

# **Structural Studies and Protein Engineering of Human O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase**

THÈSE N° 5343 (2012)

PRÉSENTÉE LE 27 AVRIL 2012

À LA FACULTÉ DES SCIENCES DE BASE

LABORATOIRE D'INGÉNIERIE DES PROTÉINES

PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

**Birgit MOLLWITZ**

acceptée sur proposition du jury:

Prof. J. A. Hubbell, président du jury

Prof. K. Johnsson, directeur de thèse

Dr J. C. Rain, rapporteur

Prof. U. Röthlisberger, rapporteur

Dr Y.-H. Song, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2012



## Abstract

The specific labeling of proteins with synthetic probes is a powerful approach to study protein function and protein tags have been widely used for this purpose. A well-established example for a self-labeling protein tag is SNAP-tag. It specifically reacts with a wide variety of *O*<sup>6</sup>-benzylguanine derivatives (BG-derivatives) and was derived from the human *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase (hAGT) by protein engineering. Relative to hAGT, SNAP-tag possesses a 52-fold higher reactivity towards BG-derivatives, does not bind to DNA and expresses well in cells as on cell surfaces. It is known that alkylation of hAGT results in protein unfolding and degradation. However, an increased degradation of SNAP-tag fusion proteins after labeling has not been observed.

The first part of this work focused on the structural basis underlying the differences in protein stability between SNAP-tag and hAGT. A detailed biochemical and structural analysis was performed to determine (i) the interaction of SNAP-tag with its substrate, (ii) the factors responsible for its increased reactivity and (iii) how the labeling affected the stability of the protein. Besides an increased reactivity with BG-derivatives the superior stability of SNAP-tag compared to the parent protein hAGT could be confirmed. Whereas wild-type hAGT was rapidly degraded in cells after alkyl transfer, benzylated SNAP-tag showed a higher stability against proteolytic degradation. Moreover, the combination of our crystallographic and computational data provided further insight into the structural basis for the improved properties. The data indicated that the intrinsic stability of a key alpha helix was an important factor in triggering the unfolding and degradation of wild-type hAGT and provided new insights into the structure-function relationship of this DNA repair protein.

The second part was aiming for the generation of a new SNAP-tag-based inhibitor complex. It was envisaged that this complex would interact with the target protein via amino acid loops and decrease its function only upon labeling with BG-inhibitor molecules. Therefore, SNAP-tag was modified by the insertion of stretches of randomized amino acids and the generated protein libraries were screened for binding affinity. The utilization of two yeast-based systems, the yeast three-hybrid and two hybrid technologies, allowed for the differentiation of small-molecule dependent and independent binding interactions. It could be demonstrated that a specific protein-loop interaction can be generated by this approach. It could further be shown that inhibition of the catalytic activity of the target protein *E.coli* dihydrofolate reductase by a SNAP-loop mutant was possible.

In summary this work revealed new insights into the stability of hAGT and SNAP-tag and the structure-function relationship of AGTs in general. Further, SNAP-tag engineering generated a new protein-binder whose affinity towards the target protein was leading to protein inhibition.

Keywords: human *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase, protein stability, reactivity, self-labeling protein tag, *O*<sup>6</sup>-benzylguanine, SNAP-tag, structure-function relationship, protein-protein interaction, yeast two-hybrid, yeast three-hybrid, dihydrofolate reductase, human polo-like kinase 4.

## Zusammenfassung

Um die Funktionsweise von Proteinen genauer zu untersuchen ist das spezifische Markieren von Proteinen mit synthetischen Molekülen ein viel verwendeter Ansatz. Kleine Proteine, sogenannte Protein-Tags, kommen für diesen Zweck häufig zur Anwendung. Ein bekanntes Beispiel eines solchen selbst markierenden Protein-Tags ist SNAP-tag. Dieser wurde von der humanen  $O^6$ -Alkylguanin-DNA-Alkyltransferase (hAGT) durch Protein Engineering entwickelt und reagiert spezifisch mit Derivaten von  $O^6$ -Benzylguanin (BG-Derivate). Verglichen mit hAGT reagiert SNAP-tag 52-mal schneller mit BG-Derivaten, bindet nicht an DNA und lässt sich gut in Zellen als auch auf der Zelloberfläche exprimieren. Es ist bekannt, dass das Alkylieren von hAGT die Proteinstabilität beeinflusst und es dadurch zum intrazellulären Abbau des Proteins kommt. Für SNAP-tag und an SNAP-tag fusionierte Proteine wurde ein solcher Abbau allerdings nicht beobachtet.

Der erste Teil dieser Arbeit konzentrierte sich daher auf die strukturellen Unterschiede zwischen hAGT und SNAP-tag. Es wurde eine detaillierte biochemische und strukturelle Analyse durchgeführt um folgende Aspekte genauer zu betrachten; (i) die Interaktion von SNAP-tag mit seinem Substrat BG, (ii) die Faktoren, die für die gesteigerte Reaktivität von SNAP-tag verantwortlich sind und (iii) der Einfluss der Alkylierung auf die Proteinstabilität. Durch diese Untersuchungen konnte neben der gesteigerten Reaktivität mit BG-Derivaten auch die höhere Stabilität von SNAP-tag bestätigt werden. Der Wildtyp hAGT wurde in unseren Experimenten nach der Alkylierung schnell abgebaut, wohingegen SNAP-tag eine höhere Stabilität gegen den Proteinabbau aufwies. Die Kombination unserer kristallographischen und computergestützten Daten führte zu weiteren Erkenntnissen bezüglich der verbesserten Eigenschaften. Unsere Daten zeigten, dass die intrinsische Stabilität einer alpha-helikalen Region eine wichtige Rolle für den Proteinabbau von

hAGT spielt und führten zu neuen Schlussfolgerungen über die Zusammenhänge von Struktur und Funktion in diesem Protein.

Das Ziel des zweiten Teils war es ein neuartigen, auf SNAP-tag basierenden Inhibitorkomplex herzustellen. Dieser Komplex sollte an ein Zielprotein durch neu eingeführte Protein-Loops binden und nur nach der Reaktion mit einem BG-Inhibitormolekül die Funktion des Zielproteins inhibieren. Dazu wurden randomisierte Peptidsequenzen in SNAP-tag eingeführt und die so hergestellten Proteinbibliotheken bezüglich ihrer Bindungsaffinität selektiert. Die Verwendung von zwei auf Hefe beruhenden Selektionssystemen, die Yeast Two-Hybrid und Three-Hybrid Systeme, erlaubten die Unterscheidung von BG-Inhibitor abhängigen und -unabhängigen Interaktionen. Mit diesen Ansätzen war es möglich eine spezifische Interaktion zu generieren und es konnte gezeigt werden, dass eine Mutante von SNAP-tag die Eigenschaft besitzen kann die katalytische Aktivität des Zielproteins *E.coli* Dihydrofolatreduktase zu inhibieren.

Zusammenfassend konnten in dieser Arbeit neue Erkenntnisse bezüglich der Proteinstabilität von hAGT und SNAP-tag gewonnen werden und Rückschlüsse bezüglich der Beziehung von Struktur und Funktion für Alkyltransferasen im Allgemeinen gezogen werden. Es konnte darüber hinaus gezeigt werden, dass durch SNAP-Tag Engineering ein neuartiger Protein-Binder generiert werden konnte, welcher die Aktivität des Zielproteins inhibierte.

Schlüsselwörter: humane  $O^6$ -Alkylguanine-DNA Alkyltransferase, Proteinstabilität, Reaktivität, Selbstmarkierender Protein-Tag,  $O^6$ -Benzylguanin, SNAP-Tag, Struktur-Funktionsbeziehung, Protein-Proteininteraktion, Yeast Two-Hybrid, Yeast Three-Hybrid, Dihydrofolatreduktase, Polo-like Kinase 4.

## Abbreviations

AD	Activation Domain
APE <sub>x</sub>	Anchored Periplasmic Expression
AT	3-Amino-1,2,4-triazole
BC	Benzylcytosine
BD	Binding Domain
BG	O <sup>6</sup> -benzylguanine
bp	Base pair
BSA	Bovine serum albumin
CDR	Complementary Determining Region
cfu	Colony forming units
CM	Complete minimal medium
CP	Chloropyrimidine
Da	Molecular mass unit (g/mol)
DARPin	Designed Ankrin-Repeat Proteins
DHFR	Dihydrofolate reductase
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
Fab	Antigen-binding fragment
FEP	Free Energy Perturbation
FL	Fluorescein
FLAsH	Fluorescein Arsenical Helix binder
GDP	Guanosine-diphosphate
GST	Glutathione S-transferase
GTP	Guanosine-triphosphate
hAGT	human O <sup>6</sup> -alkylguanine-DNA alkyltransferase
HEK293	Human Embryonic Kidney Cell Line 293

His-tag	Hexa-histidine-tag
IC <sub>50</sub>	Half-maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KD	Kinase domain
M	Molar (=moles/litre) (unit)
MASPIT	Mammalian Small Molecule-Protein Interaction Trap
MD	Molecular Dynamics
Mtx	Methotrexate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
Ni-NTA	Nickel-Nitrilotriacetic acid
OD	Optical density
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethylenimine
Plk4	Polo-like kinase 4
RT	Room temperature
TI	Thermodynamic Integration
wt	Wildtype
xg	Multiples of the gravitational force
Y2H	Yeast-Two Hybrid
Y3H	Yeast-Three Hybrid
YPAD	Rich growth medium (Yeast extract, Peptone, Adenine and Dextrose)







## Table of Content

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Engineering for the Directed Evolution of Proteins .....	1
1.2	Protein-tags and the Evolution of SNAP-Tag from human O <sup>6</sup> -Alkylguanine DNA-Alkyltransferase .....	3
1.2.1	The human O <sup>6</sup> -Alkylguanine DNA-Alkyltransferase (hAGT).....	5
1.2.2	Protein Stability .....	6
1.3	The Importance of Protein-Protein Interactions .....	7
1.4	Antibodies.....	
1.5	Alternative Scaffolds .....	9
1.6	Overview of Selection Systems.....	1
1.6.1	Cell-Dependent Systems.....	11
1.6.2	Cell-Free Systems .....	14
1.6.3	Yeast Two-Hybrid .....	15
1.7	Aims of the Project.....	1
<b>2</b>	<b>Directed Evolution of hAGT to SNAP-tag – a Study on the Structure-Function Relationship.....</b>	
2.1	Structural Considerations .....	
2.2	Studies on Protein Stability.....	2
2.2.1	Proteolysis Experiments.....	22
2.2.2	Thermal Denaturation Assay .....	23
2.2.3	Stability of Labeled hAGT and SNAP-tag in Living Cells.....	25
2.3	Structural Analysis.....	2
2.3.1	Substrate Binding .....	27
2.3.2	Increased Stability.....	28
2.3.3	Molecular Dynamics Simulations.....	30
<b>3</b>	<b>Engineering of SNAP-based Inhibitors of Protein Function .....</b>	<b>3</b>
3.1	Introduction .....	3

