the protein termini (Table 5-2).

To generate the above proposed mCitrine-RGD mutants, the RGD motif was integrated using the amino acid composition GRGDS. The RGD was flanked with additional amino acids, because the RGD sequence alone loses both affinity and specificity when taken out of its natural surrounded environment. However, the activity can be retained by adding the natural flanking amino acids: RGD (inactive) < RGDS < GRGDSP.36,44,51

The strategy to insert the GRGDS sequence in mCitrine was performed by site-directed mutagenesis. Using this approach, all DNA constructs were successfully obtained, except for one position 190/194 (Table 5-2). It was not revealed why this GRGDS mutant could not be generated. However, engineering 22 out of 23 plasmid constructs, the evaluation and characterization of many different RGD positions regarding integrin affinity and specificity was possible. As control for later integrin binding studies, wild-type mCitrine was generated containing an N-terminal cysteine.
Table 5-2: Amino acid positions for GRGDS or cysteine insertion / exchange. Underlined amino acids: inserted / exchanged amino acids. *Mutation that could not be obtained.

<table>
<thead>
<tr>
<th>GRGDS inserted</th>
<th>Before mutagenesis</th>
<th>After mutagenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/1</td>
<td>RGSVSK</td>
<td>RSGRGDSVSK</td>
</tr>
<tr>
<td>49/50</td>
<td>FICTTG</td>
<td>FICGRGDSSTTG</td>
</tr>
<tr>
<td>78/79</td>
<td>DHMKQH</td>
<td>DHMGRGDSKQH</td>
</tr>
<tr>
<td>116/117</td>
<td>FEGDTL</td>
<td>FEGGRGDSDTL</td>
</tr>
<tr>
<td>134/135</td>
<td>EDGNIL</td>
<td>EDGGRGDSNIL</td>
</tr>
<tr>
<td>140/141</td>
<td>GHKLEY</td>
<td>GHKGRGDSLEY</td>
</tr>
<tr>
<td>157/158</td>
<td>DKOQNG</td>
<td>DKOQGRGDSKNG</td>
</tr>
<tr>
<td>172/173</td>
<td>NIEDGS</td>
<td>NIEGRGDSDG5</td>
</tr>
<tr>
<td>189/190</td>
<td>PIGDGP</td>
<td>PIGGRGDSDG5</td>
</tr>
<tr>
<td>194/195</td>
<td>PVLLPD</td>
<td>PVLGRGDSLPD</td>
</tr>
<tr>
<td>213/214</td>
<td>PNEKRD</td>
<td>PNEGGRGDSKRD</td>
</tr>
<tr>
<td>240/241</td>
<td>PQFEK</td>
<td>PQFEKGRGDS</td>
</tr>
<tr>
<td>GRGDS exchanged</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49/53</td>
<td>TTGKL</td>
<td>GRGDS</td>
</tr>
<tr>
<td>78/82</td>
<td>MKQHD</td>
<td>GRGDS</td>
</tr>
<tr>
<td>116/121</td>
<td>GDTLVL</td>
<td>GRGDS</td>
</tr>
<tr>
<td>135/138</td>
<td>GNILTG</td>
<td>GRGDS</td>
</tr>
<tr>
<td>140/144</td>
<td>KLEYGN</td>
<td>GRGDS</td>
</tr>
<tr>
<td>156/160</td>
<td>KQKNGL</td>
<td>GRGDS</td>
</tr>
<tr>
<td>171/174</td>
<td>IEDGS</td>
<td>GRGDS</td>
</tr>
<tr>
<td>190/194</td>
<td>DGPVGL</td>
<td>DGPVGL*</td>
</tr>
<tr>
<td>194/198</td>
<td>LLPDNN</td>
<td>GRGDS</td>
</tr>
<tr>
<td>211/215</td>
<td>PNEKR</td>
<td>GRGDS</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C48V</td>
<td>KFICTTTG</td>
<td>KFIYTTG</td>
</tr>
</tbody>
</table>

5.2.2 Expression and purification of mCitrine mutants

All modified mCitrine constructs were successfully expressed as fusion proteins in the soluble fraction of E.coli, containing a N-terminal His-tag for purification and a C-terminal Strep-tag for purification and future surface immobilization experiments. The expression of the respective mCitrine constructs was confirmed by SDS-PAGE (Figure 5-2). The purification was carried out via nickel-nitrioltriacetic acid (Ni-NTA) affinity chromatography resulting in mCitrine-RGD/cysteine constructs with high purity (Figure 5-2). The high purity was achieved by very stringent washing conditions (40 mM imidazole) making a second purification step unnecessary. However the rigorous washing conditions led to a slight reduction of the overall mCitrine-RGD yield, compensated by the excellent expression characteristics of mCitrine in E.coli, resulting in yields between 0.5-28 mg per liter LB media. The broad range in yields reflected the influence of the RGD modifications on mCitrine structure. It indicated that different sites for RGD insertion are not equally tolerated regarding protein expression and stability. A decrease in the expression...
level revealed the negative structural effect of certain RGD positions, either by the creation of secondary structures that hinder optimal ribosome binding\textsuperscript{52,53} or the loss of the secondary structure resulting in rapid protein degradation. This means in turn if certain positions have a negative structural influence, other RGD positions might have a positive structural effect, which could result in mCitrine-RGD constructs with different integrin affinity as well as specificities.


5.2.3 \textit{Fluorescence of mCitrine mutants containing RGD or cysteine motifs}

The use of FPs as platform for generating fluorescent RGD motifs requires simultaneously an active FP and an accessible as well as correct folded RGD sequence. The integration of additional amino acids in mCitrine might reduce the intrinsic fluorescence of mCitrine, even when the integrated RGD motif is presented correctly. To exclude any interference of the RGD motif with the intrinsic fluorescent properties of mCitrine, all mutants were tested regarding their spectroscopic characteristics (Table 5-3). Intrinsic fluorescence could be detected for almost all RGD-mCitrine constructs (Table 5-3), indicating that the integration of a RGD motif is not altering the mCitrine structure or the correct chromophore formation. However non-fluorescent RGD-mCitrine variants were detected if the wild-type mCitrine sequence was exchanged into an RGD motif in the regions ranging from position 116 to 144 (Table 5-3). In contrast, the integration of five additional amino acids (GRGDS) in this protein region had no influence on the fluorescent characteristics of mCitrine, except for position 134/135 (Table 5-3). These findings were surprising but indicate that changes in this region alter the global structure of the FP or affect the correct orientation of the conserved chromogenic tripeptide (Ser-Tyr-Gly).

The positions that diminish the intrinsic fluorescence of mCitrine (region 116 to 144) are located in two different protein loops connected by one \(\beta\)-sheet (Figure 5-1). Mutations in the loop regions might destroy the rigid structure of the FP by placing the loop connecting \(\beta\)-sheet incorrectly into the \(\beta\)-barrel. Also when the chromophore protecting \(\beta\)-barrel is formed changes in the loops might alter the positioning of the chromophore generating amino acids to each other. This might
influence the essential chromophore cyclization reaction or alters the \( \pi \)-electron structure of the tripeptide that can result in the loss of yellow emission after excitation. Crystal structures or the step-wise substitution of amino acids in this protein region might reveal the chromophore structure and provide concepts for tuning the spectral properties of mCitrine as well as the integrin affinity/specificity of the RGD motif.

Table 5-3: Fluorescence of all expressed and purified mCitrine-RGD/cysteine constructs determined at 528 nm. The excitation of all mutants was carried out at 515 nm, concentration 1 \( \mu \)M in PBS. The protein concentration was determined at 280 nm.

<table>
<thead>
<tr>
<th></th>
<th>Emission intensity at 528 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRGDS inserted</strong></td>
<td></td>
</tr>
<tr>
<td>0/1</td>
<td>0</td>
</tr>
<tr>
<td>49/50</td>
<td>445</td>
</tr>
<tr>
<td>78/79</td>
<td>342</td>
</tr>
<tr>
<td>116/117</td>
<td>425</td>
</tr>
<tr>
<td>134/135</td>
<td>2</td>
</tr>
<tr>
<td>140/141</td>
<td>393</td>
</tr>
<tr>
<td>157/158</td>
<td>372</td>
</tr>
<tr>
<td>172/173</td>
<td>393</td>
</tr>
<tr>
<td>189/190</td>
<td>377</td>
</tr>
<tr>
<td>194/195</td>
<td>240</td>
</tr>
<tr>
<td>213/214</td>
<td>405</td>
</tr>
<tr>
<td>240/241</td>
<td>392</td>
</tr>
<tr>
<td><strong>GRGDS exchanged</strong></td>
<td></td>
</tr>
<tr>
<td>49/53</td>
<td>270</td>
</tr>
<tr>
<td>78/82</td>
<td>38</td>
</tr>
<tr>
<td>116/121</td>
<td>1</td>
</tr>
<tr>
<td>135/138</td>
<td>1</td>
</tr>
<tr>
<td>140/144</td>
<td>0</td>
</tr>
<tr>
<td>156/160</td>
<td>479</td>
</tr>
<tr>
<td>171/174</td>
<td>392</td>
</tr>
<tr>
<td>190/194</td>
<td>not expressed</td>
</tr>
<tr>
<td>194/198</td>
<td>235</td>
</tr>
<tr>
<td>211/215</td>
<td>192</td>
</tr>
<tr>
<td><strong>Negative control</strong></td>
<td></td>
</tr>
<tr>
<td>C48V</td>
<td>398</td>
</tr>
</tbody>
</table>

5.2.4 Determining the binding capability of mCitrine-RGD constructs to integrin receptors

The binding capability of all purified mCitrine-RGD proteins to several different integrin receptors was investigated via a microtiter-plate based integrin binding assay established for this purpose. The design of the assay is similar to an indirect sandwich enzyme linked immuno sorbent assay (ELISA), thereby the capture antibody is replaced by the integrin receptor and the target antigen by mCitrine-RGD featuring a N-terminal His-tag. By using a primary goat-anti-His
antibody, the mCitrine-RGD can be detected. Subsequently an anti-goat antibody conjugated to horseradish peroxidase (HRPO) recognizes the primary antibody, visualized by tetramethylbenzidine (TMB) substrate conversion (Figure 5-3). The level of substrate conversion is directly proportional to the concentration of bound mCitrine-RGD to the respective integrin receptor.

The microtiter-plate based integrin binding assay was established using the mCitrine-RGD protein 140/141 and mCitrine C48V (negative control) as initial model proteins. To achieve optimal readouts, antibody concentrations, mCitrine-RGD concentrations and blocking reagents were optimized. Good signal to background ratios could be obtained using 15 µg mCitrine-RGD and 0.5% bovine serum albumine (BSA) in combination with 1:1000 diluted primary and 1:10000 diluted secondary antibody (Figure 5-4). However the non-specific background signal of the mCitrine C48V protein was high under these conditions. To further improve the signal to noise ratio, three different blocking reagents were tested regarding their background reduction properties; fetale bovine serum (FBS), BSA and milk powder. Best results were obtained with 0.1% (w/v) milk powder, with almost no unspecific background signal (Figure 5-4). These conditions were subsequently used for screening all mCitrine-RGD constructs against four different human integrin receptors.

![Figure 5-3: Schematic representation of the microtiter-plate based integrin binding assay. Integrins receptors were immobilized on the surface of multi-well plates and incubated with RGD-mCitrine. Binding of RGD-mCitrine (containing a His-tag) was detected using primary anti-His antibodies. The signal was amplified using a secondary antibody conjugated to horseradish peroxidase (HRPO). Following TMB substrate conversion by HRPO, the binding was quantified at 450 nm.](image-url)
Figure 5-4: Determination of optimal conditions for the microtiter-plate based integrin binding assay. 

**Top:** Comparison of different mCitrine and primary antibody concentration on plates coated with human integrin receptor αVβ3 (0.05 µg / well in 96 well-plates), blocking reagent 0.5% BSA, 

**Middle:** Comparison of different blocking reagents, primary antibody 1:1000 diluted, well-plates not coated. 

**Bottom:** Comparison of the blocking reagents FBS and milk powder on plates coated with human integrin receptor: αVβ3 (0.05 µg / well using 96 well-plates), primary antibody 1:1000 diluted. mCitrine-RGD 140/141 was used as positive control, mCitrine C48V was used as negative control. 