<b>3.2 Construction of SNAP-Tag Loop Libraries.....</b>	<b>3</b>
<b>3.3 Optimization of Screening Conditions.....</b>	<b>4</b>
3.3.1 Bait-Proteins .....	40
3.3.2 Bait-Protein Expression in Yeast .....	41
3.3.3 Bait Auto-Activation.....	41
3.3.4 Pilot Screens.....	43
3.3.5 Growth Dependency of eDHFR and DHFR-Mutants.....	44
3.3.6 Small Molecule Labeling Efficiency in Yeast.....	46
3.3.7 Test Screening .....	48
<b>3.4 Screening against eDHFR.....</b>	
3.4.1 Bait Specificity of eDHFR Hits in Yeast .....	52
<b>3.5 Screening against Plk4 .....</b>	<b>5</b>
3.5.1 Bait-Specificity of Plk4 Hits in Yeast .....	56
<b>3.6 <i>In vitro</i> Analysis of Confirmed Hits .....</b>	
3.6.1 Labeling of SNAP-Mutants with BG-Substrates.....	60
3.6.2 GST Pull-Down with C-terminal SNAP-Tag Mutants .....	61
3.6.3 Further Characterization of Mutant Loop1 .....	62
<b>4 Discussion.....</b>	
<b>4.1 New Insights on Structural Aspects of hAGT and SNAP-Tag .....</b>	
4.1.1 Conclusions and Outlook.....	70
<b>4.2 Generation of Novel Protein Binders Based on the SNAP-Tag Scaffold .....</b>	<b>7</b>
4.2.1 Lessons from Library Design.....	71
4.2.2 Conclusions and Outlook.....	73
<b>5 Materials and Methods .....</b>	
<b>5.1 Structural Analysis of hAGT and Mutants .....</b>	
<b>5.2 Engineering of SNAP Loop-Mutants .....</b>	
5.2.1 Library Construction and Screening Set-Up .....	81
5.2.2 Yeast Screening .....	85
5.2.3 Hit-Validation.....	88
<b>5.3 List of Primers .....</b>	<b>91</b>
<b>6 Acknowledgments .....</b>	<b>93</b>
<b>7 References.....</b>	<b>95</b>
<b>8 Curriculum Vitae .....</b>	<b>111</b>



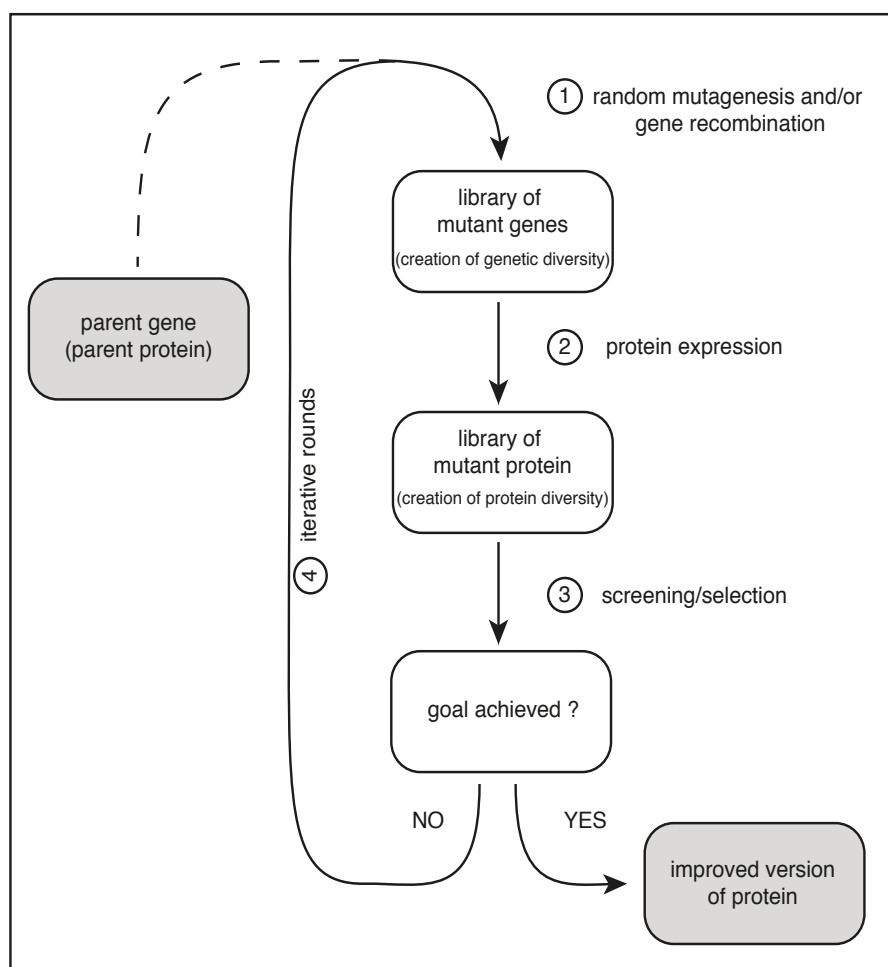


# 1 Introduction

## 1.1 Engineering for the directed evolution of proteins

Directed evolution mimics natural evolution as random mutagenesis is applied to a protein and improved variants with the desired qualities are picked out by diverse selection strategies. Tailoring the catalytic activity or stability of a protein has become a subject of interest among bioengineers as the application of proteins in the medical or industrial field has strongly increased over the last decade<sup>1-3</sup>. The results of these experiments also advanced our understanding of the forces that shape protein evolution and gave us a detailed knowledge about the structural aspects that are important for protein function<sup>4</sup>.

The set-up of experiments may vary widely, however all directed evolution experiments share the same basic evolutionary principles (Figure 1). The starting point is the parent protein and the engineering goal, for example the increase in protein reactivity. The gene of the parent protein is then modified, either by error-prone PCR<sup>5-7</sup>, DNA shuffling<sup>8-13</sup> or saturation mutagenesis<sup>14-16</sup>. In that way a library of different proteins is produced and subsequently screened for the desired function, mostly upon expression from an adequate expression system (bacteria, yeast or mammalian cells). Improved proteins are separated, amplified and may be resubmitted to further rounds of protein evolution, until the protein exhibits a sufficiently high level of the desired property.



**Figure 1:** General scheme of directed evolution. The parent gene of a protein of interest serves as starting point. The first step is the modification by random mutagenesis and/or gene recombination (1). Depending on the screening system of choice this gene pool library is expressed in a host system (2). The screening/selection step enriches the positive colonies that carry improved versions of the protein of choice (3). Most often several iterative rounds of gene randomization are needed until a satisfying degree of optimization has been achieved (4). The improved proteins are isolated and expressed in bigger quantity to further analyze their new properties in additional experiments. Illustration taken and modified from Johannes *et al.* <sup>3</sup>.

Saturation mutagenesis was found to be advantageous in exploring non-conservative amino acid substitutions in protein engineering. Random point mutagenesis on the protease subtilisin S41 identified Lys211 and Arg212 as a pair of residues that could improve protein stability<sup>16</sup>. Saturation mutagenesis of these residues selected highly hydrophobic residues for these positions and a significant number of mutants surpassed the stability of the point mutants. The selected amino acids were only accessible by multiple (two to three) base substitutions in a single codon and would



have been extremely rare in a point mutation library or in natural evolution. This study illustrated the power of laboratory protein evolution and emphasized the potential of saturation mutagenesis for the improvement of target proteins as a pathway that has been rarely explored in nature, but might offer important improvements as an alternative group of amino acids is made available<sup>16</sup>.

## 1.2 Protein-tags and the evolution of SNAP-tag from human *O*<sup>6</sup>-alkylguanine DNA-alkyltransferase

The possibility to select for certain protein features enabled the development of protein-tags. Protein tags are peptide sequences that are attached to proteins for various purposes. Amongst them, auto-fluorescent proteins represent an important group of protein-tags. They have been widely used for studying various aspects of protein function, from intracellular localization to rates of protein turnover<sup>17</sup>. Even though important advancements have been made in the field of auto-fluorescent proteins<sup>18-21</sup>, alternative protein-based tags that can be modified with a large variety of small molecule-derivatives remain an important alternative for the study of both extra- and intracellular proteins<sup>22,23</sup>.

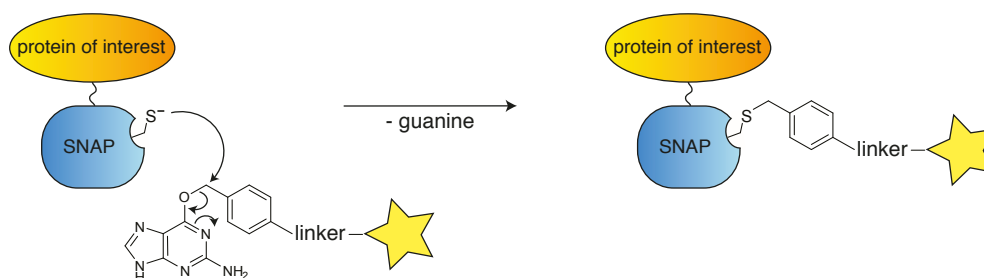
An example for a fluorescent protein tag of very small size is the FAsH-tag (Fluorescein Arsenical Helix binder). A tetrahistidine-tag can be labeled with biarsenical fluorophores to visualize intracellular proteins and the formation of biological complexes in living cells<sup>24,25</sup>. However, its main disadvantage is the unspecific binding as the probe can cross-react with thiols present in other proteins and cofactors such as lipoic acid<sup>26</sup> which is leading to high background signals.

Besides these two tag-systems, several protein-based labeling-techniques have been developed that offer a broader variety in the labeling with small molecules<sup>22</sup>. Those can be broadly divided into two categories: (i) enzyme-mediated tag labeling that requires the presence of a recognition sequence in the target protein and co-expression of an enzyme to perform the labeling reaction and (ii) self-labeling proteins that can transfer the functional part of a small molecule probe to the active site residue without any need of cofactors. The main advantage of the first approach is the small tag size

and the high specificity of an enzyme catalyzed reaction. However, it is often restricted to the labeling of cell surface proteins as the enzyme expression inside the cell might interfere with intracellular processes<sup>27,28</sup>.

In contrast, self-labeling proteins offer the advantage to function without any special co-factors and can generally be used for intracellular applications and *in vivo* experiments. The first developed self-labeling tag was the SNAP-tag<sup>29</sup>, which will be discussed in more detail. Other well-known examples of self-labeling protein tags include the orthogonal mutant CLIP-tag<sup>30</sup> and HaloTag<sup>31</sup>.

The labeling mechanism of SNAP-tag is based on the covalent transfer of a functionalized benzyl group of an *O*<sup>6</sup>-benzylguanine derivative to an active site cysteine to form a covalently modified protein<sup>29,32,33</sup> (Figure 2). The development of this protein-tag permits the labeling of SNAP-tag fusion proteins with a wide variety of different synthetic probes<sup>22</sup>.

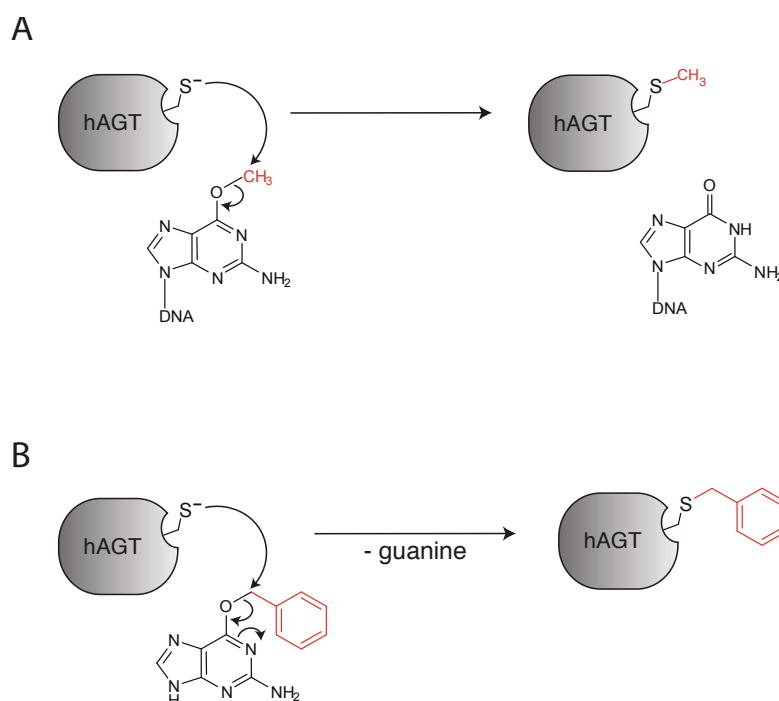


**Figure 2:** SNAP-tag reaction scheme with *O*<sup>6</sup>-benzylguanine derivative. The reactive cysteine 145 of SNAP-tag reacts with the benzyl moiety of the compound. After the reaction SNAP-tag is covalently labeled with the derivatized benzyl-part. SNAP-tag expression as fusion to a protein of interest enables the covalent labeling with a broad variety of compounds for cellular experiments.

SNAP-tag was generated in a stepwise manner from human *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase (hAGT) by introducing a total of 19 point mutations and deletion of 25 C-terminal residues. During directed evolution of the protein an emphasis was put on the increase of reactivity towards *O*<sup>6</sup>-benzylguanine-derivatives (BG-derivatives). Additionally, non-essential cysteines were removed and further mutations were introduced to abolish DNA-binding<sup>26,34,35</sup>. The next paragraph focuses on the wild-type protein hAGT, its role in nature and reveals important aspects on hAGT stability.

### 1.2.1 The human $O^6$ -alkylguanine DNA-alkyltransferase (hAGT).

The wild-type protein hAGT belongs to the group of DNA-repair proteins, which remove alkyl groups from the  $O^6$ -position of guanine and the  $O^4$ -position of thymine residues in DNA (Figure 3A). Alkylation at these positions represents an endogenous damage of DNA and leads to G:C to A:T transition in DNA base-pairing. In chemotherapies that are based on alkylating agents such as bischloroethylnitrosourea (BCNU), hAGT plays a significant role in the development of resistance<sup>36</sup>.  $O^6$ -benzylguanine is one of the best-known inhibitors of alkyltransferase and has been used as a very potent enhancer for chemotherapeutic agents<sup>37</sup>(Figure 3B). As for SNAP-tag, that follows the same reaction mechanism than hAGT, the alkyl-, or benzyl transfer leads to a covalent bond formation and is irreversible. As hAGT is a single turnover protein, nature developed an efficient way of hAGT removal after alkylation. Several studies have shown that alkylated hAGT is degraded very rapidly by the ubiquitin-proteasome pathway<sup>38,39</sup>. This is grounded in the fact that the alkylation event triggers a conformational change, which renders the protein more sensitive to proteolysis by proteases<sup>40,41</sup>.



**Figure 3:** hAGT reaction mechanism (A) on  $O^6$ -methylated guanine bases and (B) with the potent alkyltransferases inhibitor  $O^6$ -benzylguanine. Both alkylation reactions induce structural changes in hAGT leading to protein unfolding and increased ubiquitination and degradation via the proteasome pathway.

### 1.2.2 Protein stability

Understanding the factors that control protein stability and therefore the ability of the protein to function under harsh conditions *e.g.* high temperature or organic solvents, is of scientific and commercial interest<sup>1</sup>.

Proteases and peroxidases were amongst the first enzymes to be evolved for increased stability, mainly because of the need for stable and functional variants for detergent and laundry application<sup>42-44</sup>.

A general method to design the stability of a protein through protein engineering was developed by Schmid and coworkers<sup>45</sup>. In the reported approach named Protein Stability Increased by Directed Evolution (ProSIDE), filamentous phage is used to assess the stability of a protein library by determination of the proteolytic susceptibility<sup>1,46</sup>. A similar approach was published earlier by Kristensen *et al.*<sup>47</sup> The protein sequence to be evolved is sandwiched between two domains of the g3p protein, which is important for phage infectivity. Phages carrying clones that due to the inserted mutants do not fold properly are not able to propagate and are therefore eliminated from the screen. The propagated phages are further selected for high stability mutants by applying a selection pressure (either high temperature or guanidinium hydrochloride salt) followed by a protease treatment, which destroys phages with non-correctly folded proteins. This screening method assumes a direct correlation between proteolytic resistance and increased stability but it could be shown by other groups that this correlation cannot always be applied<sup>48,49</sup>.

Besides phage-based screenings the use of thermophilic microorganisms and selection by auxotrophic markers turned out to be a powerful approach to improve protein stability<sup>50</sup>.

Despite a large number of studies on the stabilization of protein-folds by rational or random approaches the proposition on strict design rules that would be applicable to new targets is difficult<sup>1,51</sup>. It rather became clear, that there are many structural ways that can be employed to increase stability<sup>52</sup>. However, some general conclusions can be drawn from the studies on protein stability. First, enzymes that were evolved for stability mostly carry mutations located on the surface of the protein, rather than in secondary structural elements. This might be due to a higher tolerance for structural changes as surface residues are rarely decisive for the protein fold, neither directly involved in catalytic activity<sup>1</sup>. Moreover, it has been suggested that high stability is dependent on rigidity<sup>53</sup>. This is unfavorable for enzyme activity and could explain the

fact why the same proteins, one derived from a thermophilic organism, the other from a mesophilic source, have the tendency to be similarly active at their respective optimal temperatures. In the thermophile protein the beneficial effect of temperature on the catalytic rates is counterbalanced by the increased rigidity<sup>54,55</sup>. This principle is less true for proteins that were evolved in the laboratory, as selection pressure for improvements in catalytic activity can be uphold while improving protein stability as well<sup>56</sup>.

### 1.3 The importance of protein-protein interactions

An important field for protein engineering is the design of novel protein binders. These proteins have become increasingly important over the last years for research, diagnostic- and therapeutic applications. Here, the importance of protein-protein interactions in general will be discussed. Further, an overview of antibody engineering and the development of alternative scaffolds will be given.

Proteins are essential for the function of an organism and participate in virtually every process within a cell. They mediate complex networks of interactions in all processes of life and molecular recognition by proteins ensures the specific interaction with other bio-molecules. An important example of interactions occurs during DNA replication, where large molecular machines are built through protein-protein interactions that carry out this important molecular process.

In general, protein-protein interactions can be classified in stable interactions, and transient interactions. Stable protein interactions occur when the proteins are part of big macromolecular machines. Transient interactions are generally less conserved but of highest importance for a cell, *e.g.* as part of a signal transduction network<sup>57</sup>.

Examples for a stable protein interactions are ATP synthase or protein transporters, such as the 50 MDa nuclear pore complex that selectively ensures the transport of cargo across the nuclear envelope<sup>58</sup>. These complexes have long lasting protein-protein interactions to build up the functional structures and usually dissociate only when degraded by the proteasome. On the other hand, transient interactions as in the membrane-associated heterotrimeric G proteins can shift their equilibrium between oligomeric and monomeric state by a molecular trigger. Upon guanosine-

triphosphate/-diphosphate exchange (GTP/GDP exchange) the affinity between the subunits  $G\alpha$  and  $G\beta\gamma$  of the G-protein complex changes about 1000-fold, which permits the complex to function as regulatory switch for the effective control of dynamic protein networks<sup>59</sup>.

All the above-mentioned examples show the importance of proteins and their role in living organisms. The wealth of proteins and the numerous possibilities for modification makes them very sensitive regulatory elements essential for the specific network signaling in living systems. This is why the elucidation of protein interaction-networks has become an important field of research<sup>60</sup> and large-scale determination of proteome-wide protein interactions using two-hybrid analysis has allowed great progress in understanding biological systems<sup>61-66</sup>. Interaction maps derived from these studies serve as unique resource for further analysis and use in the medical field. Over many years this knowledge has been useful for medical applications for the targeted disruption of intracellular signaling pathways by chemotherapeutic treatments using chemical small-molecules. More recently, the research on protein-protein interactions has led to the development of protein-based drugs, the so-called “biologics” and more and more of these new drugs are entering clinical trials. Protein scaffolds based on immunoglobulins were amongst the most successful ones, proven by the more than 20 different drugs that have been approved so far<sup>67</sup>. In parallel, important advances in the field of combinatorial protein design led to the development of a new generation of affinity proteins that are no longer based on the immunoglobulin scaffold<sup>68</sup>. These alternative proteins possess many advantages over classical antibodies and have an impressive therapeutic potential<sup>69</sup>.