Primary antibody: goat-anti-His, secondary antibody: anti-goat HRPO. All antibodies were diluted in 0.1% blocking reagent, the secondary antibody was diluted 1:10000. **Optimal determined concentrations:** mCitrine: 15 µg / well, primary antibody: 1:1000, secondary antibody: 1:10000, Blocking reagent: milk powder 0.1%.
In order to establish optimal conditions for the detection of mCitrine-RGD / integrin interaction, all expressed mutants were screened against four different human integrin receptors: αVβ3, αVβ5, α5β1 and α1β1. The screening revealed two strong integrin binders, 140/141 and 194/195 and several other good binders (Figure 5-5). Weak integrin interactions were mostly restricted to the integrin receptor αVβ3, with low binding found at the positions 116/117, 134/135, 172/173 and 213/214 for the integrated GRGDS motif and at the position 78/82 and 135/138 for the inserted GRGDS sequence (Figure 5-5). In the case of mutant 78/82, constant weak binding was detected to all four tested integrin receptors (Figure 5-5). However, the observed binding of the 78/82 construct is most probably due to nonspecific binding, because the integrin receptor α1β1 does not bind its ligand in an RGD-dependent fashion. All other weak integrin binders showed no binding to the screened RGD binding integrin receptors: αVβ5, α5β1 and α1β1, except for mutant 134/135 and the non-fluorescent mutant 135/138 (Figure 5-5). This strengthened the assumption that the displayed binding of 78/82 is due to nonspecific binding.

The 134/135 mutant as well as the non-fluorescent 135/138 mutant displayed weak selective binding to the integrin receptor αVβ3 (Figure 5-5). The observation could indicate an RGD structure with exclusive binding properties to the αVβ3 integrin receptor. While tuning the RGD structure regarding affinity and specificity, mCitrine-RGD mutants might develop with better as well as higher selectivity and affinity. Different conformations of the RGD motif are known not only to affect the integrin affinity, but also the integrin specificity. This has been reported for the cyclo(RGDfV) peptide, featuring a higher affinity towards αVβ3 in comparison to α5β1 and α1β1.\textsuperscript{54} The importance of the RGD ligand structure for efficient integrin binding is demonstrated by the integration of the GRGDS sequence at the N- and C-terminus of mCitrine. Both protein termini are not part of the rigid structure of FPs (Figure 5-1) resulting in mCitrine mutants containing a non-structured RGD motif. Missing the correct conformation for efficient integrin binding, both mutants feature no integrin binding, as illustrated in Figure 5-5.

The two binders 140/141 and 194/195 bound strongly to the human integrin receptor αVβ3 and αVβ5, but not to α5β1 or α1β1. The strong binding to both integrin receptors was independent of the integration strategy used; insertion or exchange (Figure 5-5). However, integration of the GRGDS motif by insertion showed higher integrin binding in comparison to integration by exchange (Figure 5-5). These results indicated an excellent surface exposition of the GRGDS sequence on the mCitrine scaffold. The enhanced binding of the mCitrine mutants featuring an inserted GRGDS motif, in comparison to an exchanged motif, might be due to a better surface exposition of the RGD motif caused by five extra amino acids in the protein sequence. To further investigate the specificity of the integrin binding at the position 140/141 and 194/195, the GRGDS motif was altered to a GRADS sequence. The modification of the RGD motif to an RAD motif diminished any integrin binding, confirming the selective binding of the integrated RGD motif to the integrin receptors αVβ3 and αVβ5 (Figure 5-6).

The two most effective GRGDS insertion / exchange positions (140 and 194) are located opposite to each other in the mCitrine structure (Figure 5-1). The different positions in mCitrine can be advantageous, regarding accessibility of the RGD motif after immobilization onto a
Fluorescent proteins with integrated RGD motifs

surface. Further it indicates optimal surface exposition of the RGD motif for good integrin binding. However, when both RGD positions in mCitrine (140/141 and 194/195) are optimally surface exposed for integrin binding, it was surprising to detect no binding to the integrin receptor α5β1 (Figure 5-5). The integrin receptor α5β1 is a typical RGD binder and similar affinities of GRGDS containing peptides are reported towards αVβ3 and α5β1.44

To increase the overall binding of the two mCitrine-RGD mutants (140/141 and 194/195) towards integrin receptors, a double mutant was expressed containing the GRGDS motif at position 140/141 as well as at position 194/195. To exclude any interference of the two RGD motif with the intrinsic fluorescent properties of mCitrine, single and double mutants were tested spectroscopically regarding their fluorescence properties (Figure 5-7). The emission spectra of the mCitrine-RGD mutants 140/141 and 194/195 showed adequate fluorescence, confirming that mCitrine is tolerating major changes in its amino acid composition at these positions (Figure 5-7). However, the double mutant 140_194 showed very weak fluorescence, indicating no correct chromophore formation. This was further confirmed by no absorbance at 515 nm (Figure 5-7, Inset). Additionally, the generated double mutant 140/194 was screened against all four tested human integrins. No increase in integrin binding in comparison to the single mutants 140/141 and 194/195 was observed (Figure 5-6). This could be due to the opposite location of the two GRGDS insertion sites in mCitrine, preventing simultaneous binding to two integrins: (Figure 5-1) when the first RGD motif is bound to the integrin receptor, the rigid structure of the β-barrel may prevent the binding of the oppositely located second RGD motif. No binding of the double mutant was observed when the RGD motif was substituted to an RAD motif (Figure 5-6), confirming RGD mediated integrin binding.
Chapter 5

Figure 5-5: Screening overview of mCitrine-RGD mutants. All mCitrine-RGD mutants were screened against four human integrin receptors: αVβ3, αVβ5, α5β1 and α1β1. The mCitrine-RGD mutant 140/141 and 194/195 bind strongly to human integrin αVβ3 and αVβ5. mCitrine-RGD mutant 78/82 binds to all four integrin receptors. Left: GRGDS sequence integrated into the wild-type mCitrine sequence at the given position, Right: wild-type mCitrine sequence exchanged into GRGDS sequence at the given position.
Figure 5-6: Screening overview of single and double mCitrine-RGD mutants in comparison to the respective mCitrine-RAD mutants. All mutants were screened against four human integrin receptors: αVβ3, αVβ5, α5β1 and α1β1. No increase in binding of the double mutant was observed versus the single mutants. **Left**: GRGDS mutants, **Right**: GRADS mutants.
Figure 5-7: Emission spectra of mCitrine-RGD mutants, 140/141, 194/195 and 140_195. The excitation of mCitrine-RGD mutants (1 μM in PBS) was carried out at 515 nm. Inset: Absorption at 280 nm and 515 nm.

5.3 Conclusions

Fluorescent proteins featuring genetically encoded RGD motifs are versatile tools for the evaluation of structure induced binding properties. By selecting mCitrine positions reported not to influence the fluorescence characteristics of the protein, 22 different mCitrine-RGD mutants were successfully generated. The RGD motif was integrated into the fluorescent protein either by direct insertion or by exchange of the wild-type sequence. Using Ni-NTA chromatography, all mCitrine-RGD variants could be obtained on a scale up to 28 mg in high purity. Excitation of the expressed mCitrine-RGD proteins led to strong fluorescent emission in the yellow spectral range. A major reduction of the fluorescence was detected in mutants featuring exchanged RGD motifs in the amino acid region 116 to 144. In contrast, integration of the RGD motif in the same protein region did not diminish the fluorescence of mCitrine.

After establishing a microtiter-plate integrin assay, all protein constructs were screened against four human integrin receptors (αVβ3, αVβ5, α5β1 and α1β1.) This initial in-vitro screening revealed two strong binders, 140/141 and 194/195 and two prospective selective binders 134/135 and 135/138. The two selective binders towards the human integrin receptor αVβ3, featuring highly promising opportunities for studies regarding the structure function relationship between RGD geometry and integrin binding. Developing RGD conformations featuring a high affinity as well as specificity are highly promising and have great potential in view of targeted imaging or cell immobilization.

The two strong binders 140/141 and 194/195 showed excellent binding towards the human integrin receptors αVβ3 and αVβ5. Both mutants were highly fluorescent and are potential alternatives to fluorescent labeled RGD peptides but future applications will depend on the possibility to bind integrin receptors in vivo. Crystallization of both mutants is therefore highly
Fluorescent proteins with integrated RGD motifs

recommended to determine the RGD conformation at these positions. If the RGD structure is elucidated, it might reveal how geometry can induce integrin specificity.

5.4 Experimental

Materials and methods

All materials and compounds not mentioned below have been purchased as described in chapter 2. Bugbuster and benzonase were purchased from Novagen. MaxiSorb plates were obtained from Nunc. All integrin variants were purchased from Millipore in Triton X-100 formulation. Goat anti-6-His antibody was obtained from Bethyl. Donkey Anti-Goat IgG (HRPO) antibody was bought from Leinco. TMB substrate and stopper solution were obtained from Cell Signaling. Magnesium chloride (MgCl2) and bovine serum albumin were purchased from Sigma. Manganese (II) chloride (MnCl2) and Tween20 were obtained from Merck. Milk powder was bought from Campina and fetal bovine serum (FBS) was obtained from Invitrogen.

Construction of pHT293 containing the construct His-mCitrine-Strep:
The plasmid was obtained from Dr. Hoang D. Nguyen.

Site-directed mutagenesis of His-mCitrine-Strep:
The mutations in His-mCitrine-Strep were carried out by Jurgen Schill, using the site-directed mutagenesis kit from Stratagene and primers listed below.

Primer to insert or exchange to cysteines or GRGDS / GRADS sequences:

<table>
<thead>
<tr>
<th>Primer to insert the GRGDS sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/1_F: acttaagggagag ggatcttaggg cggcaacctc accatatc attctctc</td>
</tr>
<tr>
<td>0/1_R: gagagtatgtatg atgtatgt atgtatgtatgtagttccggc gcggcatag cattctctctttaatt</td>
</tr>
<tr>
<td>240/241_F: gaggccatcgc aattttgaaa aaggggcggag accttaacttgac gctccggg</td>
</tr>
<tr>
<td>240/241_R: cccgggacgtta aagagttccgc gccctttttaaatggtggtgagcttcc</td>
</tr>
<tr>
<td>49/50_F: aagttcatcgaccggcgggacctacccggaacttc gccaagctg ccc</td>
</tr>
<tr>
<td>49/50_R: gggcagttgccggtatagtttcccgg cgcgggttg cagatgatgatc</td>
</tr>
<tr>
<td>78/79_F: ctaccccgacaccatg gggcgggac gcttaacgca cagcaaccttt</td>
</tr>
<tr>
<td>78/79_R: agaaatctgtctgtctgta gttccgacg cggcggcacttt</td>
</tr>
<tr>
<td>116/117_F: ggtaagttccagc aggccggcggg cgggaactgtgc ccaacccggac</td>
</tr>
<tr>
<td>116/117_R: gttccacagggtgtgctagcagtgcccgacggcgcctcggcctc</td>
</tr>
<tr>
<td>140/141_F: ctcctggggccacaagggcgggac gactctg ggtgactacaactaca</td>
</tr>
<tr>
<td>140/141_R: tgtgaggactacgagtc cggcgggacg cgccttgtgc gcccaggatg</td>
</tr>
<tr>
<td>134/135_F: ttcagggagagggcggg cgggaactta accttggggcgc</td>
</tr>
</tbody>
</table>

105
Chapter 5

| 134/135_R | gtgcccaaggatgttagagtccccggtcctcttgaa |
| 157/158_F | cattgcgcacaaggggcggggggcggggagaactctcagaagggcactgaag |
| 157/158_R | cttgtgatgcggatcaggtgctggcggccctgccatggtggcctgtg |
| 172/173_F | gccacaactcaggcggggcggggctcgcctcctgaaagaggttgacttggccag |
| 172/173_R | tcgcaactcaggcggggctcgcctcctgacagagggcctttg |
| 194/195_F | ggccccggtgctggggcggggactctctgcccgacaac |
| 194/195_R | gtgtgagcggagaggtccccgcgccccagcggtgg |
| 189/190_F | cccccctgcgcggggcggcccgtctctgactgacgag |
| 189/190_R | acgggggcggcccgtgcgtaggg |
| 213/214_F | caaatgcccccaacgaggggcggggctcgcctcctgaaagaggttgacttggccag |
| 213/214_R | cctcgctggtgctggcggcgtcggccagaggttgacttggccag |