### 1.4 Antibodies:

Antibodies are one of the most important protein classes of the immune system, essential for the identification and neutralization of foreign objects as bacteria or viruses. They recognize specifically a unique part of their target protein, the antigen, and bind to it with very high affinity. This remarkable capability is grounded in the architecture of antibodies, more precisely in the arrangement of polypeptide chains in the antigen-binding fragment (Fab)<sup>70</sup>. The antigen binding-site situated at the N-

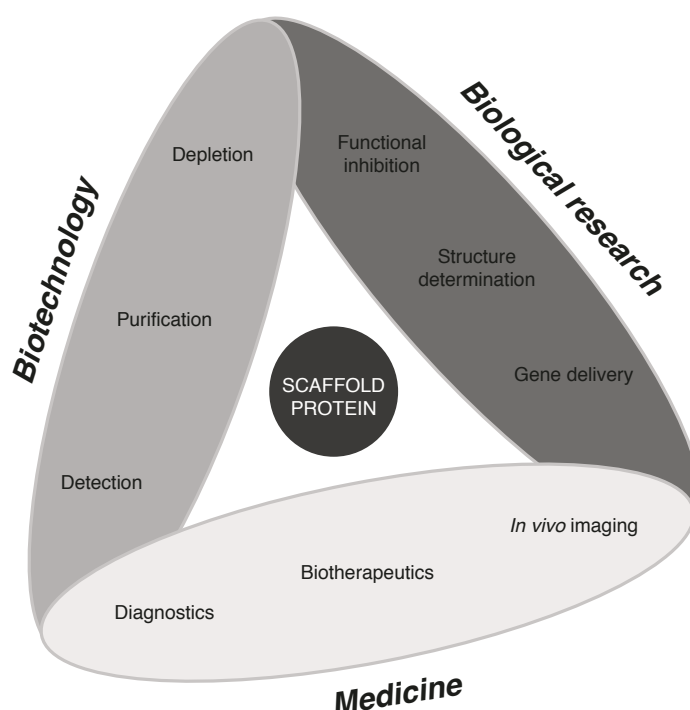
terminal end of the Fab-fragment contains the so-called Complementary Determining Regions (CDRs). These are flexible peptide loops interconnecting the intertwined beta-sheets that make up the antigen binding-site. These loops are exposed on the surface and have the capability to complement the antigen's shape, which determines the protein affinity and specificity. Due to somatic recombination, antibodies show extremely high variability of their amino acid sequence, especially in the CDRs. The large pool of possible amino acid combinations ensures that antibodies can be raised against virtually any possible antigen. Besides their important role in protecting host organisms against infections, antibodies turned out to be extremely useful for therapeutic applications and in biotechnology. Especially the development of techniques to produce monoclonal antibodies that can uniformly recognize one specific site on a target-protein increased the application of antibodies in medical therapy and biochemistry<sup>71,72</sup>. Typical fields of indications are rheumatoid arthritis, multiple sclerosis and different types of cancer but taking into account the recent development in antibody engineering many more indications are to come in the future<sup>73,74</sup>.

## 1.5 Alternative scaffolds:

The emergence of bioengineering techniques to screen large protein libraries accelerated the engineering efforts on diverse protein scaffolds different from antibodies. They have been successfully used to design very potent protein binders and are used as investigational tools, for diagnostics and therapeutic applications<sup>67,75-77</sup>(Figure 4).

As shown in elegant work by Plückthun *et al.* the use of an engineered form of ankrin-repeat motives yields very selective, high-affinity protein binders against diverse targets<sup>78,79</sup>. Designed Ank<sup>rin</sup>-Repeat Proteins (DARPin)s bind very specifically to the human epidermal growth factor receptor 2 (Her2). This receptor is an important target for cancer therapy and diagnosis and DARPin)s against Her2 have been used to determine the status of Her2 over-expression in different carcinomas and may also have a potential use in targeted therapy. Additionally, alternative protein scaffolds were developed. The most popular ones are monobodies<sup>80</sup>, Kunitz

domains<sup>81</sup> and anticalins<sup>82</sup>. Even though monobodies were derived from fibronectins and not immunoglobulins, their structure is based on a  $\beta$ -sandwich and thus shows a similar binding mode to that of antibodies<sup>80</sup>. Anticalins are not homologous to the immunoglobulin superfamily. Besides the selective binding to protein structures they can recognize molecules of lower molecular weight, as toxins or dioxigenin<sup>83,84</sup>. The discovery of a high structural plasticity in certain parts of the protein and an induced-fit upon target binding were two features so far being considered as typical for the binding sites of antibodies but were also found in anticalin scaffolds<sup>67</sup>. Kunitz domains are composed of a disulfide rich alpha-/beta-fold and have a size of about 6 kDa. Due to their long peptide loop gathering at the tip of this wedge-shaped proteins, Kunitz domains tend to bind targets with concave surfaces such as clefts for enzyme substrates or pockets for natural ligands<sup>85,86</sup>.



**Figure 4:** Application possibilities for non-immunoglobulin scaffolds that were engineered for specific protein binding by diverse protein-engineering approaches. Figure taken from Grönwall *et al.*<sup>87</sup>.

These examples show how engineered proteins derived from different protein scaffolds have become valuable tools for medical applications, diagnosis, molecular biology research and biotechnology. The creation of new target-specific binding



proteins through the creation of novel protein libraries and the application of diverse screening techniques has been very successful in the past. With the development of even more sensitive systems and the help of structural data coming from crystallographic and computational efforts, their range of application will continue to grow<sup>67</sup>. An overview of the most prominent screening techniques will be given in the next paragraph.

## **1.6 Overview of selection systems used to identify new affinity proteins:**

Despite our knowledge about proteins the rational design of new proteins is still a difficult task. The *in silico* prediction of the effects of various mutations is challenging and experimental re-optimization is needed in most cases<sup>51</sup>. To identify protein variants with improved functions, the screening of protein libraries has been the most promising approach over decades. Improvements in protein stability, reactivity or substrate specificity, as well as the isolation of proteins with novel binding specificities have been achieved by these selection systems. All methods share the common principle to link the phenotype of the protein with its genotype. As sequencing of proteins is still difficult this principle enables easy identification and amplification of the selected polypeptides via the nucleic acids. All selection systems follow the principle of diversification, selection and amplification: After the creation of a protein library the proteins that show the new property are selected, amplified and tested. It is possible that several rounds of diversification, selection and amplification are done prior to characterization in order to enrich for the right proteins. The different selection systems described below focus on the selection of new affinity proteins and can be divided in three categories; cell-dependent display systems, cell-free display systems and non-display systems.

### **1.6.1 Cell-dependent systems:**

In cell-dependent systems as phage-display<sup>88</sup>, the library proteins are displayed on the surface of cells or in a cellular compartment. It is possible to select positive clones with powerful flow cytometric techniques without any need for elution from the target protein. Major disadvantages are the limitation of library size due to the

transformation efficiencies of DNA and the need for the target protein to be available in large quantities, preferably in its native form.

### **Phage display.**

Phage display uses bacteriophages to connect the proteins with the encoding genetic information. Most commonly filamentous phages are used<sup>87</sup>. For the display of foreign proteins on the surface of phage the gene encoding the protein of interest is fused to the phage gene of one of the major coat proteins pVIII or pIII. pVIII is produced in 2700 copies and covers the whole phage particle and only short peptides are tolerated in selection<sup>89</sup>. pIII is only present in five copies at the tip of the phage and can be used for the fusion of larger proteins, as antibodies. During virus assembly the produced fusion proteins are incorporated into the phage particle enabling the protein display on the phage surface and the selection for affinity binders against the protein of interest<sup>90</sup>. Assembly of phage particles displaying the protein library is done by infection of *E.coli* cells carrying the DNA information of the library with helper phage. After incubation of phage particles with the target protein interacting phage is captured and unbound phage washed away. After elution the phage is used to infect new *E.coli* cells to amplify selected clones. This panning process is typically repeated between three- to four times to insure a good enrichment of high affinity binders.

The technology was originally discovered in 1985 by Smith *et al.*<sup>91</sup> He demonstrated that selection for peptides that were genetically fused to the gene of pIII of filamentous phage was possible, if an antibody against these peptides was available. Since then, this technology has been further developed and improved and was successfully used, *i.a.* antibody engineering<sup>92-98</sup>. The method has equally been used for display of libraries of alternative non-immunoglobulin protein scaffolds<sup>78,99-101</sup> and for drug discovery<sup>102-104</sup>.

### **Yeast display.**

Yeast display has been one of the first alternatives to phage display, first reported by Boder and Wittrup<sup>105</sup>, its major advantage being the possibility of mammalian protein

expression in a eukaryotic host system. Similar to phage display, proteins or peptides are displayed on the yeast surface. Therefore, the proteins are displayed as fusion proteins with the Aga2p subunit, a protein needed for the yeast to mediate cell-cell contacts during the mating process. Aga2p attaches to the yeast cell wall by interaction with the Aga1p protein. For yeast display, both proteins are under the control of inducible promoters. After induction, approximately  $10^4$  to  $10^5$  copies of the recombinant proteins are displayed per yeast cell<sup>105,106</sup>. A big advantage of yeast display over other selection techniques is the possibility to use Fluorescence-Activated Cell Sorting (FACS) for the isolation of antigen-binding cells. For this, fluorescent-labeled target is added to the cell-displayed protein library prior to FACS screening. This method allows the monitoring and quantification of the relative affinities for the target of each library member already during the selection process. The introduction of two protein-tags, an N-terminal HA-tag and a C-terminal c-myc-tag in fusion to the displayed proteins, allows for the quantification of full-length expression efficiency and the amount of displayed protein on the yeast surface<sup>107</sup>.

Due to a lower transformation efficiency in yeast compared to *E.coli* the construction of large naïve libraries can be more difficult. Differences in the glycosylation pattern of yeast-expressed proteins compared to mammalian cells can be disadvantageous when using this technology, but it has not limited the success of the approach for a number of applications. For instance, using this technology, Buonpane *et al.* succeeded in the isolation of a femtomolar affinity binder<sup>108,109</sup>. It has also been shown that the combination of directed evolution and yeast display could improve the affinity of an integrin  $\alpha$ -subunit to the intracellular adhesion molecule 1 by 200,000-fold<sup>110</sup>.

### **Bacterial display.**

Similarly to yeast cells, bacteria can also be used for displaying proteins. Mostly used host system is a Gram-negative bacterium *E.coli*, its major advantages being the rapid growth rate, easy handling and the possibility of making large libraries of up to  $10^{11}$  different variants being screened<sup>111</sup>. The use of gram-positive bacteria as *Staphylococcus carnosus* have been reported<sup>112,113</sup>. Major limitation of this method is the presentation of correctly folded protein on the cell-surface. Problems can arise due to the choice of carrier protein, the amount of disulfide-bonds, as well as the size of

the displayed protein. For the facilitated display of larger, more complex molecules on *E.coli* a variation of the classical bacterial display called the Anchored Periplasmic Expression (APEX) has been developed<sup>114,115</sup>. For this, proteins are displayed on the periplasmic-side of the inner membrane. During the selection process, the membrane is disrupted or permeabilized and the fluorescent-labeled ligand can bind to the displayed protein. As for yeast display, sorting of positive clones via FACS analysis is the method of choice. Bacterial display was recently applied to the successful isolation of a high affinity affibody targeting TNF- $\alpha$  from a naïve affibody library previously enriched with one round of phage display<sup>116</sup>. The successful application of gram-positive bacteria *S. carnosus* for large-scale epitope mapping of antibodies through the display of antigen libraries on the surface of the cells and subsequent FACS analysis has been reported<sup>117,118</sup>.

### 1.6.2 Cell-free systems:

Large libraries with up to  $10^{13}$  mutants can be screened using cell-free systems as ribosomal display or mRNA display. Library construction is based on *in vitro* transcription and translation, which also enables the possibility to introduce *in vitro* mutagenesis during amplification rounds.

#### **Ribosomal display.**

In order to circumvent restrictions in library size due to limitations in DNA transformation efficiencies, cell-free systems as the ribosomal display have been developed<sup>119,120</sup>. The crucial point in this system is the fusion of the library DNA to a spacer sequence lacking a stop codon. During *in vitro* translation this spacer sequence stays attached to the peptidyl tRNA and occupies the ribosomal tunnel. As this sequence is C-terminal to the amino acid sequence of the library protein it has already been translated that it protrudes out of the ribosome and folds. The resulting complex of protein, mRNA and ribosome can then be subjected to panning against the target protein. Low temperatures and high  $Mg^{2+}$  concentration can further stabilize the mRNA-ribosome-protein complex<sup>121</sup>. After elution of bound protein from the target, the mRNA can be reverse transcribed to cDNA and further analyzed. At this stage, an additional round of mutagenesis can be introduced to increase the selective pressure.

**mRNA display.**

A higher stability of the mRNA-protein complex can be reached by mRNA display, where the mRNA is covalently linked to the polypeptide through the antibiotic puromycin<sup>122</sup>. The structure of puromycin resembles the 3' end of an aminoacyl-tRNA molecule and can readily enter the ribosomal A site and be incorporated into the nascent peptide. The mRNA-polypeptide fusion is then released from the ribosome.

The possibility to use large libraries with up to  $10^{15}$  different members is a strong advantage of the *in vitro* translation methods<sup>123</sup>. Non-natural amino acids can be used during translation and *in vitro* mutagenesis can be easily incorporated during the individual selection rounds to increase the genetic variability in the sequence pool. The covalent binding and the smaller size of the linker fragment puromycin are advantageous over mRNA-protein linkage via the ribosome as in ribosomal display<sup>124</sup>.

**1.6.3 Yeast two-hybrid:**

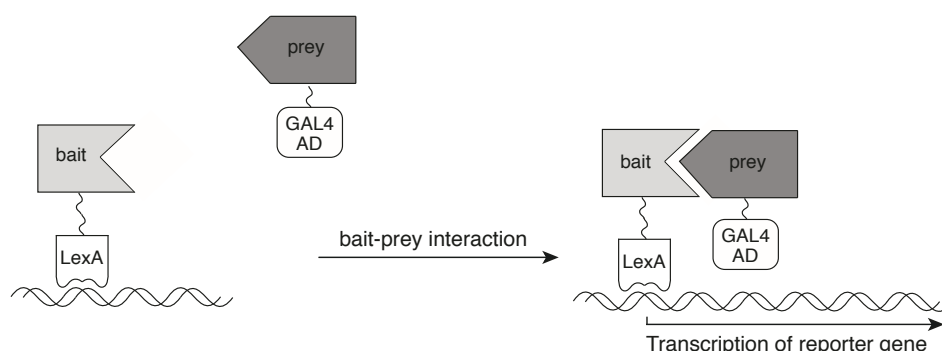
An alternative system for the direct evaluation of protein-protein interactions is the yeast two-hybrid (Y2H) screening that was first applied by Chien *et al.* in 1991<sup>125</sup>. The system is based on the modular organisation of transcription factors that when in close proximity trigger reporter gene transcription. The first protein to be used for this was the yeast protein GAL4. Nowadays, the DNA-binding domain of the *E.coli* protein LexA is used. GAL4 has a DNA-Binding Domain (BD) and an Activation Domain (AD). When GAL4 binds to its cognate binding site, the activation domain is brought close to the promoter, allowing the activation domain to interact with the transcription machine and gene transcription.

The first basic experiment that opened the way to library screening was the discovery that the N-terminal fusion of the DNA repressor LexA from *E.coli* to the yeast activation-domain GAL4 could trigger protein transcription, but only when a LexA operator was present near the transcription starting site<sup>126</sup>.

Further pioneering work was done by Fields *et al.* as they used GAL4 transcriptional activation for the detection of protein-protein interactions<sup>127</sup>. They showed that the reconstitution of proximity of the GAL4 BD and the GAL4 AD could be mediated via

the interaction of the proteins SNF1 and SNF4. The interaction of these two fusion proteins would lead to the activation of a reporter gene downstream of the GAL4 binding site. The authors envisioned to replace one fusion protein by a genomic library of cDNA sequences fused to the GAL4 AD. A target protein could then be fused to the GAL4 BD. In case of protein interaction both GAL4 domains would be brought in close proximity and trigger reporter gene transcription. Cells expressing two interacting proteins would then be selected based on reporter gene transcription (auxotrophic marker or colorimetric assay). The first Y2H screening against Sir4-binding proteins was published two years later<sup>125</sup> in which the authors confirmed the known homodimerization of Sir4.

Since then, Y2H screenings have become a highly used method for the detection of protein-protein interactions (Figure 5) and large-scale protein interaction mapping projects have been realized for several organisms (*S.cerevisiae*<sup>62,66</sup>, *C. elegans*<sup>63</sup>, *Drosophila melanogaster*<sup>61</sup> and recently human<sup>65,128</sup>).



**Figure 5:** Yeast two-hybrid detection principle. Bait-LexA DNA-binding domain fusion protein is co-expressed with prey-GAL4 activation domain fusion protein in yeast cells. In the case of bait-prey interaction, the transcription of a reporter gene (e.g. HIS3 reporter gene) is activated. The transcription product allows the yeast cell to survive on auxotrophic media and colonies of the interacting protein pair will be detected in a screening.

A crucial point for these large-scale screening projects is the quality of the cDNA libraries. Historically, cDNA libraries have been produced by more or less random fragmentation of genomic DNA or random primed cDNAs were prepared from the mRNA of diverse tissues<sup>129</sup>. The disadvantage of libraries created in this way is the uncontrolled fashion in which the coding sequences of the inserts are attached to the coding sequence of the transcriptional activator. Some consequences of this approach

include the expression of the wrong reading-frame or from sequences of the untranslated regions of the mRNA. The resulting non-natural proteins provide a rich source for non-specific interactions that add to the number of false positives. Recently, more and more cDNA libraries comprising the full-length open reading frame (ORF) of the mRNA are generated for several species<sup>130-132</sup>, in part in dedicated efforts to provide new resources for Y2H screening<sup>133-135</sup>.

A modification of the yeast two-hybrid screening is the yeast three-hybrid system that has been used successfully for the identification of small molecule-protein interactions. In this assay the two fusion proteins can only interact via a small molecule-anchor. Studies using methotrexate-DHFR<sup>136</sup> or  $\beta$ -estradiol-biotin anchors<sup>137</sup> to identify new protein-drug interactions have been published.

Similar systems based on alternative reporter interactions have also been developed, as the split-ubiquitin assay that is based on the reconstitution of split-ubiquitin leading to the liberation or the reporter fragment in case of protein-protein interaction<sup>138</sup> or small molecule-protein interaction<sup>139</sup>. The MAmmalian Small Molecule-Protein Interaction Trap (MASPIT)<sup>140,141</sup> permits the detection of both modification-independent and phosphorylation-dependent interactions in human cells.

## 1.7 AIMS OF THE PROJECT:

Aim1: SNAP-tag is more reactive towards BG-derivatives and more stable than wild-type hAGT. In this work, the structural differences between these two proteins were studied in more detail. Crystal structures of hAGT and SNAP-tag in their benzylated and non-benzylated state were the starting point from which further studies on the structure-function relationship were developed. The overall aim of this work was to obtain a detailed understanding of the impact of individual point mutations on SNAP-tag stability and reactivity.

Aim 2: The second part of this work was focused on protein engineering. The aim was to create a SNAP-tag-based inhibition complex, which interacts with a target protein through amino acid loops and decreases its function only upon labeling with BG-inhibitor molecules. Therefore, randomized amino acid loops were introduced at distinct positions of SNAP-tag and mutants were selected for small molecule dependent binding-affinity using Y3H-screening. Identified hits were tested for *in vitro* binding and target protein inhibition.