**Primers to exchange original mCitrine sequence to GRGDS sequence**

| 49/53 FE | gacccgaagtttcatctggtgcggggagatcgcggccggagcctggcctgggccac |
| 49/53 RE | ggtggacgcccagggcggggagatcgcggccggagcctggcctgggccac |
| 78/82 FE | ttcgcccgtcactccccccacacaggggagagggatcggccggccggagcctggcctgggccac |
| 78/82 RE | ggcgatggcagaggttgtgataatcgcggccggagcctggcctgggccac |
| 116/121 FE | gcgcggcaggggtgaagttctcggggcggtggatctctctcggccggccggagcctggcctgggccac |
| 116/121 RE | tgccttcctcactgatcgccctgctggctggccggccggagcctggcctgggccac |
| 140/144 FE | cggcaacacttgggggagcgggggagcggaggtggatcgcggccggagcctggcctgggccac |
| 140/144 FR | gacgctttggtgctgtgtctgctgcccccagggatgtggcctgctggccggagcctggcctgggccac |
| 135/138 FE | aagggccagcggaggtggatcgcggccggagcctggcctgggccac |
| 135/138 FR | ggtgctttggtgctgtgtctgctgcccccagggatgtggcctgctggccggagcctggcctgggccac |
| 156/160 FE | ccacaacgttatatagctggcaggggagcggaggtggatcgcggccggagcctggcctgggccac |
| 156/160 RE | cggatctttggtgctgtgtctgctgcccccagggatgtggcctgctggccggagcctggcctgggccac |
| 171/174 FE | cttcaagatccggcaacgaggggagcggaggtggatcgcggccggagcctggcctgggccac |
| 171/174 RE | tgtggggaggtgctgtgtctgctgcccccagggatgtggcctgctggccggagcctggcctgggccac |
| 194/198 FE | gcaacgagggggggagcggaggtggatcgcggccggagcctggcctgggccac |
| 194/198 RE | agcttaaggttgtgctgtgtctgctgcccccagggatgtggcctgctggccggagcctggcctgggccac |
| 190/194 FE | gaacacccccatgcgggggagcggaggtggatcgcggccggagcctggcctgggccac |

106
Fluorescent proteins with integrated RGD motifs

Primers to exchange the original mCitrine sequence to a GRADS sequence

140/141GRADS_ F  cgcggctgctgctgctgctgctgctgcgcgcggactct
140/141GRADS_ R tgtagttgtactccagagagtccgcgcgccccttgtgccccaggatg
194/195GRADS_ F  ggccccgtgctggggcgcgcggactctctgcgcccgacaac
194/195GRADS_ R  tgtgctgctgctgctgctgctgctgcgcggactctctgcgcccgacaac

Primer for mCitrine negative control

C48V_F  gcaagctgaccctgaagttcatcgtgaccaccggcaagctg
C48V_R  cagcttgccggtggtcacgatgaacttcagggtcagcttgc

Amino acid sequence of His-mCitrine-Strep:

-His-tag: underlined and italic sequence
-mCitrine: bold sequence
-GRGDS insertion positions: bold and underlined sequence
-exchange to GRGDS: bold and green box
-cysteine insertion / exchange positions: grey box
-Strep-tag: underlined sequence

MHHHHHHHSGLVPGRS5VSKGEELFTGVVPIVLVEDGDVNGHKFSVSGEFGDATYGKLTLKFICGGTGLPVWPPLLVTTFGYGLMCFARYPDHMKQHDVF
FKSAMPEGVYQRTIFFKDGDNYKTRAEVKFEQDTLYGNRIELKGDIFKEDGNILCGLKLEYNYSHNVIYMADKQKNIKVNFKIRHNFQEDGVQLADHY
QQNTPIGDGPVLLPDNHLYSRYQSAESKDFNEKRDHMVLLEFVTAAGISRV
DVASHPQFEKStop

Expression of His-mCitrine-Strep:
The expression was carried out by Jurgen Schill. The different pH293 constructs were transformed in E.coli BL21 (DE3) cells. Clones containing the respective plasmid were inoculated in 20 ml Luria Bertani (LB) culture containing 50 µg / ml ampicillin. The culture have been grown in a shaking incubator at 37°C overnight. On the next day, 10 ml of the overnight culture were used to inoculate 1 liter LB media containing 50 µg / ml ampicillin. The culture was placed on a shaking incubator at 37°C for about 2.5 hours. At an optical density of A600 = 0.9, the expression was induced with IPTG, final concentration 1 mM. The culture was placed back into the shaking incubator and incubated at 25°C overnight. Subsequently, the culture was harvested and pelleted by centrifugation at 10000 g, 10 min, at 4°C. The cell pellet was stored at -80°C.

Purification of His-mCitrine-Strep:
The purification was carried out by Jurgen Schill. The cell pellet was thawed at room temperature and resuspended in 5-10 ml Bugbuster containing 1 µl / ml benzonase. To lyse the cells, the mixture was incubated
30 min at 4°C on a shaker and centrifuged at 10000 g for 10 min. The soluble fraction was collected and loaded onto 5 ml Ni-NTA resin equilibrated with 5 column volumes binding buffer (20 mM Tris, 500 mM NaCl, pH 7.8). Subsequently the resin was washed with 5 column volumes wash buffer (binding buffer, 40 mM imidazole) and eluted with 5 column volumes elution buffer (binding buffer, 300 mM imidazole). All obtained fractions (loading, washing, eluting) were analyzed by SDS-PAGE gel electrophoresis and LC-MS. The eluent fraction was dialyzed 4 times against PBS and concentrated via Amicon filters (cut-off 10 kDa) to a final volume of 0.5 – 2 ml.

**SDS-PAGE:**
SDS-PAGE gel electrophoresis was performed as described in chapter 2.

**Microtiter-plate based integrin binding assay:**
MaxiSorb plates were coated with 100 μl integrin (0.5 μg / ml dissolved in: 50 mM Tris, 150 mM NaCl, 1 mM MgCl₂, 2 mM MnCl₂, pH 7.4) on a shaker for 24 h at room temperature. Subsequently each well was washed three times 10 min with 100 μl blocking buffer (50 mM Tris, 150 mM NaCl, 1 mM MgCl₂, 2 mM MnCl₂, pH 7.4, 0.1% milk powder (w/v)). The plates were incubated 24 h at 4°C in the dark with 100 μl His-mCitrine-Strep (150 μg / ml diluted in blocking buffer) and washed three times 10 min with 100 μl wash buffer (50 mM Tris, 150 mM NaCl, 1 mM MgCl₂, 2 mM MnCl₂, pH 7.4, 0.05% Tween 20 (w/v)). The washed plate was blocked three times 10 min with 100 μl blocking buffer and subsequently incubated with 100 μl Goat anti-6-His antibody (1:1000 diluted in blocking buffer). After an incubation time of 24 h at 4°C in the dark, the antibody was removed and the plate washed three times for 10min with 100 μl wash buffer as well as 100 μl blocking buffer. After blocking, 100 μl of the second antibody (Donkey Anti-Goat IgG (HRPO) 1:10000 diluted in blocking buffer) was added and the plate was incubated 24 h at 4°C in the dark. Subsequently, the second antibody was removed and the plates washed three times for 10 min with 100 μl wash buffer as well as 100 μl blocking buffer. 100 μl TMB substrate was added and the plate was incubated for 15 min at 37°C in the dark. The enzymatic substrate conversion was stopped by the addition of 100 μl stopper solution and the absorption was read out at 450 nm.

**Fluorescent spectroscopy:**
The fluorescence spectra (λ<sub>ex</sub> = 515 nm, λ<sub>em</sub> = 520 – 650 nm) of all His-mCitrine-Strep mutants were measured in Varian Cary Eclipse fluorescence equipped with a Perkin-Elmer PTP-1 Peltier temperature control system. All measurements were perform in quartz cuvettes (10 mm light path), 2 ml minimal volume, at 20°C, concentration 1 μM in PBS.

**Absorbance spectroscopy:**
The UV-VIS absorbance spectra of all His-mCitrine-Strep mutants were recorded in a Jasco V-650 spectrophotometer equipped with a Perkin-Elmer PTP-1 Peltier temperature control system. All measurements were perform in quartz cuvettes (10 mm light path), 2 ml minimal volume, at 20°C, concentration 1 μM in PBS.
5.5 References


Abstract. Dynamic and reversible protein interactions play a key role in a number of biological processes. The emergence of synthetic and in water dynamically self-assembling supramolecular architectures provides an attractive entry to study non-covalent interactions in living biological systems. Interactions of dynamic supramolecular systems with cellular proteins provide a novel approach to tune protein function and assembly. Decorating non-cell permeable but dynamically self-assembling amphiphiles (discotics) with amine functionalities led to endocytic uptake into living cells. By using the SNAP-tag technology in combination with benzylguanine functionalized discotics, initial experiments were carried out for receptor mediated uptake into the cell as well as electroporation mediated uptake.


*equally contributing
6.1 Introduction

Supramolecular chemistry, the chemistry that goes beyond the covalent bond and the individual molecule\(^1\) has experienced an enormous development in the last decade. Starting with the fundamental work of Jean-Marie Lehn, Donald J. Cram and Charles J Petersen regarding crown ethers, cryptands or spherands,\(^2\) supramolecular chemistry has spread from material science via physics to chemistry and biology. In contrast to molecular chemistry that can produce a diverse number of highly sophisticated molecules from atoms linked by covalent bonds, supramolecular chemistry is generating complex systems based on self-assembly and self-recognition of building blocks held together via non-covalent bonds.\(^7,8\) The supramolecular systems created assemble through weak and reversible intermolecular interactions such as hydrogen bonding, ionic interactions, \(\pi-\pi\) stacking or Van-der-Waals interactions.\(^2\) Although individual non-covalent interactions are typically weak, “the summation of a large number of weak bonds [...] can lead to strong overall bonding in the supermolecule”.\(^9\) The reversible properties of the non-covalent connections introduce a dynamic character to the system by a constant flux of bond making and breaking events.\(^9\) This dynamic process can be influenced by environmental factors giving access to both, diversity and adaptability.\(^7\)

Supramolecular materials, dynamic by nature, are switchable by external stimuli because non-covalent interactions, in contrast to covalent bonds, can be dependent on pH, temperature, solvent and concentration.\(^10\) The degree of polymerization in supramolecular polymers, which are polymers generated by the self-assembly of monomers held together by complementary recognition groups, is the result of an equilibrium reaction between the assembly and disassembly of the monomeric units, directly regulated by the monomer concentration.\(^11\) The reversible cross-linking and self-organizing properties results in attractive applications such as autonomous self-healing polymers to seal structural defects\(^12\) or supramolecular hydrogels with high potential in regenerative medicine.\(^13\) Both applications are very interesting, scientifically as well as commercially.

Supramolecular concepts, initially inspired by nature, are nowadays mainly found in the field of material science such as supramolecular switches, electronics or polymers.\(^14,15\) Most of the supramolecular systems only show their dynamic self-assembly properties in organic solvents.\(^15\) However, all natural supramolecular polymers such as microtubulin,\(^16\) actin filaments\(^17\) or uncounted protein-protein interactions are occurring in an aqueous environment. Water provides two major challenges for supramolecular polymer scientists: a) the generation of water soluble building blocks and b) minimization or avoidance of the involvement or interference of water with the non-covalent interactions.\(^18\) On the other hand supramolecular chemistry “promise[s to reduce] the enormous complexity of biological systems to manageable levels”\(^19\) and significant progress has been made in recent years, resulting in many different supramolecular architectures able to assemble in aqueous media\(^18\) such as supramolecular self-assembling amphiphiles developed in the Brunsveld group.\(^19\) The main driving force for the association of supramolecular elements in water is the hydrophobic effect,\(^20\) mostly resulting from a strong hydrophobic
assembling component in the supramolecular polymer in combination with a structuring polar element.\textsuperscript{15}

To interface supramolecular chemistry with biological systems, which would potentially provide access to applications such as drug delivery, imaging or analytical tools, three important requirements must be fulfilled:

a) Self-assembly of the supramolecular system in aqueous media with association constants in the regime of bio-molecules, to enable an effective interaction with the biologically relevant target.

b) Directional and bio-orthogonal assembly of the supramolecular system, to enable selective and controllable interactions.

c) The ability to link different ligands orthogonal to the supramolecular scaffold without disturbing the assembled system in order to provide a wide spread applicability to diverse biological systems.\textsuperscript{15}

6.1.1 The supramolecular polymer of choice – a self-assembling fluorescent unit

A promising supramolecular polymer, which self-assembles in water, has recently been developed.\textsuperscript{19,21,22} It is based on the assembly of so called discotics into columnar architectures. This supramolecular platform has been shown to enable controlled association of bio-macromolecules along the dynamic scaffold and to feature a strong auto-fluorescence upon assembly, useful for detection in biological studies. The discotics can be functionalized with a number of ligands, for example via acylation or Huisgen [2+3] cycloaddition.\textsuperscript{23,24} These supramolecular polymers fulfill the three criteria formulated above: water solubility, orthogonal dynamic self-assembly and simple attachment of a diverse number of ligands and are therefore ideal candidates to target biologically relevant questions in which polyvalent recognition interactions are occurring in biological dynamic supramolecular assemblies.