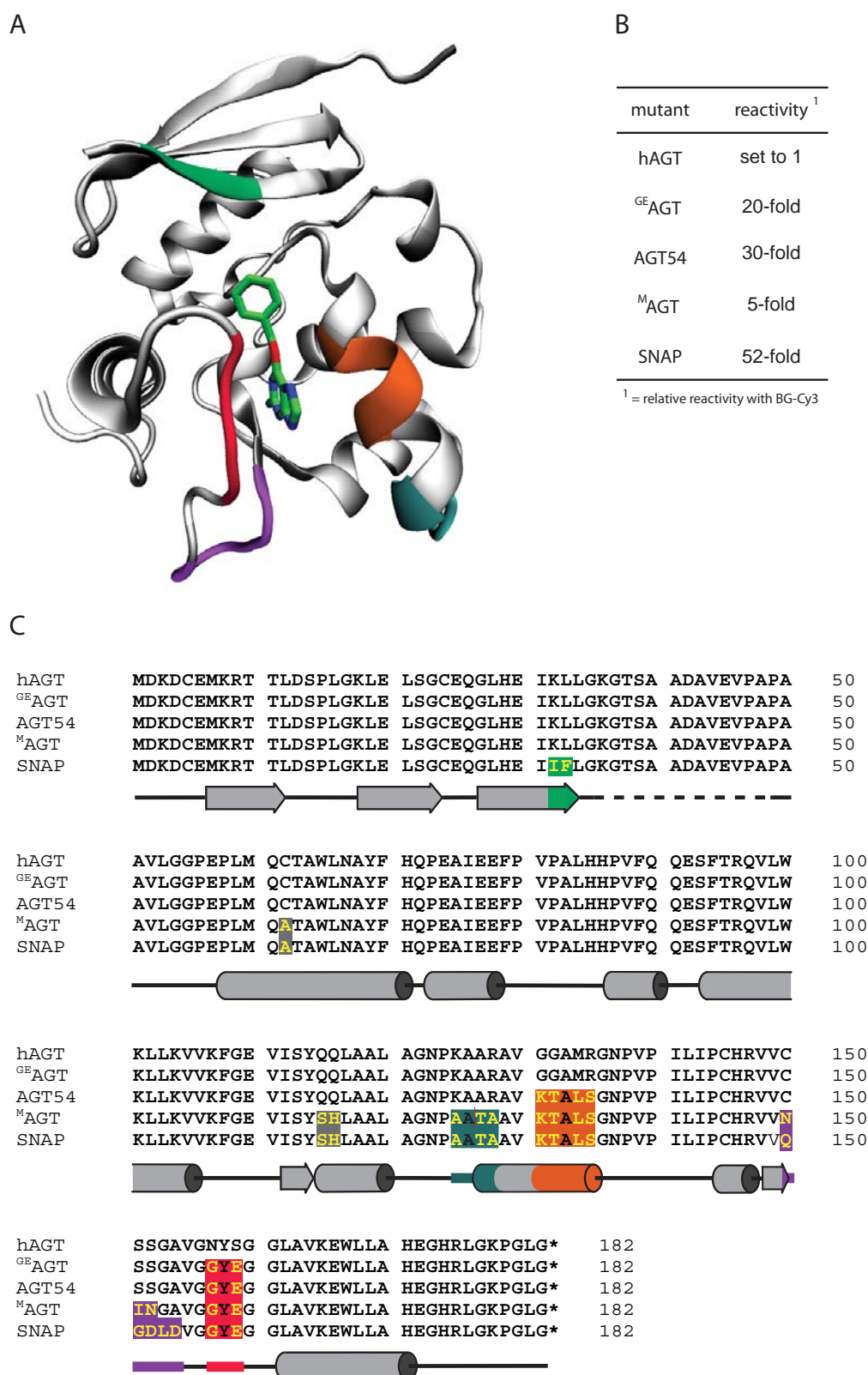


## **2 Directed evolution of hAGT to SNAP-tag – a study on the structure-function relationship**

In this chapter a biophysical, structural and computational analysis of the directed evolution of hAGT to SNAP-tag is presented. Relative to hAGT, SNAP-tag possesses a 52-fold higher reactivity towards BG derivatives, does not bind to DNA and expresses well in cells as on cell surfaces. However, we had a low understanding of how the introduced mutations affected protein activity and protein stability. We therefore performed a detailed study to understand the effect of the introduced mutations in terms of protein stability and reactivity. The results are presented in this chapter.

### **2.1 Structural considerations**

SNAP-tag was generated in a stepwise manner from human O<sup>6</sup>-alkylguanine-DNA alkyltransferase (hAGT) by introducing a total of 19-point mutations (Figure 6) and deleting 25 C-terminal residues. Saturation mutagenesis of four active-site residues followed by phage display and selection for activity against BG derivatives resulted in <sup>GE</sup>AGT, a mutant with 20-fold increased activity towards such substrates (Figure 6B)<sup>34</sup>. Subsequent saturation mutagenesis of four additional residues involved in substrate binding followed by phage selections resulted in AGT-54, a mutant with 1.5-fold higher activity than <sup>GE</sup>AGT.



**Figure 6:** Overview of the directed evolution of hAGT to SNAP-tag: (A) Crystal structures of SNAP-tag C145A mutant co-crystallized with BG. Mutagenized parts of the protein are highlighted in color. (B) Relative reactivity of hAGT and mutants with BG-Cy3. (C) Sequence alignment of hAGT with intermediate mutants and SNAP-tag. Colors that highlight mutations correspond to those used in (A).

To further optimize the protein for applications in protein labeling, mutations were introduced to suppress DNA binding and reactivity towards nucleosides, to remove non-essential cysteines and to truncate the last 25 residues<sup>35</sup>. The resulting mutant <sup>M</sup>AGT displayed relatively low activity towards BG derivatives (Figure 6B) (5-fold higher than hAGT). To rescue the activity of <sup>M</sup>AGT against BG derivatives, an additional round of saturation mutagenesis (residues 150-154 and 32-33) followed by phage display was performed, resulting in SNAP-tag<sup>142</sup>.

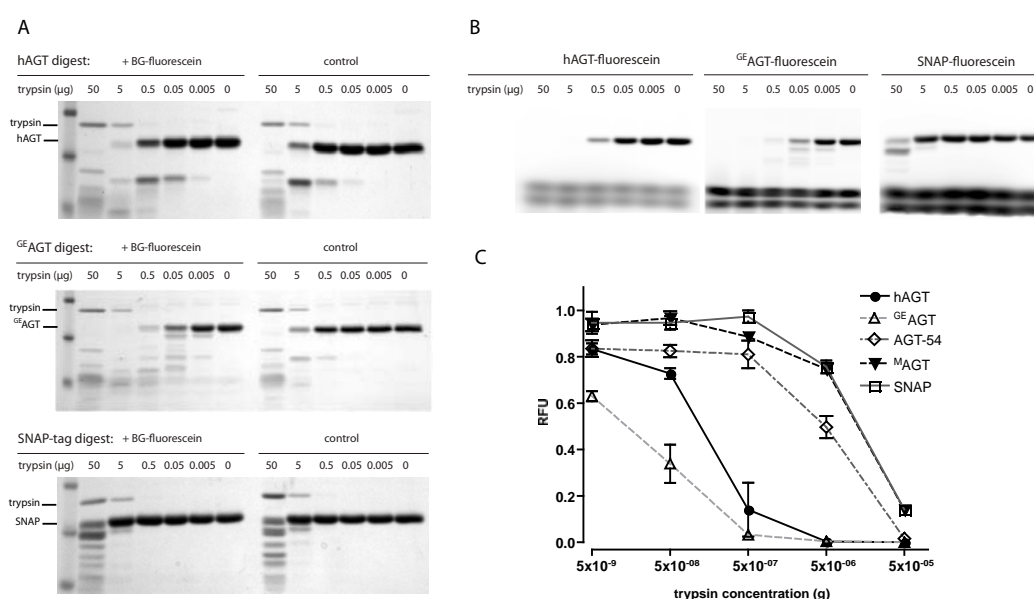
A unique feature of O<sup>6</sup>-alkylguanine-DNA alkyltransferases is that the protein is not regenerated but degraded after DNA repair<sup>38</sup>. In mammalian cells, it is believed that alkyl transfer triggers a conformational change in hAGT, leading to ubiquitination and degradation of the alkylated protein<sup>38</sup>. Supporting this hypothesis is that alkylation of hAGT increases its sensitivity towards proteolysis *in vitro*<sup>40</sup>. Furthermore, structural analysis of hAGT before and after alkylation revealed that alkylation leads to sterically unfavorable interactions that result in partial unfolding of the protein<sup>143</sup>. Interestingly, an increased degradation of SNAP-tag fusion proteins after labeling has not been observed<sup>144</sup>.

## 2.2 Studies on protein stability

The directed evolution of hAGT to SNAP-tag was done in several evolutionary steps. The major selection criteria were changes in reaction speed. In order to understand the impact on protein reactivity and protein stability in more detail a structural study has been performed using experimental as well as computational methods to analyze the impact of the individual point mutations. All computational experiments have been done in cooperation with the laboratory of Prof. U. Rothlisberger, by E. Brunk and are presented here for further insight on the experimental results.

### 2.2.1 Proteolysis experiments:

It has been previously reported that alkylation of wild-type hAGT increases the susceptibility of the protein towards proteases such as trypsin<sup>40</sup>. The increased protease susceptibility can be interpreted as a decreased stability of the alkylated protein. As a first measure of how the stability of hAGT changed in the course of its directed evolution into SNAP-tag, we therefore measured the susceptibility of the different hAGT mutants towards trypsin before and after labeling with BG-fluorescein.



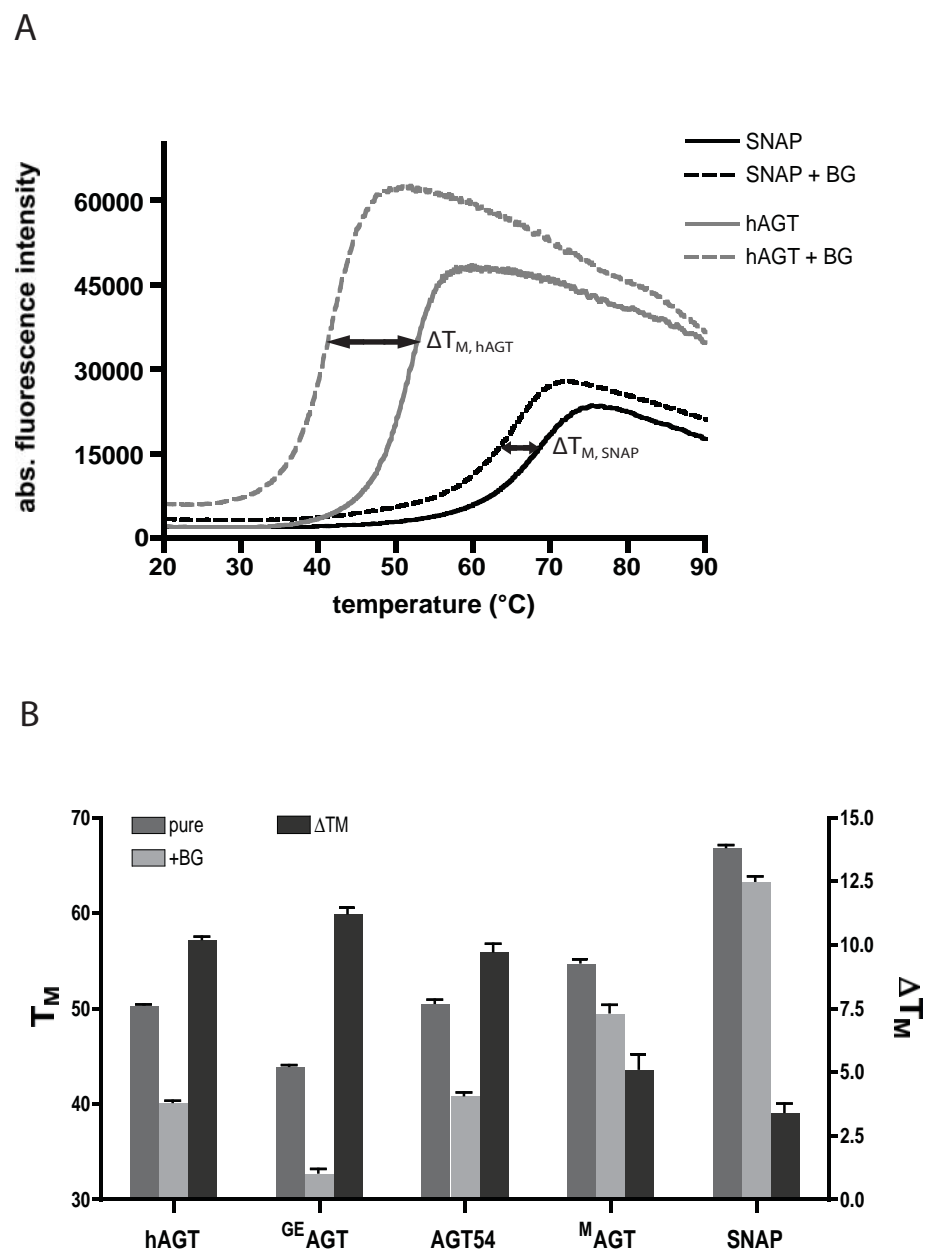
**Figure 7:** Determination of protein stability by trypsin digestion. Strong differences were observed for hAGT, SNAP-tag and intermediate mutants. (A) SDS-PAGE of hAGT, <sup>GE</sup>AGT and SNAP-tag digestion mixes. (B) Fluorescence gel-scan used for quantification. (C) Quantification of fluorescence gel band intensities of all mutants at different trypsin concentrations.