The synthesis of the disc-shaped discotics is based on linking a central benzene-1,3,5-tricarbonyl unit to three 2,2'-bipyridine-3,3'-diamine molecules (Figure 6-1). This inner core element is additionally surrounded by three gallic acid moieties functionalized with PEG chains for water solubility. The termini of the PEG chains can be synthesized with an amine (Figure 6-1) or azide functionality for the attachment of biologically interesting ligands or orthogonal substrates such as a benzylguanine (BG) (Figure 6-7). The number of ligands can be varied between one and nine, depending on the functionalization of gallic acid with a diversity of PEG chains. For more detailed information regarding the synthesis of the discotics, the interested reader is referred to the PhD theses of Marion Müller\textsuperscript{25} and Katja Petkau-Milroy.\textsuperscript{26}
6.1.2 Supramolecular discotics – a dynamic tool for biology

“Living cells are highly dynamic systems”\(^{28}\) that create stable and diverse biological assemblies via supramolecular interactions.\(^{29}\) The spontaneous organization of simple building blocks into highly ordered structures by dynamic and reversible bonding enables the cell to create intelligent architecture, “self-assembled in metastable states dependent on an external supply of energy”.\(^{30}\) By using scaffold proteins as platform, the natural cell is organizing the enormous number of molecules, in space and time.\(^{31}\) This elegant concept, controlling protein activity and function upon assembly, might be suited for interfacing with synthetic supramolecular systems. By taking control over bio-macromolecules in- or outside the cell via supramolecular elements (Figure 6-2), the creation of new, or the modulation of natural cellular properties might be possible. This approach of “understanding by building”\(^{28}\) by using artificial protein assembly platforms should provide unique insights how individual protein interactions fit together in one complex system, the cell.\(^{28}\)

Here three different supramolecular self-assembling amphiphiles (discotics) will be evaluated using adherent HeLa cells; a) amino functionalized discotics, functionalized with one, three or nine amino groups, b) fluorescein labeled discotics and c) biotin and benzylguanine functionalized discotics. The amino functionalized discotics are envisaged to trigger charge mediated non-specific cellular uptake and if successful should be used as carrier for non-cell permeable discotics, such as fluorescein or biotin discotics.

Mono-functionalized benzylguanine discotics should not enter the cell via endocytosis. The benzylguanine discotics will be used for labeling and clustering of plasma membrane proteins featuring a SNAP-tag. The covalent attachment of the benzylguanine discotic to the SNAP-tag labeled protein is envisaged to result in endocytosis of the complex and maybe release into the cytoplasm (Figure 6-2). Furthermore electroporation mediated cellular uptake was investigated
using amino and benzylguanine functionalized discotics. Electroporation mediated uptake guarantees the free accessibility of the functionalized discotics inside the cell and might enable intracellular interactions such as between SNAP-tag labeled proteins and benzylguanine discotics, resulting in protein dimerization or sub-compartment targeting (Figure 6-2).

Figure 6-2: Supramolecular chemistry and biology. The graphic illustrates possible applications for supramolecular discotics inside and outside the cell. A) Ligand functionalized discotics bind to plasma membrane receptors, which results in receptor clustering and triggering receptor signaling or endocytic uptake. B) Direct cell penetration results in the uptake of the discotics into the cytosol, the free accessibility of the discotic might result in intracellular protein recruitment / dimerization / localization. L = Ligand, Rec = Receptor, P = Protein of interest.

6.2 Results and Discussion

6.2.1 Charge mediated cell uptake

The prerequisite of artificial supramolecular interactions inside the cell is the intracellular delivery of the synthetic compound. The nature of the plasma membrane, essential for cellular survival and functionality, prevents the spontaneous uptake of big molecules like supramolecular discotics. To overcome the selective barrier, membrane impermeable molecules can make use of active transport mechanisms such as specific (receptor-mediated) endocytosis or non-specific endocytosis.

Cell-penetrating peptides (CPPs) are often used for the delivery of molecules from the extracellular space into the cell via non-specific endocytosis. The translocation across the
membrane is mediated via the positive charge of the CPPs, which facilitates the electrostatic interactions to the negatively charged proteoglycans (membrane bound), resulting in uptake via endocytosis.\textsuperscript{35,36} By decorating the above described non-cell permeable discotics with amino groups, a positive charge is generated resulting in excellent uptake efficiencies of the 3/9NH\textsubscript{2}-discotics into HeLa cells (Figure 6-3).\textsuperscript{21} No difference could be detected in the uptake efficiency between 3NH\textsubscript{2}- and 9NH\textsubscript{2}-discotics. In contrast, no uptake of 1NH\textsubscript{2}-discotics could be observed at equal discotic concentrations (0.5-5 \textmu M). This might be explained by the overall lower number of positive charges generated by one amino group per monomer. Time-lapse confocal microscopy could further reveal fast binding of 3/9NH\textsubscript{2}-discotics to the outside of the plasma membrane within the first 10 min. After 30 min membrane bound discotics as well as internalized discotics could be detected in the cytosol, and after 60-90 min all discotic fluorescence was localized around the nucleus (Figure 6-4).\textsuperscript{26}

Figure 6-3: Confocal microscopy of HeLa cells treated with (5 \textmu M) 1NH\textsubscript{2}-discotics, 3NH\textsubscript{2}-discotics and 9NH\textsubscript{2}-discotics.\textsuperscript{21} Cellular uptake of 3/9NH\textsubscript{2}-discotics can be observed, whereas no cellular uptake of 1NH\textsubscript{2}-discotics can be detected.\textsuperscript{21}

Figure 6-4: Cellular uptake of 3NH\textsubscript{2}-Disc (5 \textmu M) by HeLa cells over time.\textsuperscript{21}

The dynamic character of the supramolecular discotics was proven by intermixing cell-permeable 3/9NH\textsubscript{2}-discotics with non-permeable biotin- or fluorescein discotics (Figure 6-5).\textsuperscript{21} The amine functionalized discotics served thereby as carrier molecules to enable cellular uptake of non-permeable biotin- or fluorescein discotics. The carrier function is generated through the dynamic co-assembly of the discotic stacks without the need of extended synthesis, illustrating the unique and powerful properties of supramolecular polymers in comparison to conventional polymers.
Figure 6-5: Confocal microscopy of fixed HeLa cells incubated for 1 h with a 5 µM mixture of 9NH2-Disc, fluorescein-Disc, and biotin-Disc (80:10:10). After 24 h, the cells were fixed and stained with a Cy5-labeled anti-biotin antibody.\(^{21}\)

In summary, the efficient endocytic uptake of amino functionalized discotic was shown, which was mediated by the positive charge of the amine groups. The amino functionalized discotics could thereby serve as carrier for otherwise non-cell permeable discotics, demonstrated by the uptake of mixed biotin- or fluorescein discotics. The uptake of the discotics mediated by charge is cell unspecific. To increase the cell target specificity, receptor mediated uptake was explored.

### 6.2.2 Receptor mediated cell uptake

The described auto-fluorescent supramolecular scaffold (section 6.2.1) was used as a platform to explore receptor mediated uptake into living cells. A highly biological relevant target, the epidermal growth factor receptor (EGFR), was chosen. To generate a covalent interaction between EGFR and the supramolecular scaffold, the SNAP-tag technology was used.\(^{37}\) By introducing an extracellular SNAP-tag into the EGFR via protein engineering and using discotics featuring one benzylguanine (BG) moiety, chemical labeling of the EGFR as well as receptor mediated internalization was envisaged (Figure 6-6).

Figure 6-6: Labeling and internalization of epidermal growth factor receptor (EGFR) with BG-discotics. Cellular localization of EGFR can be detected via fusion to mCherry, a fluorescent protein. Cellular localization of the BG-discotics can be detected using their intrinsic auto-fluorescence.
The benzylguanine / SNAP-tag technology was chosen as a reactive orthogonal functionality. The 21 kDa SNAP-tag, developed by K. Johnsson & co-workers, reacts fast, specifically and covalently with O\(^6\)-benzylguanine (BG) by irreversibly transferring the benzyl group from BG to a specific cysteine residue in the SNAP-tag.\(^{28,39}\) The genetically encoded SNAP-tag can be fused to any protein of interest, providing entry to site-selective extracellular or intracellular labeling.\(^{37,40}\) The discotic monomers featured one BG moiety per monomer.\(^{15,26}\) The accessibility and reactivity of the BG functionality on the supramolecular polymer was tested by incubating the expressed SNAP-tag fusion protein SNAP-ECFP with BG-discotics. The conjugation of BG-discotics to SNAP-ECFP was analyzed by SDS-PAGE. The SDS-PAGE of the mixture showed a slight gel shift upwards compared to the non-ligated SNAP-ECFP, confirming covalent attachment of the BG-discotic monomer (3 kDa in size) to the SNAP-ECFP protein (Figure 6-7). By exciting the BG-discotic via UV-light, only one band with the higher mass displayed fluorescence (Figure 6-7), confirming the irreversible attachment.\(^{26}\)

The epidermal growth factor receptor (EGFR) was chosen as trans-membrane protein target and genetically modified, featuring one extracellular SNAP-tag as well as an intracellular red fluorescent protein (mCherry) for detection. The resulting protein fusion construct SNAP-EGFR-mCherry was transfected and expressed in HeLa cells. Correct membrane localization was confirmed by confocal microscopy using the fluorescent properties of the mCherry domain, fused to the intracellular site of EGFR (Figure 6-8). Next to membrane associated SNAP-EGFR-mCherry,
intracellular EGFR was detected by mCherry fluorescence in the golgi apparatus (Figure 6-8). The aggregation of SNAP-EGFR-mCherry protein in the golgi is due to over-expression in HeLa cells. The accessibility and reactivity of the SNAP-tag on the cells was tested with a small fluorescent probe, Alexa647 labeled benzylguanine (BG-Alexa647). Incubation of HeLa cells expressing SNAP-EGFR-mCherry with BG-Alexa647 showed co-localization of BG-Alexa647 with SNAP-EGFR-mCherry on the plasma membrane as well as inside the cell (Figure 6-8). The result confirmed the accessibility and the reactivity of the SNAP-tag in the SNAP-EGFR-mCherry construct. BG-Alexa647 labeling of non SNAP-EGFR-mCherry expressing cells was not detected (Figure 6-8).

Figure 6-8: Confocal microscopy of HeLa cells transfected with SNAP-ErbB1-mCherry and incubated with BG-Alexa647 for 10 min. Arrows indicate boundaries of the fluorescent cell area. Left: Brightfield, Middle: mCherry channel, Right: Alexa647 channel.

Incubation of HeLa cells expressing SNAP-EGFR-mCherry with BG-discotics revealed no labeling after a short incubation time (5 min). However, incubation for a longer period (3 days) showed co-localization of the BG-discotics with SNAP-EGFR-mCherry inside the cell (Figure 6-9). Furthermore, cellular labeling with BG-discotics could only be achieved when the discotics were dissolved in DMSO prior to cell incubation. BG-discotic uptake could not be observed in cells not expressing SNAP-EGFR-mCherry (Figure 6-9), illustrating the specificity of the uptake.

The co-localization of the BG-discotics with SNAP-EGFR-mCherry inside the cell demonstrates the successful SNAP-BG interaction between discotics and the EGFR as well as the EGFR mediated internalization. The observed slower SNAP-BG reaction rate was expected since the stacking properties of the BG-discotics could lead to steric hindrance in comparison to small molecules. That the reaction rate is dependent on the nature of the molecules was experienced by the SNAP-ECFP / BG-discotics coupling reaction, which needed about 3 h at 37°C (Figure 6-7), a reaction rate much slower than reported for unmodified BG substrates in solution (k~2×10^4 s^-1 m^-1). However an increase to three days is unusual but might reflect the sterically more challenging environment of the cellular membrane. The need for DMSO prior to cell labeling might additionally indicate clustering of the supramolecular elements into larger assemblies under the selected molecular conditions, which further reduces the accessibility of the BG moiety. Also the linker length between the BG group and the discotics could be too short for bridging the
distance to the SNAP-EGFR on the cell surface. Both aggregation and inadequate linker length, might lead to the slow labeling of HeLa cells expressing SNAP-EGFR-mCherry.

Figure 6-9: Confocal microscopy of HeLa cells transfected with SNAP-ErbB1-mCherry and incubated with BG-discotics for 72 h. Left: mCherry channel, Middle: discotic channel, Right: overlay of both channels with the brightfield image.

To test the hypothesis regarding linker length, a PEG₁₂ spacer was successfully introduced between the discotics and BG moiety via standard maleimide and NHS-chemistry (Figure 6-10). By coupling benzylguanine-maleimide to the thiol functionality of SH-PEG₁₂-COOH (PEG-CT), BG-PEG₁₂-COOH was generated in quantitative yields. Direct coupling of the terminal carboxyl moiety to the amino functionalized 1NH₂-discotics via in-situ activation with HBTU was not successful. However, NHS activation of the carboxylic acid was partially achieved by using a large excess of NHS. During size-exclusion purification, the significant difference in retention time of the NHS activated compared to the non-reacted BG-PEG₁₂-COOH linker indicated back-folding of the PEG₁₂-linker due to strong hydrogen bonding between the amine rich benzylguanine and the carboxylic acid. This on the one hand hampered the activation of the carboxylic acid, but on the other hand enabled separation of two molecules with a mass difference of 100 using a size-exclusion column with a separation limit of 1000. The isolated BG-PEG₁₂-NHS was successfully conjugated to the amino functionalized 1NH₂-discotics, yielding in the final product: BG-PEG₁₂-discotic (Figure 6-10). Although complete conversion could not be achieved (Figure 6-10), sufficient material was synthesized for first cell experiments.

However initial incubation experiments with HeLa cells expressing SNAP-EGFR-mCherry and BG-PEG₁₂-discotics were not successful. Using larger excess of BG-PEG₁₂-discotics or extend the incubation time to 5 days did not lead to labeling of SNAP-EGFR-mCherry receptor on HeLa cells. This result indicates that not the linker length is the limiting factor, but rather the self-assembly process of the discotics itself.

To investigate the effect of temperature on the size / aggregation of the BG-discotics, Dynamic light scattering (DLS) measurements were performed. The autocorrelation function of BG-discotics at 20°C showed a single-exponential decay (Figure 6-11, left, blue). CONTIN analysis of this autocorrelation function revealed a unimodal size-distribution (Figure 6-11, right, inset, blue). The diffusion coefficient of the self-assembled BG-discotics was determined to be $D = 2.6 \times 10^{-8}$ cm²/s using angular dependent measurements. After heating the sample to 40°C the formation of
larger aggregates was observed (Figure 6-11, left, red). CONTIN analysis of the autocorrelation function, which did not show a single-exponential decay anymore, revealed a bimodal size distribution (Figure 6-11, right, inset, red). Next to particles with a diffusion coefficient of \( D = 2.9 \times 10^{-8} \text{ cm}^2 / \text{s} \), which are similar to those observed at 20°C, larger particles with a diffusion coefficient of \( D = 1.3 \times 10^{-10} \text{ cm}^2 / \text{s} \) were observed. The larger particles probably result from the aggregating of BG-discotics at higher temperatures (40°C). Since all cell labeling experiments were performed at 37°C, BG-discotics aggregation might hinder efficient EGFR labeling.

Figure 6-10: Structure and mass spectrum of BG-PEG12-discotics. **Left:** Structure of BG-PEG12-discotics. **Bottom right:** MALDI mass spectrum of BG-PEG12-discotics, peak1 = small trace of unreacted and modified 1NH2 discotics; expected mass 3308, observed mass 3360. The mass addition of +52 is elusive. Peak2 = BG-PEG12-discotics; expected mass BG-PEG12-discotics: 4414, expected mass BG-PEG12-discotics + sodium: 4437, observed mass 4437.

Figure 6-11: Dynamic light scattering (DLS) measurements of BG-discotics (10 \( \mu \text{M} \)) at 20°C and 40°C. **Left:** Plot of the autocorrelation function of BG-discotic at 20°C (blue) and 40°C (red) at an angle of 90°. **Right:** \( \Gamma \) versus \( q^2 \) plot. Inset shows the CONTIN size distributions for the 90° angle.
In summary, receptor mediated uptake of BG-discotics was shown. The internalization was mediated by endocytosis of the SNAP-tag labeled EGFR. Non SNAP-tag labeled cells showed no labeling or uptake of BG-discotics demonstrating the specificity of receptor-mediated uptake. Possible temperature induced aggregation of the BG-discotics reduced the labeling efficiency. However labeling and internalization were sufficient for detection by confocal microscopy. Considerable improvements are expected by performing cell labeling experiments at 20°C and in combination with intermixed 1NH2-/ BG-discotics, to reduce aggregation of the BG-discotics and to achieve faster labeling.

6.2.3 Electroporation mediated cell uptake

Electroporation is a broadly used methods to introduce DNA, RNA, proteins, peptides, dyes and drugs into the cell.\textsuperscript{43} The physical technique that reversibly permeabilizes cell membranes is a valuable tool in genetic engineering and provides an alternative to classical transformation methods.\textsuperscript{43} It is based on the generation of high intensity pulse(s) creating transient pores in the cell membrane facilitating the uptake of exogenous material such as DNA.\textsuperscript{43} After removal of the electric field, the membrane pores reseal allowing the survival of electroporated cells.\textsuperscript{44} The Cellaxess CX1 system used in the present work contains an electrode in conjunction with a capillary (Figure 6-12). The combination generates a) a locally restricted site of pore formation by the electrode and b) the capillary delivers a high intra- to extracellular concentration ratio of cell loading agent to the site of pore formation.\textsuperscript{42} Both properties facilitate the overall uptake of exogenous material. Furthermore, no trypsinization of adherent cells is necessary prior to electroporation. This is advantageous for adherent cell lines such as HeLa cells, avoiding problems with trypsinization and re-seeding and allows direct imaging using fluorescence microscopy.

Figure 6-12: Schematic representation of electroporation device Cellaxess CX1 adopted from reference.\textsuperscript{45}

Controlled intracellular protein recruitment or targeting via supramolecular polymers requires the free accessibility of their functional moieties in the cytosol. Here the direct electroporation of 1NH2-discotics as well as BG-discotics into HeLa cells is explored.

Optimal electroporation conditions and discotic concentration were evaluated using 1NH2-discotics because cellular uptake of 1NH2-discotics via endocytosis does not occur, as shown
controlled protein assemblies inside the cell via the supramolecular elements. 

Enable specific intracellular protein recruitment and co-intracellular outcome; entrapment into vesicles or cytoplasmic accessibility. If possible, it would supra-applications as supramolecular platform inside the cell. Due to the dynamic behavior of the discotics is linked to the self-molecular stacks, intermixing of 1NH₂ - and BG-discotics might be an option to tune the intracellular outcome; entrapment into vesicles or cytoplasmic accessibility. If possible, it would enable specific intracellular protein recruitment and could be used to generate spatially controlled protein assemblies inside the cell via the supramolecular elements.

Subsequently the optimized electroporation conditions for 1NH₂-discotics were successfully applied to BG-discotics, but required a longer pulse length, from 25 ms (1NH₂-discotics) to 50 ms for BG-discotics (see experimental section). The increase in pulse length, the time in which the cell is permeabilized, indicates an increase of size of the BG-discotics in comparison to the 1NH₂-discotics. This size increase is due to ligand alteration caused by the lack of charge repulsion. Substitution of the amino group on 1NH₂-discotics to a BG moiety removes the peripheral positive charge, eliminating the charge repulsion effect that limits the stacking size. This property, controlling stacking size by ionic interactions, was used in C₃-symmetrical benzene-1,3,5-tricarboxamide (BTA) based supramolecular polymers. ⁴⁶

The intracellular distribution of the 1NH₂-discotics and BG-discotics after electroporation was monitored over 24 hours. The time-lapse experiment showed for both systems homogeneous cytoplasmic distributions directly after electroporation, with no discotics signal detected inside the nucleus (Figure 6-13). After 1 h incubation, BG-discotics got sorted into vesicles and 24 h later were located around the perinuclear region (Figure 6-13). The localization around the nucleus was similar to the observed endocytic uptake of 3NH₂- or 9NH₂ discotics in HeLa cells. ²¹ Because co-staining of 3NH₂- or 9NH₂ discotics with endosomal marker revealed localization in endosomes, ²¹ it indicates as well the uptake of the BG-discotics into endosomes. In contrast to the BG-discotics, the cellular distribution of 1NH₂ discotics showed no entrapment into intracellular vesicles. Rather an even distribution through the whole cell. Additionally diffusion into the nucleus could be detected after 24 h, demonstrating the free diffusion of these supramolecular polymers inside the cell (Figure 6-13). The different distribution behavior illustrates the impact of small changes at the discotics periphery on the physiological outcome in vivo. The presence of the auto-fluorescence of discotics inside the cells over the time course of 24 h indicates the stability of these self-assembled polymers in the intracellular environment, since the strong fluorescence of these discotics is linked to the self-assembled state. This stability is an important feature for future applications as supramolecular platform inside the cell. Due to the dynamic behavior of the supramolecular stacks, intermixing of 1NH₂ - and BG-discotics might be an option to tune the intracellular outcome; entrapment into vesicles or cytoplasmic accessibility. If possible, it would enable specific intracellular protein recruitment and could be used to generate spatially controlled protein assemblies inside the cell via the supramolecular elements.
Figure 6-13: Fluorescent microscopy of HeLa cells electroporated with 30 µM 1NH₂-discotics (top) or BG-discotics (bottom), respectively. Time-course from 0-24 h after electroporation.

6.3 Conclusions

The possibility to use the interplay of supramolecular elements with proteins may open up an opportunity to influence and control biological processes in- and outside the cell and investigate the cellular response to them. By choosing auto-fluorescent amphiphiles decorated with amino groups, charge mediated uptake into living cells was demonstrated. By combining the cell permeable amine functionalized discotics with membrane impermeable discotic, such as biotin- or fluorescein discotics, cellular uptake of the otherwise non-permeable discotics was induced.

Using benzylguanine functionalized discotic, successful labeling and internalization of the EGFR on HeLa cells was achieved. Optimizing the labeling conditions is required to reduce BG-discotic aggregation. Incubation at lower temperatures as well as intermixing of the BG-discotic with positively charged amino functionalized discotics should prevent aggregation and might increase the local discotic concentration on the membrane due to charge interactions. By optimizing the system in this respect, controlled receptor clustering and cellular uptake of the supramolecular BG-discotic is envisaged.

After establishing electroporation conditions for the direct cellular uptake of 1NH₂- and BG-discotics, supramolecular discotics could be visualized for the first time as free accessible self-assemblies in the cytoplasm of a cell. Thereby different cellular distributions were detected. 1NH₂-discotics showed an even distribution through the whole cell, including the nucleus after 24 h. In contrast, BG-discotics were quickly removed from the cytoplasm by endosomal uptake and entrapped into vesicles, demonstrating the high influence of the connected ligands on the intracellular behavior of the discotics. By using intermixed 1NH₂- / BG-discotics, tuning between
freely available cytoplasmic and vesicle entrapped discotics is envisaged. This should open up the possibility to create synthetic supramolecular platforms inside the cell, used for protein recruitment, as a dimerization platform or as a synthetic intracellular carrier.

### 6.4 Experimental

#### Materials and methods

All materials and compounds not mentioned below have been purchased as described in chapters 2 to 5. Carboxyl- and sulfhydryl-terminated PEG\textsubscript{12} (PEG-CT) was purchased from Pierce. Benzylguanine-maleimide was obtained from NEB. N-(3-Dimethylaminopropyl)-N\textprime-ethylcarbodiimide hydrochloride (EDC), N-Hydroxsuccinimide (NHS), Triethylamine (TEA) and N,N\textprime-Dicyclohexylcarbodiimide (DCC) were purchased from Sigma. Extra dry N,N-Dimethylformamide (DMF) and dichloromethane (DCM) were obtained from Biosolve B.V.. Glass bottom dishes (35 mm) were obtained from MatTek. The Cellaxess CX1 system was bought from Cellectron. The plasmid SNAP-EGFR-mCherry was obtained from Jenny Ibach. All discotics (1NH\textsubscript{2}, 3NH\textsubscript{2}, 9NH\textsubscript{2}, fluorescein-, biotin-, BG-discotics) were obtained from Katja Petkau-Milroy.\textsuperscript{26} The anti-biotin antibody (Cy5-labelled) was purchased from Jackson ImmunoResearch.

#### Synthesis of Benzylguanine-PEG\textsubscript{12}-discotics:

1NH\textsubscript{2} discotics were obtained from Katja Petkau-Milroy. The synthesis is described in reference.\textsuperscript{26} To introduce a PEG spacer between its functionality and the discotics, a carboxyl- and sulfhydryl-terminated, SH-PEG\textsubscript{12}-COOH (PEG-CT) was used.

8.2 µmol PEG-CT was dissolved in DMF and reduced with 27.6 µmol TCEP for 5 min at room temperature. The reduced PEG-CT was mixed with 8.2 µmol benzylguanine-maleimide (dissolved in DMF), the pH adjusted to 7.5 and incubated at 30°C for 1 h. Product formation was confirmed by LC-MS analysis, LC-MS(ESI): $R_t=6.9$ min, $m/z$ calculated 1123.5, found 1123.8. Subsequently the product was purified by reversed phase high pressure liquid chromatography (RP-HPLC). RP-HPLC was performed on a Shimadzu LC-8A HPLC system by using a Gemini 5u C18 column. A gradient of water in acetonitrile, both containing 0.1% formic acid was used to elute the product. Detection was performed by a Shimadzu SPD-10AV UV-detector ($\lambda_\text{max}=280$ nm).