Proteins were titrated with increasing amounts of trypsin and analyzed on SDS-PAGE by coomassie staining (Figure 7, A) and by quantification of in-gel fluorescence (Figure 7, B). As reported previously, hAGT showed an increased susceptibility towards trypsin, especially upon labelling with BG-fluorescein. In contrast, no increased sensitivity of SNAP-tag could be observed. Furthermore, an almost 100-fold higher trypsin concentration had to be used to degrade SNAP-tag to a similar extent as hAGT, indicating an increased stability of SNAP-tag. Mutations introduced in <sup>GE</sup>AGT (Asn157Gly, Ser159Glu) had a destabilizing effect, especially on the labeled protein. However, the trypsin resistance significantly increased for AGT-54

and <sup>M</sup>AGT. These data showed that during the directed evolution of hAGT for higher reactivity towards BG the stability of the protein, especially after benzylation increased drastically. It should be noted that the mutations introduced throughout the directed evolution of hAGT might not only affect the susceptibility towards trypsin by changing the stability of the protein but also via the introduction or removal of trypsin cleavage sites. We therefore sought to confirm these results through an independent assay.

### **2.2.2 Thermal denaturation assay:**

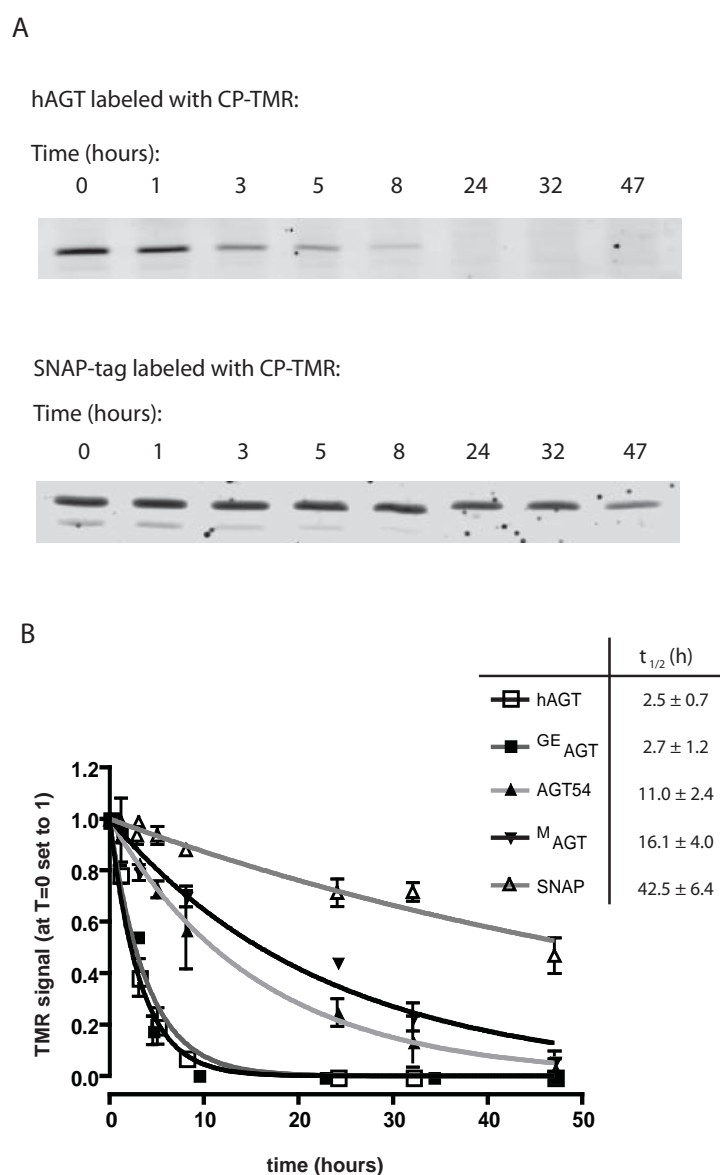
In order to confirm independently the results obtained in the proteolysis experiment, we evaluated the stability of hAGT and mutants by thermal denaturation <sup>145</sup>. The melting temperatures ( $T_M$ ) were determined with help of a fluorogenic molecule (SYPRO Orange®) that shows a strong increase in fluorescence upon binding to hydrophobic regions of a protein. Fluorescence intensity reached a maximum and then started to decrease, probably due to precipitation of the complex of the fluorescent probe and the denatured protein <sup>146</sup>. We determined the melting temperatures of all proteins in the presence and absence of BG to assess the impact of benzylation on protein stability (Figure 8). We observed that the  $T_M$  of SNAP-tag was increased by 17 °C relative to hAGT. The directed evolution for higher reactivity towards BG derivatives led at the same time to an increased stability of the labeled protein relative to the unmodified protein (see  $\Delta T_M$  in Figure 8, A and B). A 10.2 °C decrease in melting temperature was observed for benzylated hAGT, whereas  $T_M$  of SNAP-tag was reduced by only 3.5°C upon benzylation. These results were in agreement with the data obtained for the proteolysis experiments and could confirm the increased stability throughout protein evolution.



**Figure 8:** Melting point ( $T_M$ ) measurement for hAGT, SNAP-tag and mutants. (A) Raw data of thermofluor assay presented for hAGT and SNAP-tag before and after reaction with BG. Differences in  $\Delta T_M$  before and after benzylation are highlighted by arrows. (B) Melting point analysis. Shown are  $T_M$  of pure protein (grey) and benzylated (light grey) as well as  $\Delta T_M$  before and after benzylation (dark grey).

### 2.2.3 Stability of labeled hAGT and SNAP-tag in living cells:

Our data on the sensitivity towards proteases and thermal denaturation showed a significantly increased stability of SNAP-tag relative to wild-type hAGT, in particular for the labeled protein. As hAGT is known to be rapidly degraded upon alkyl transfer in living cells<sup>38</sup> we investigated to what extent the increased *in vitro* stability of SNAP-tag would also translate into an increased intracellular half-life of the labeled protein. Therefore, we performed pulse-chase experiments with all mutants (Figure 9).



**Figure 9:** Pulse-chase experiment of hAGT, SNAP-tag and mutants in HEK 293 cells. (A) Fluorescence gel scan of SDS-PAGE of hAGT and SNAP-tag pulse-chase experiments. Samples were analyzed at indicated time points for amount of fluorescence labeled protein present in each sample. (B) Plot of the relative fluorescence signal intensities for analyzed proteins as a function of time.

Suspension cultures of HEK 293 cells expressing the different mutants were incubated for 15 minutes with 0.65  $\mu$ M CP-TMRstar (NEB) and the reaction was quenched with 100  $\mu$ M of BG. Aliquots were taken at distinct time-points and analyzed on SDS-PAGE and by fluorescence gel scanning (Figure 9, A). Labeled hAGT showed a half-life of around 3 hours, whereas the apparent half-life of labeled SNAP-tag was determined to be approximately 42 hours (Figure 9, B). It should be noted that the actual half-life of SNAP-tag should be even higher, since the observed signal decrease results most likely from a dilution of labeled protein due to continuous cell growth. With respect to the intermediate mutants, labeled <sup>GE</sup>AGT showed similar stability as wild type hAGT (half-life of 2.7 hours), whereas AGT-54 showed an increased half-life of 11 hours and mutations introduced in <sup>M</sup>AGT resulted in a further prolonged the half-life of 16 hours (Figure 9, B). These data clearly show that labeled SNAP-tag is not degraded to a significant degree in living cells and that the directed evolution for higher reactivity resulted in the generation of mutants with increased *in vitro* stability that translates into an increased intracellular protein half-life.

### 2.3 Structural analysis:

In order to obtain further insights into how the mutations influenced the reactivity and stability of SNAP-tag, we obtained the crystal structures of (i) SNAP-tag, (ii) SNAP-tag mutant Cys145Ala with BG bound and (iii) benzylated SNAP-tag. These structures were solved at 1.9 Å, 1.89 Å and 1.7 Å resolution, respectively (pdb entries 3KZY, 3KZZ and 3LOO). The mutation C145A of the active site cysteine was necessary to prevent the reaction of protein with its substrate.

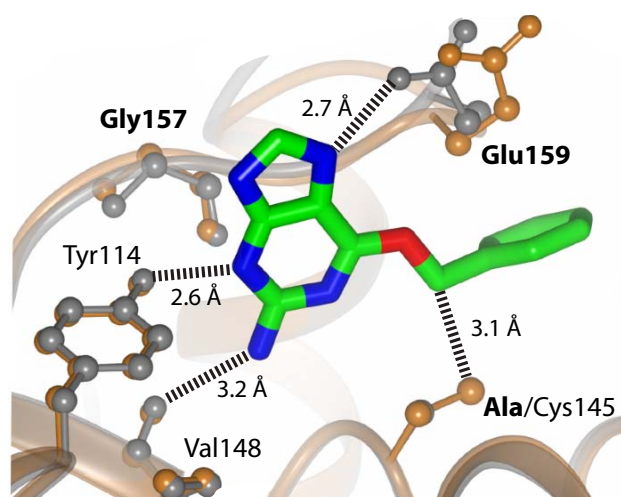
Like hAGT, SNAP-tag consists of two domains and in all three SNAP-tag structures a Zn<sup>2+</sup> ion was located in the N-terminal domain. As observed in the hAGT structure, a large flexible loop within the N-terminal domain of SNAP-tag was not resolved in the electron density maps (SNAP-tag: Lys36-Pro49, SNAP-tag benzylated: Gly35-Leu53, SNAP-BG: Gly35-Leu53). A comparison of the backbone structures of hAGT (PDB entry 1EH6) and SNAP-tag (PDB entry 3KZY) gave rise to a root mean square deviation (rmsd) of 0.794 Å calculated between C $\alpha$ -positions for residues 1-182 of SNAP-tag and hAGT (EMBL-EBI, Secondary Structure Matching).