HPLC pure benzylguanine-PEG\textsubscript{12}-COOH was activated in dry DMF under Argon flux with 0.9 equivalents of NHS, 10 equivalents of DCC, 10 equivalent EDC and a few drops TEA. The reaction was continued at room temperature overnight and LC-MS confirmed conversion. Excess NHS and DCC was removed via size exclusion chromatography (BIO RAD BioBeads S-X1 (200-400 mesh) in a long glass column (1.2 m) under atmospheric pressure (in DCM) pressure and a flow rate less than 1 ml / min in DMF or in DCM. The solvent was evaporated and 1NH\textsubscript{2} discotics were coupled to benzylguanine-PEG\textsubscript{12}-NHS (ratio 1:1). The product formation was confirmed by Matrix assisted laser desorption/ionisation time of flight mass spectrum (MALDI-TOF-MS). MALDI-TOF-MS was measured on a PerSeptive Biosystems Voyager-DE PRO spectrometer with a Biospectrometry workstation using 2-{[(2E)-3-(4-tert-butyphenyl)-2-methylprop-2-enyldenede]malononitrile (DCTB) and α-Cyano-4-hydroxycinnamic acid (CHCA) as matrix material and chloroform as solvent. M/z values are given in g/mol. Remaining 1NH\textsubscript{2} discotics were removed via size exclusion chromatography as described above.

#### LC-MS analysis:

Samples were analyzed using a Shimadzu SCL-10 AD VP series HPLC coupled to a diode array detector
Chapter 6

(Finnigan Surveyor PDA Plus detector, Thermo Electron Corporation) and an Ion-Trap (LCQ Fleet, Thermo Scientific). Analyses were performed using a reversed phase HPLC column (GraceSmart PP18, 50 mm x 2.1 mm, 3 µm), using an injection volume of 1-4 µl, a flow rate of 0.20 ml / min and typically a gradient (5% to 100% in 10 minutes, held at 100% for 1 more minute) of acetonitrile in water (both containing 0.1% formic acid) at 25°C.

**Dynamic light scattering experiments (DLS):**
DLS measurements were performed by Katja Petkau-Milroy on an ALVCGS-3 Compact Goniometer at 90°. To evaluate the angular dependence of $\Gamma$, the scattering intensity was recorded at $30^\circ \leq \theta \leq 140^\circ$ in intervals of 10°, in six independent runs of 60 s. The incident beam was produced by a solid state laser, operating at 532 nm. The intensity signal was sent to an ALV5000 digital correlator, using a typical acquisition time of 60 s for each angle.

**Transfection of SNAP-ErbB1-mCherry:**
The transfection is described in chapter 3.

**Discotic uptake via electroporation:**
The electroporation of 1NH₂- or benzylguanine (BG) functionalized discotics were carried out with Cellaxess CX1 system. HeLa cells were seeded in Glass Bottom Dishes (35 mm) and electroporated at a confluency of 70 – 80%. The discotics (in water) were diluted in DMEN media (FBS and antibiotic free) to a concentration 30-50 µM. The settings for the electroporation have been:

<table>
<thead>
<tr>
<th>Name</th>
<th>Dispense Volume (µl)</th>
<th>Dispense rate (µl / min)</th>
<th>Voltage (V)</th>
<th>Pulse length (ms)</th>
<th>Interval (s)</th>
<th>Number of pulses</th>
<th>Start (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td>10</td>
<td>1</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>1NH₂-discotics</td>
<td>20</td>
<td>20</td>
<td>50-10</td>
<td>25</td>
<td>1</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>1BG-discotics</td>
<td>20</td>
<td>20</td>
<td>80-100</td>
<td>50</td>
<td>1</td>
<td>25</td>
<td>40</td>
</tr>
</tbody>
</table>

**Two-photon / Confocal microscopy:**
For confocal imaging, cells were seeded in uncoated 24 glass bottom wells purchased from MatTeK. The cells were imaged on a Leica TCS SP5 AOBS equipped with an HCX PL APO CS x63/1.2 NA water immersion lens and a temperature-controlled incubation chamber maintained at 37°C. Discotics were excited with a Chameleon Multiphoton laser using 760 nm laser pulses. Alexa647, mCherry, fluorescein and Cy5 were excited with a Whitelight laser (Alexa647 at 647 nm, mCherry at 587 nm, fluorescein at 488 nm, Cy5 at 633 nm). During imaging with Multiphoton laser the pinhole was fully opened, whereas otherwise it was closed to 1 airy unit. Fluorophore emission bands were detected in the following ranges: discotics, 500-550 nm; Alexa647, 665-699 nm; mCherry 600–632 nm; fluorescein, 496–520 nm; Cy5, 665-715 nm.
Fluorescent microscopy:

HeLa cells were imaged on a Zeiss AX10 equipped with a 40x/1.3 oil immersion lens. Discotics were excited with a HXP120 lamp. Bandpass filter discotics: excitation 357/44, emission 525/50. The photos were taken with an AxioCam MRm camera.

Antibody staining:

When staining with antibody and fixation was required, HeLa cells were seeded on glass slides in 24-well plates and cultured in DMEM media supplemented with 10% FBS and 1% Pen/Strep at 37°C, 5% CO₂. When the confluency reached about 70%, the media was removed and replaced by discotics dissolved in media. After 1 hour of incubation, disc-containing media was removed. After washing the cells 5 times with 2 ml PBS, the cells were fixed with 4% paraformaldehyde in PBS for 30 minutes at RT. The supernatant was discarded and the cells were washed 3 times with 2 ml Tris (50 mM, pH 8.0), NaCl (100 mM) and 1 time 10 min with 2 ml 20 mM Glycine in PBS. The cells were permeabilized for 15 minutes at room temperature with 0.1 % TritonX-100 in PBS and consequently washed 3 times for 5 minutes with 2 ml PBS.

Afterwards the cells were blocked with 5% BSA in PBS for 30 minutes at room temperature. To stain the 3Biotin-Disc the supernatant was removed and Cy5-conjugated monoclonal mouse anti-biotin antibody (1:600 diluted in PBS) was added and incubated with the cells at room temperature in the dark for 45 minutes. After the staining, the cells were washed 5 times with PBS, mounted with aqua polymer on a glass slide and dried for 1 hour in the dark at room temperature before imaging.

6.5 References

(10) Schmuck, C. Nachrichten aus der Chemie 2012, 60, 969-1062.
(18) Oshovsky, G. V.; Reinholdt, D. N.; Verboom, W. Angew Chem Int Ed Engl 2007, 46, 2366-93.
(24) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. Angew Chem Int Ed Engl 2002, 41, 2596-9.
(45) Cellectricon.
7

Epilogue
Chapter 7

7.1 Overview of the achievements described in this thesis

The expression of hEGF containing orthogonal ligation sites was developed and scaled up to a multi-milligram scale. To achieve this, the intein technology was applied that offered an attractive alternative for the selective removal of the purification tag by auto-excision without the use of additional enzymes. The biological activity of the cysteine rich hEGF protein was not influenced by the introduction of additional cysteines as site-selective ligation points. Remarkably, the use of one thiazolidine ring protected N-terminal cysteine enabled the introduction of a second unprotected cysteine at the C-terminus, which allowed the sequential double labeling of hEGF. This promising strategy for site-directed labeling of proteins featuring native cysteines was further employed on TGFα and the nano-body EGB4.

Furthermore, mCitrine proteins were generated featuring RGD motifs at different positions in the protein scaffold. Screening experiments against four different human integrin receptors revealed two high affinity binders as well as two potential selective binders. The mCitrine scaffold served as platform to induce the integrin affinity and selectivity. The use of supramolecular scaffolds as dynamic tools in biology was explored by self-assembling autofluorescent discotics. Charge and receptor mediated uptake into cells was shown, triggered by the dynamic intermixing of non-cell-permeable and permeable 3/9NH2- or BG-discotics. The uptake resulted in entrapment of the discotics into vesicles, which limited their free availability. Initial electroporation experiments revealed that cytoplasmic accessibility depends on the functionality attached at the periphery of the supramolecular element.

In this chapter possible future applications of the developed molecules are discussed, as well as their benefit in comparison to the natural unmodified molecules.

7.2 Application of hEGF containing orthogonal ligation sites

As discussed in chapters 2 to 4, EGFR is an attractive and biologically relevant target since its over-expression or mis-regulation is responsible for many types of cancer. Upon stimulation with hEGF, a conformational change of the EGFR occurs resulting in receptor dimerization and clustering. The localization of the EGFR into lipid rafts is believed to influence its ligand binding and phosphorylation and consequently its clustering behavior. However, the role of plasma membrane resident proteins and lipids which modulate the EGFR activity as well as questions about the “size and ligand-occupancy” of the EGFR signaling complex are unclear and are frequently disputed in recent research. This type of information is difficult to obtain because dynamic protein clustering and oligomerization occur on a nano-scale level. Attempts to address this challenging problem by local stimulation via latex or magnetic beads decorated with hEGF, uncovered details about the role of cholesterol and gangliosides as well as signal spreading during EGFR activation.

Applying quantum dots coupled to hEGF in microscopy revealed first details about EGFR dimerization, clustering and EGFR mobility. However to accomplish this, control over the interfacing EGF-EGFR interaction is necessary. This process is expensive and experimentally
challenging. Providing ligands such as hEGF with orthogonal handles would facilitate these types of experiments, enabling controlled stimulation, aggregation or manipulation of EGFR at specific plasma membrane locations. The probes described in chapter 2-4 are featuring these handles and can be site-specifically attached to dyes or surfaces. This would allow nano-actuation experiments with ligands in which the probe orientation and density is well-defined. Reversible and locally restricted EGFR actuations experiments are envisaged by functionalizing AFM tips with the developed hEGF (Figure 7-1). Attaching the stimulating ligands via the N- or C-terminus, the orientation of the biomolecules would be controlled. The applied ligand density could be tuned by generating scaffolds featuring single or multiple copies of hEGF. One possibility to obtain such scaffolds is using multivalent PEG spacers or streptavidin featuring different biotin valencies. Streptavidin bearing different valencies is described in chapter 3. Important advantages of functionalized AFM tips compared to hEGF coated beads are the reversibility of the EGFR stimulation and control over the number of stimuli per cell. Two orthogonal ligation sites on hEGF enable fluorescent labeling and immobilization simultaneously. As a consequence, surface functionalization could be simply visualized via the fluorescence of the attached dye. Alternatively, hEGF can be functionalized with two different fluorescent dyes (FRET pair or fluorescent dye and quencher), generating EGFR sensors to map the mechanical strain of the EGFR internalization upon stimulation, as described by Stabley et al.\textsuperscript{7}

The developed expression protocol for hEGF featuring orthogonal handles is envisaged to be applicable for cysteine rich proteins in general, which in turn should result in a variety of different probes containing site-specific ligation points (TGF\textalpha, VEGF, nano-bodies etc.). When having in hand, stimulation experiments with single or multiple copies of different ligands of choice are possible, an option that enables the simultaneous activation of several plasma membrane proteins, which might reveal information about the dynamic interplay inside the plasma membrane.

![Figure 7-1: Envisaged nano-actuation experiment with a hEGF functionalized AFM tip. By visualizing the EGFR on the plasma membrane upon stimulation, information about its mobility, dimerization and clustering behavior might be extracted. Left: A single hEGF molecule is coupled site-specifically to an AFM tip. Right: Scaffold featuring different numbers of hEGF molecules is coupled to an AFM tip. As scaffold could be used multifunctional PEG spacer or streptavidin containing different biotin valencies.](image)
7.3 Application of RGD-mCitrine proteins

RGD motifs genetically engineered into fluorescent proteins (FP) provide an attractive alternative to fluorescent labeled RGD peptides and might offer specific, genetically encoded, integrin binders. Screening against four human integrins revealed two integrin high affinity binders (140/141 and 194/195). The incubation with mammalian cells might confirm their in vitro binding ability. Their binding efficiencies could be evaluated by fluorescent microscopy or fluorescence activated cell sorting (FACS). Furthermore, crystallization with different integrin receptors is envisaged, which might provide new structural information about the integrin / RGD interaction. These insights could serve as template for the development of better and more specific integrin binders. The proposed FP scaffold might thereby induce integrin selectivity.

The immobilization of integrin binding RGD-mCitrine proteins onto synthetic surfaces is attractive for the development of biomaterials with cell adhering properties. Using the genetically incorporated Strep-tag in the RGD-mCitrine proteins, various surface patterns can be created for cell directed adherence and migration. By combining RGD-mCitrine proteins with supramolecular elements, such as the dimerizer Q8, surfaces with self-adaptable cell properties can be generated, as described by Neirynck et al.. The correct surface pattern can be confirmed by fluorescence microscopy or AFM (Figure 7-2).

7.4 Application of supramolecular discotics inside and outside the cell

Multivalent and dynamically reversible interactions play a fundamental role in many biological processes. Scaffold proteins, usually composed of “multiple modular interaction domains or motifs”, organize molecules autonomously into higher ordered systems. Mimicking these architectures by synthetic self-assembling systems such as supramolecular discotics (chapter 6) might facilitate artificially induced and controlled protein clustering inside and outside the cell. Due to their unique adaptive and responsive nature, supramolecular scaffolds adjust themselves
to their environment, enabling self-organization of protein interactions, ranging from single biomolecule interactions to protein clustering / oligomerization.

Exploring self-assembling auto-fluorescent discotics with different amine content showed effective cellular uptake of 3NH- and the 9NH2-discotics (chapter 6). Intermixing of cell permeable discotics with non-cell permeable biotin and fluorescein discotics led to cellular uptake of all three types of discotics. However, the charge mediated endocytosis resulted in endosomal entrapment of the discotics into vesicles, limiting their use as intracellular supramolecular platform.

Targeting plasma membrane proteins, cellular uptake of the supramolecular system is not required. We envisage modified discotics with mono- or multivalent reactive functionalities such as ligands or peptides that would allow reversible ligand induced receptor binding, clustering and activation (Figure 6-2, chapter 6). Initial experiments with benzylguanine (BG) functionalized discotics and EGFR exposing an extracellular SNAP-tag yielded co-localization of discotics with EGFR (chapter 6). However, the EGFR labeling efficiency was hampered due to discotic aggregation. Intermixing of BG-discotics with 1NH2-discotics will minimize aggregation due to charge repulsion. Additionally, intermixing with 1NH2-discotics will increase the positive charge on the supramolecular assembly, which in turn will increase the attraction and concentration of the BG-discotics onto the negatively charged membrane. When optimal intermixing conditions are established, receptor actuating discotics can be generated, i.e.: functionalized with hEGF (chapter 2-4). The density of the displayed bio-active ligand can be tuned by changing the composition of the supramolecular scaffold, simply by intermixing with inert or 1NH2-discotics. Appling to cells would lead to dynamic receptor clustering, signaling and receptor internalization. Extra information about the receptor engulfing vesicles could be obtained by incorporating additional ligands into the scaffold, such as pH sensitive probes. In case the scaffold is connected to an AFM tip, reversible and localized receptor stimulation would be possible. This locally restricted stimulation with a supramolecular probe might enable dynamic clustering and de-clustering of the plasma membrane receptors, which may provide new insights into the dynamic adaptability of the plasma membrane.

To exploit this method inside the cell, cellular uptake of the supramolecular assembly is required. As discussed above, charge mediated uptake led to endosomal entrapment of discotics, and their release would be difficult to achieve. However, this limitation can be overcome by microinjection or electroporation. Initial electroporation experiments of 1NH2-discotics and BG-discotics led to cellular uptake, resulting in free cytoplasmic accessibility or fast clearance into vesicles, respectively. Free accessibility of the BG-discotics in the cytoplasm might be triggered by intermixing with 1NH2-discotics. If successful, the SNAP-tag / BG system could be used for exploring supramolecular elements inside the cell. By covalently linking BG-discotics to SNAP-EYFP conjugates (EYFP=enhanced yellow fluorescent protein) featuring specific localization sequences, targeting of the BG-discotics to cellular sub-compartment might occur (Figure 7-3). SNAP-EYFP proteins will be expressed intracellularly and localized in their respective compartment (nucleus or membrane). If cytoplasmic BG-discotics are simultaneously
available, their covalent bonding to the SNAP-EYFP will occur, resulting in co-localization in the cellular sub-compartment (Figure 7-3). Both, SNAP-EYFP and BG-discotic, can be visualized by fluorescent microscopy, directly confirming their possible interactions. When successfully established, the supramolecular discotics might be used as synthetic scaffold to modulate signal transduction pathways. This might result in fundamental new insights into the complex signaling pathways used by cells, by either activation of intracellular signal molecules through the supramolecular scaffold or artificially coupling them together.

![Figure 7-3: Envisioned strategy for targeting non-cell permeable BG-discotics to different intracellular compartments (membrane and nucleus). Introduction of BG-discotics with SNAP-EYFP conjugates featuring localization sequences for nucleus or membrane, might enable protein recruitment and localization.](image)

7.5 References

Chemical Biology Approaches Targeting Cell Membrane Proteins

The plasma membrane is a dynamic and living boundary that characterizes a cell by separating it from its extracellular environment while simultaneously sensing and converting extracellular signals into the cell. Cell surface receptors in the plasma membrane are playing a central role in these signal transduction events and are actively targeted by specific signaling molecules, leading to a dynamic and reversible association and dissociation process of proteins and other components within the plasma membrane. However, the understanding of these complex and dynamic processes is still rudimentary. Targeting the dynamic nature of the plasma membrane is challenging and hampered by the lack of molecular probes, which allow control over the probe geometry and density. This type of probes is urgently needed.

This work is aiming to expand the molecular toolbox for such investigations by exploring novel biochemical molecules: extracellular signal molecules, fluorescent proteins and dynamic supramolecular scaffolds. The developed signal molecules (hEGF, EGb4) feature well-defined orthogonal ligation sites, which open up the possibility to control their orientation and density for applications such as nano-actuation, printing, protein assembly and targeting experiments. The site-specific attachment points have been engineered at the N- or C-terminus of the proteins to avoid undesired alterations in their biological activity. The recombinant expression of the signaling molecules featuring thiazolidine protected and unprotected orthogonal ligation sites was carried out in E.coli using the intein technology. The intein system allows the non-enzymatic removal of the affinity tag by auto-excision after purification. Chapter 2 describes the optimization of expression, purification, folding and intein cleavage of hEGF. Having established optimal conditions, three differently modified hEGF constructs were generated on a multi-milligram scale containing one or two additional N- or C-terminal cysteines as orthogonal handles: Cys0_hEGF, hEGF_Cys54 and Cys0_hEGF_Cys54. Insertion of an N-terminal thiazolidine protected cysteine or unprotected C-terminal cysteine in hEGF did not interfere with the disulfide formation of the native six cysteines, as receptor activation experiments demonstrated.

Chapter 3 discusses site-specific functionalization of these hEGF constructs at their protein terminus via thiol ligation chemistry. By addressing the N- and C-terminus in a stepwise fashion, sequential single or double labeling could be achieved. Cross-linking or oxidation of the additionally introduced cysteines in hEGF was prevented using thiazolidine ring protected cysteines. Full functionality of the site specifically labeled protein conjugates was proven by receptor activation experiments and live cell confocal imaging. Initial experiments with streptavidin featuring different numbers of biotin binding sites illustrated the potential use of hEGF displaying orthogonal ligation sites.
The global applicability of the established expression and purification protocol described in chapter 2 was applied to generate hEGF featuring a double cysteine containing fusion tag (CGC_hEGF). Additionally this technology was applied to other proteins containing a different number of native cysteines such as TGFα and the nano-body EGB4. CGC_hEGF was successfully produced but the modified protein was unfeasible for orthogonal labeling reaction due to potential misfolding. Optimizing the linker length of the double cysteine containing fusion tag might overcome this problem. TGFα was successfully expressed with additional N- or C-terminal ligation points but was not biologically active. To overcome this drawback the refolding conditions should be improved. Expression of the EGFR binding but non-activating nano-body EGB4 was achieved in good yields. Site-specific C-terminal labeling resulted in monovalent-labeled protein with bio-functional activity equivalent to unmodified nano-bodies.

In chapter 5, to target and visualize cell membrane integrin receptors, an alternative to fluorescently labeled RGD peptides has been developed. For this, a broad range of mCitrine proteins were generated with RGD motifs incorporated at several positions. Integration into mCitrine was envisaged to induce a diverse range of RGD conformations, based on the localization of the RGD motif in the protein. Depending on the generated conformation, different integrin affinities as well as specificities might evolve. The various positions were evaluated against four human integrin receptors by a specifically for this purpose developed ELISA assay, revealing two high affinity binders as well as two potential selective binders. Combination of the two high affinity RGD positions into one mCitrine molecule however did not increase the overall binding affinity.

Chapter 6 discusses synthetic supramolecular scaffolds as tool to study non-covalent interactions in- and outside the cell. Non-cell permeable but dynamically self-assembling auto-fluorescent amphiphiles (discotics) decorated with amine functionalities were studied with live cell confocal microscopy revealing endocytic internalization of discotics when displaying three or nine amine groups at their periphery. Intermixing of non-cell permeable fluorescein / biotin functionalized discotics with cell permeable amine discotics resulted in cellular uptake. Additionally receptor mediated uptake was shown using discotics bearing a benzylguanine (BG) moiety and cells expressing EGFR extracellular labeled with a SNAP tag. Although successful, the fast reaction kinetics of the BG-SNAP reaction observed in vitro was slower in the cellular environment, most probably due to higher order assembly of the discotics at 37°C. To enable intracellular discotic targeting, initial electroporation experiments with BG- and amine functionalized discotics were carried out, resulting in packing of the BG-discotics into intracellular vesicles shortly after uptake. Discotics decorated with amino groups remained free in the cytoplasm over 24 h. Intermixed supramolecular polymers containing BG-discotics and amino functionalized discotics might lead to the free accessibility of BG-discotics in the cytoplasm. This would open up the possibility to perform supra-molecular chemistry inside the cell, through specific targeting and localization of intra-cellular proteins.

In summary, signal molecules with orthogonal handles, fluorescent proteins with RGD motifs and dynamic supramolecular scaffolds decorated with different functional groups have been
explored as tools for targeting plasma membrane proteins. The development of an expression and purification protocol for hEGF containing site-specific ligation sites provides the opportunity to apply homogenous EGF stimuli to cells with control over the stimuli orientation and density. This level of control over the interfacing actuation process might provide new insights in the aggregation and clustering behavior of plasma membrane proteins upon stimulation. RGD containing fluorescent proteins provide an alternative to fluorescent RGD peptides. Additionally, using fluorescent proteins as scaffold, different RGD conformations can be envisioned leading to different affinities or selectivities to integrin receptors. Synthetic and self-organized supramolecular polymers allow the generation of homo- and heterovalent bioactive architectures. The ability of dynamic intermixing was demonstrated through the cellular uptake of non-cell permeable discotics induced by mixing with cell permeable discotics. Specific interaction of supramolecular elements with cell membrane proteins could be shown using the SNAP-tag technology. Combining dynamic intermixing and site-specific protein interaction on the plasma membrane should enable dynamic cross-linking or interactions of plasma membrane proteins. Introducing supramolecular systems into the cell can provide artificial signaling or dimerization platforms inside the cells. This has the potential to revolutionize our understanding about the dynamic and adaptable assemblies seen inside and outside the cell.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A431</td>
<td>Human epidermoid carcinoma cell line, EGFR over-expressing</td>
</tr>
<tr>
<td>AFM microscopy</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>Alexa647</td>
<td>Fluorescent dye</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulfate</td>
</tr>
<tr>
<td>AREG</td>
<td>Amphiregulin</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>AU</td>
<td>Absorbance unit</td>
</tr>
<tr>
<td>BCCP</td>
<td>Biotin carboxy carrier protein</td>
</tr>
<tr>
<td>BG</td>
<td>Benzylguanamine</td>
</tr>
<tr>
<td>BirA</td>
<td>Biotin ligase</td>
</tr>
<tr>
<td>BL21</td>
<td><em>E. coli</em> strain for protein expression</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BTA</td>
<td>Benzene-1,3,5-tricarboxamide</td>
</tr>
<tr>
<td>BTC</td>
<td>Betacellulin</td>
</tr>
<tr>
<td>CBD</td>
<td>Chitin binding domain</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementary determining region</td>
</tr>
<tr>
<td>CFP</td>
<td>Cyan fluorescent protein</td>
</tr>
<tr>
<td>CHCA</td>
<td>α-Cyano-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>CPPs</td>
<td>Cell-penetrating peptides</td>
</tr>
<tr>
<td>Cy3</td>
<td>Fluorescent dye, cyanine 3</td>
</tr>
<tr>
<td>Cy5</td>
<td>Fluorescent dye, cyanine 5</td>
</tr>
<tr>
<td>Cys0_hEGF_Cys54</td>
<td>hEGF containing an extra N-terminal cysteine as well as an extra C-terminal cysteine</td>
</tr>
<tr>
<td>Cys0_hEGF</td>
<td>hEGF containing an extra N-terminal cysteine</td>
</tr>
<tr>
<td>D</td>
<td>Diffusion coefficient</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DCC</td>
<td><em>N,N</em>-Dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DCTB</td>
<td>2-[(2E)-3-(4-tert-butylyphenyl)-2-methylprop-2-enylidene]-malononitrile</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DS streptavidin</td>
<td>Divalent streptavidin</td>
</tr>
<tr>
<td>DSMZ</td>
<td>Deutsche Sammlung von Mikroorganismen und Zellkulturen</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>DyLight549</td>
<td>Fluorescent dye</td>
</tr>
<tr>
<td>E.coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>ECFP</td>
<td>Enhanced Cyan Fluorescent Protein</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>EDC</td>
<td>N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGB4</td>
<td>Non-stimulating EGFR nano-body</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>EGFRvIII</td>
<td>Epidermal growth factor receptor that lacks 267 amino acids from its extracellular domain</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>EPGN</td>
<td>Epigen</td>
</tr>
<tr>
<td>EPL</td>
<td>Expressed protein ligation</td>
</tr>
<tr>
<td>ErbB1</td>
<td>Epidermal growth factor receptor-1</td>
</tr>
<tr>
<td>ErbB2</td>
<td>Epidermal growth factor receptor-2</td>
</tr>
<tr>
<td>ErbB3</td>
<td>Epidermal growth factor receptor-3</td>
</tr>
<tr>
<td>ErbB4</td>
<td>Epidermal growth factor receptor-4</td>
</tr>
<tr>
<td>EREG</td>
<td>Epiregulin</td>
</tr>
<tr>
<td>EYFP</td>
<td>Enhanced yellow fluorescent protein</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescent activated cell sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FIAsH</td>
<td>Tetracysteine dye, fluorescein-based arsenical hairpin binder</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescent protein</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescent resonance energy transfer</td>
</tr>
<tr>
<td>g</td>
<td>Gravity</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>(GGS)</td>
<td>Glycine-serine spacer</td>
</tr>
<tr>
<td>GndHCl</td>
<td>Guanidine hydrochloride</td>
</tr>
<tr>
<td>GPCR NK1</td>
<td>G-protein-coupled receptor neurokinin 1</td>
</tr>
<tr>
<td>GPCRs</td>
<td>G-protein-coupled receptors</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>hAGT</td>
<td>Human O6-alkylguanine-DNA alkyltransferase</td>
</tr>
<tr>
<td>HBEFG</td>
<td>Heparin-binding EGF-like growth factor</td>
</tr>
<tr>
<td>HcAbs</td>
<td>Heavy chain antibodies</td>
</tr>
<tr>
<td>HCl</td>
<td>Hydrochloric acid</td>
</tr>
<tr>
<td>hEGF</td>
<td>Human epidermal growth factor</td>
</tr>
<tr>
<td>hEGF_Cys54</td>
<td>hEGF containing an extra C-terminal cysteine</td>
</tr>
<tr>
<td>HeLa cells</td>
<td>Immortal human cell line derived from Henrietta Lacks</td>
</tr>
<tr>
<td>HRPO</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>HSP</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>IGF</td>
<td>Insulin-like growth factor</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G antibody</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-beta-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>Kd</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>LAP</td>
<td>Ligase acceptor peptide</td>
</tr>
<tr>
<td>LB media</td>
<td>Luria Bertani media</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography mass spectrometry</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass over charge ratio</td>
</tr>
<tr>
<td>MagicMedia</td>
<td>Autoinduction media</td>
</tr>
<tr>
<td>MALDI-TOF-MS</td>
<td>Matrix-Assisted-Laser-Desorption/Ionization Time-Of-Flight Mass-Spectrometry</td>
</tr>
<tr>
<td>MAPK pathway</td>
<td>Mitogen-activated protein kinase pathway</td>
</tr>
<tr>
<td>mEGF</td>
<td>Mouse epidermal growth factor</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>mg</td>
<td>Milligramm</td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>MnCl₂</td>
<td>Manganese (II) chloride</td>
</tr>
<tr>
<td>MS streptavidin</td>
<td>Monovalent streptavidin</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>NCL</td>
<td>Native chemical ligation</td>
</tr>
<tr>
<td>ng</td>
<td>nanogramm</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>Nickel-nitrol triacetic acid</td>
</tr>
<tr>
<td>Novablue</td>
<td>E.coli strain</td>
</tr>
<tr>
<td>NRG</td>
<td>Neuregulin</td>
</tr>
<tr>
<td>PALM microscopy</td>
<td>Photoactivated localization microscopy</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylen glycol</td>
</tr>
<tr>
<td>PEG-CT</td>
<td>Carboxy-PEG-Thiol Compound</td>
</tr>
<tr>
<td>Pen/Strep</td>
<td>Penicillin streptomycin</td>
</tr>
<tr>
<td>PI3K/AKT pathway</td>
<td>Phosphatidylinositol 3-kinase pathway</td>
</tr>
<tr>
<td>PLZF</td>
<td>Promyelocytic leukemia zinc finger</td>
</tr>
<tr>
<td>q</td>
<td>Scattering vector</td>
</tr>
<tr>
<td>ReAsH</td>
<td>Tetracysteine dye, resorufin-based arsenical hairpin binder</td>
</tr>
<tr>
<td>RGD</td>
<td>Amino acid sequence: arginine-glycine-aspartic acid (integrin recognition motif)</td>
</tr>
<tr>
<td>rGel</td>
<td>Recombinant gelonin</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>RT-HPLC</td>
<td>Reversed phase high pressure liquid chromatography</td>
</tr>
<tr>
<td>RTKs</td>
<td>Receptor tyrosine kinases</td>
</tr>
<tr>
<td>s</td>
<td>Second</td>
</tr>
<tr>
<td>SAB-Y1</td>
<td>Synthetic anti EGFR antibody</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SEC</td>
<td>Size-exclusion chromatography</td>
</tr>
<tr>
<td>SH2</td>
<td>Src homology 2, a cytoplasmic tyrosine kinase</td>
</tr>
<tr>
<td>SNAP-tag</td>
<td>Modified human O6-alkylguanine-DNA alkyltransferase</td>
</tr>
<tr>
<td>STAT pathway</td>
<td>Signal transducer and activator of transcription proteins pathway</td>
</tr>
<tr>
<td>STAT1</td>
<td>Signal transducer and activator of transcription factor 1</td>
</tr>
</tbody>
</table>
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>STED microscopy</td>
<td>Stimulated emission depletion microscopy</td>
</tr>
<tr>
<td>STORM</td>
<td>Stochastic optical reconstruction microscopy</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris-(2-carboxyethyl)-phosphine</td>
</tr>
<tr>
<td>TEA</td>
<td>Triethylamine</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethyl-ethane-1,2-diamine</td>
</tr>
<tr>
<td>TGFα</td>
<td>Transforming growth factor alpha</td>
</tr>
<tr>
<td>TKI</td>
<td>Tyrosine kinase inhibitor</td>
</tr>
<tr>
<td>TMB</td>
<td>Tetramethylbenzidine</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>Tris streptavidin</td>
<td>Trivalent streptavidin</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TS streptavidin</td>
<td>Tetravalent streptavidin</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>VHH</td>
<td>Variable heavy chain of heavy chain antibody</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow fluorescent protein</td>
</tr>
<tr>
<td>Г</td>
<td>Decay rate</td>
</tr>
<tr>
<td>µ</td>
<td>Micro</td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
</tr>
<tr>
<td>µm</td>
<td>Micrometer</td>
</tr>
</tbody>
</table>

Standard amino acid abbreviations

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>3 letter code</th>
<th>1 letter code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>R</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>
List of Publications


Sonntag, M. H.; Ibach, J.; Verveer, P.; Brunsveld, L., Site-specific protection and labeling of human epidermal growth factor, to be submitted.

Sonntag, M. H.; and Brunsveld, L., Site-specific protection and labeling of human growth factors, Invention disclosure, to be submitted.

Haase, C.; Sonntag, M. H.; Brunsveld, L., Chemical synthesis and site-specific protection of human epidermal growth factor, to be submitted.
Curriculum Vitae

Michael Sonntag was born on October 26th 1980 in Berlin, Germany. After finishing his military service in 2001, he went on to study Biotechnology at the Technische Universität Berlin. In 2004 he received his Vordiplom in Biotechnology and continued to study abroad at the University of Wollongong, Australia where he focused on genetics and molecular biology. After returning to Germany in 2005, he continued to study Biotechnology and worked part-time as student assistant at the Department of Genetics and Technical & Industrial Microbiology at the Technische Universität Berlin as well as at the Robert Koch-Institut Berlin at the HIV Diagnostics Department. In 2008 he received his Diplom-Ingenieur degree in Medical Biotechnology, after successfully finishing his research project “Detection of unknown viruses in german bats” under the supervision of Dr. A. Kurth at the Robert Koch-Institut Berlin. In 2009 he started his PhD project at the Technische Universität Eindhoven (the Netherlands) under the guidance of Prof. Dr. L. Brunsveld focusing on the expression and biological evaluation of recombinant proteins featuring orthogonal handles and the exploration of supramolecular polymers in biological systems. The most important results are presented in this dissertation.
Acknowledgement

In my last words I’m taking the opportunity to thank all the people who contributed to this work and made it an unforgettable experience for me.

First of all, I would like to express my gratitude to my promotor, Prof. Dr. Luc Brunsveld, for giving me the opportunity to work on this exciting and stimulating project. Geachte promotor, beste Luc, hartelijk dank voor jouw begeleiding en ondersteuning in de laatste 4 jaar. Je heb me echt veel vrijheid gegeven om dingen uit te proberen en te ontwikkelen. Daardoor kon ik mijn eigen ideeën volgen en het success daarvan in de praktijk testen. Jouw voortdurende enthousiasme heeft ervoor gezorgd dat ik het inein system niet opgegeven heb. Dank je voor jouw vertrouwen in mij.

I would like to thank Dr. Christian Ottmann to be my co-promotor and the members of my committee: Dr. Quentin Hanley, Prof. Dr. Menno Prins and Prof. Dr. Vinod Subramaniam for reading my manuscript and for all the valuable comments and suggestions. Dr. Peter Verveer is acknowledged for being part of my extended defensive committee.

The cooperation with Jenny Ibach and Peter Bosch introduced me to the world of microscopy and the underlying physics. This fruitful collaboration offered me a different perspective on plasma membrane proteins and readjusted my view on the cell. Jenny, danke für die unzähligen Confocal-stunden in Dortmund und die verschiedenen EGFR plasmide. Peter, hartelijk dank voor het uitleggen van de fysica van microscopie en de SNAP-EGFR cell line. I wish you both a lot of success with writing your theses.


A huge thank you goes to our analytical team, Lou Xianwen, Ralf Bovee and Joost van Dongen, for setting up machines and uncountable measurements. Dr. Christian Haase I would like to thank for countless attempts to chemically synthesize hEGF. I think we both underestimated the challenges but I’m confident that we will also bring this project to a successful end. Matthias, wir haben fast ein Jahr versucht Zellen lokal mit hEGF beschichteten Partikeln zu stimulieren. Auch
wenn die Interpretation der Ergebnisse oft schwierig und frustrierend war, haben wir uns nicht entmutigen lassen. Vielen Dank für die tolle Zusammenarbeit und viel Glück bei deiner Verteidigung.

Sascha (Dr. Fuchs) Dir möchte ich danken als Initiator für die vielen sozialen Aktivitäten in unserer Gruppe. Deine große Kommunikationsbereitschaft und unverbindlichen Art einen überall mit hin zu schleppe sorgten für unzählige Events. Fußball, Grillen und viele Verkleidungs-, Cocktails- und Tanzparties werden für immer in meiner Erinnerung bleiben.

Additionally, I would like to acknowledge the whole Brunsveld and Merkx’s group which has became my second home and family. The international environment and interdisciplinary cooperation with people from different cultures and scientific backgrounds was a great and unforgettable experience. I would like to thank: Dana, Katja, Lidia, Misha, Ralf, Jan, Melissa, Inga, Lars, Pauline, Brian, Edith, Luc, Christian, Hoang, Laurens, Loes, Sascha, Jeroen, Stijn, Dung, Lech, Ewelina, Ingrid, Matthew, Lech and Wencke. Joke I would like to thank for organizing all the official paper work and keeping an eye on all administrative obligations. Hans Damen I thank for the huge number of orders he took care of and kept track of their delivery. Wencke wil ik bedanken voor het synthetiseren en zuiveren van het Biotin peptide. Peggy I want to thank for the organization of the biolab and for taking care of all the different lab permissions.

Special thanks goes to my former office and lab mates Dr. Melissa Koay (my Paranymp) and Dr. Ewelina Burakowska-Meise. The time we spent together laughing, chatting or in science was always very calming for me and gave me the feeling that I’m not standing alone in the world. I’ve learnt a lot from you both scientifically and for life that will always be in my heart and mind. You became true friends and I hope that we will continue to share a lot of happy moments in the future. My new office mates Dr. Geert van Almen (alias DJ van Almen), Anne Hessels en Martijn van Rosmalen wil ik danken voor het vertrouwd maken met de Nederlands cultuur, de goede muziek en de relaxte atmosfeer in onze kantoor.


Zum Schluß möchte ich mich bei meiner Familie bedanken insbesondere bei meinen Eltern. Danke dass ihr mir auch in den für euch schweren Wendejahren eine so schöne und liebevolle
Acknowledgement


The last lines are dedicated to Fanni. Kedves Fanni, thanks for your love and support. I enjoy every second we spend together.

Micha