### 2.3.1 Substrate Binding:

In all crystal structures, hAGT and SNAP-tag, Tyr114 and Val148 form hydrogen bonds to the purine and position the alkylated base for efficient alkyl transfer in the active site (Figure 10) <sup>143</sup>. An overlay of SNAP-tag with SNAP-tag Cys145Ala co-crystallized with BG shows that the thiol of Cys145 is in a 3.1 Å distance to the CH<sub>2</sub> of the benzyl ring and ideally positioned for alkyl transfer.

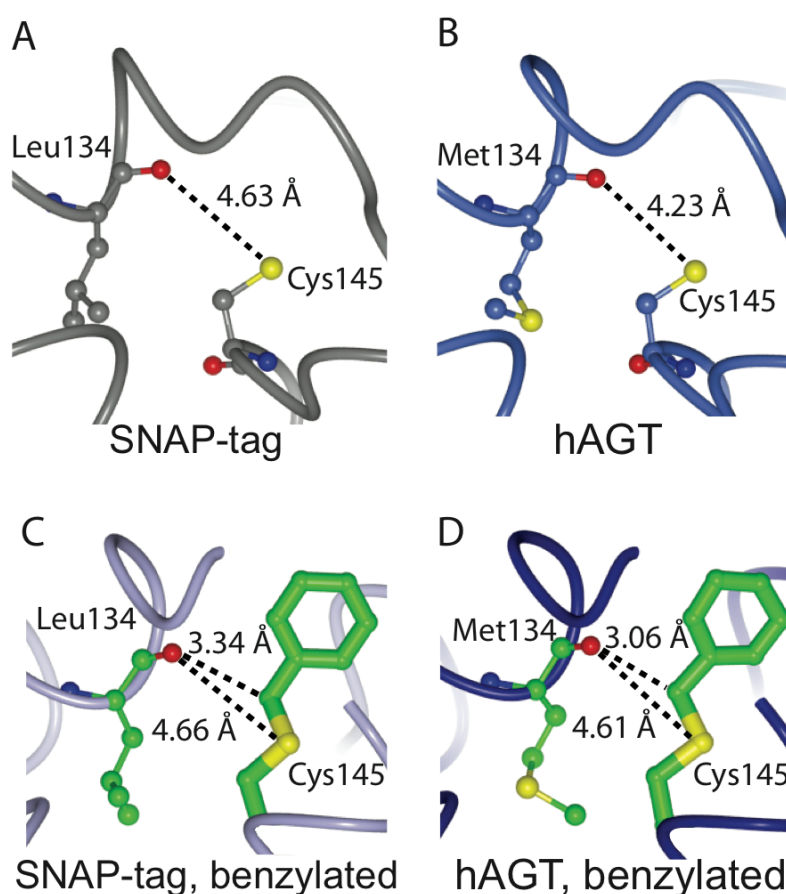
In <sup>GE</sup>AGT mutations introduced at positions 157 and 159 (see red segment in Figure 6A, C) increased significantly the reactivity towards BG. In particular, the introduction of Glu159 increased reactivity 20-fold (see <sup>GE</sup>AGT). We speculated previously that Glu159 would form a hydrogen bond with N7 of BG <sup>34</sup>. Indeed, in the crystal structure of SNAP C145A with bound BG, one of the carboxylate oxygens of Glu159 is within 2.7 Å of N7 of BG. In addition, C $\beta$  and C $\gamma$  of Glu159 make hydrophobic contacts with the benzyl ring of BG (distance approximately 3.7 Å), which is located between the glutamate side chain and Pro140 with an additional edge-on hydrophobic interaction with Tyr158. These interactions should facilitate substrate binding and stabilize the leaving group in the S<sub>N</sub>2 reaction. It is noteworthy that in the structure of SNAP-tag the side chain of Glu159 shows an alternative conformation that is not present in the structure of benzylated SNAP-tag, suggesting that binding of BG fixes the conformation of this residue.



**Figure 10:** Active site of SNAP-tag C145A mutant (in grey) co-crystallized with BG; free SNAP-tag (in gold) is overlaid to highlight Cys145 positioning in reference to BG. Ala145, Gly157 and Glu159 (in bold) have been mutated compared to wild-type hAGT. Tyr114, Val148 and Glu159 form hydrogen bonds with the substrate. In the presence of BG Glu159 changes position and forms hydrogen bonding interaction with BG, C $\beta$  and C $\gamma$  of Glu159 make hydrophobic contacts with the benzyl ring of BG. Cys145 is in 3.1 Å distance to the CH<sub>2</sub> of the benzyl ring and ideally positioned to perform the alkyl transfer.

### 2.3.2 Increased stability:

It has been suggested that after alkyl transfer in hAGT, sterically unfavorable interactions between the benzylic CH<sub>2</sub> group and the carbonyl oxygen of Met134 cause a displacement of alpha helix 127-136<sup>147</sup>. This displacement of the alpha helix triggers unfolding of the alkylated protein. In the structures of hAGT and benzylated hAGT, the displacement of helix 127-136 upon alkylation also manifests itself in an increased distance of the sulfur of Cys145 and Cys145benz to the carbonyl oxygen of Met134 from 4.2 to 4.6 Å (Figure 11B and D). In SNAP-tag, structural changes in the active site avoid unfavorable steric interactions upon benzylation (Figure 11 A and C).



**Figure 11:** Distances between carbonyl oxygen of Met134 (in hAGT) or Leu134 (in SNAP-tag) and sulfur of Cys145 (A and B) or sulfur of benzylated Cys145 and CH<sub>2</sub> of benzyl ring (C and D). In SNAP-tag distances upon benzylation remained virtually unchanged contrary to hAGT where the distance increases due to the movement of alpha helix 127-136 and the close proximity especially to the Met134 carbonyl group.

When overlaying the structures of SNAP-tag and benzylated SNAP-tag, a displacement of alpha helix 127-136 cannot be detected. Also, the distance between the sulfur atom of Cys145 and the carbonyl oxygen of Leu134 is 4.6 Å in SNAP-tag and this distance remains virtually unchanged upon benzylation (Figure 11 A and C). The absence of structural information on single point mutants makes an unambiguous identification of the key residues responsible for this structural change difficult, but we believe that particular mutations in the alpha helix 127-136 are important. Mutations introduced at this position of SNAP-tag shorten hydrogen bonds within the alpha helix considerably (Table 1): the distance between O131 and N135 is 3.0 Å in SNAP-tag but 3.8 Å in hAGT. Glycine residues are known to destabilize alpha helices if located in the middle of a helix<sup>148,149</sup> and the mutations in SNAP-tag result in a more compact alpha helix and increased available space in the active site.

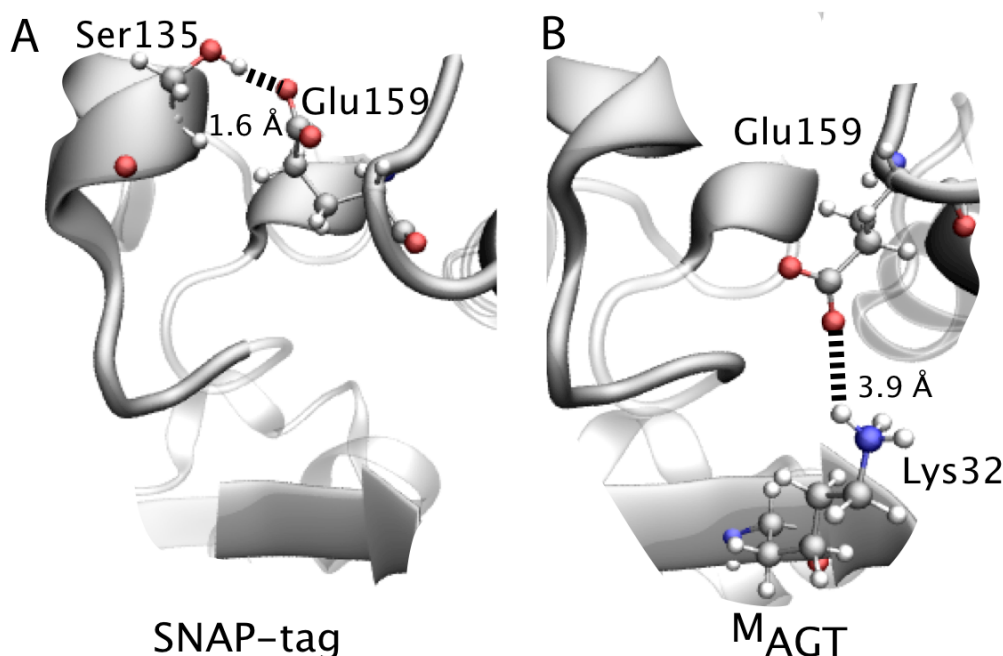
	hAGT [Å]	hAGT, benzylated [Å]	SNAP [Å]	SNAP, benzylated [Å]
O126 – N130	2.96	2.99	3.00	2.93
O127 – N131	2.98	2.94	3.01	2.98
O128 – N132	3.07	3.01	2.99	3.05
O129 – N133	2.94	2.80	2.95	2.89
O130 – N134	3.04	3.09	2.89	2.84
O131 – N135	3.81	3.87	3.03	3.03
O132 – N136	3.33	3.73	3.17	3.35
O133 – N137	3.25	3.54	3.19	3.48

**Table 1:** Length of hydrogen bonding in alpha helix 126-137 for hAGT, benzylated hAGT, SNAP-tag and benzylated SNAP-tag. Mutations introduced in SNAP-tag lead to a shortening of the alpha helix hydrogen bonds, especially the distance of O131 – N135 changes a lot comparing hAGT with SNAP-tag.

Especially Gly131 but also Gly132 are well conserved among O<sup>6</sup>-alkylguanine-DNA alkyltransferases<sup>150</sup> and it has been suggested that for steric reasons the flipping of alkylated bases out of double-stranded DNA requires glycines at these positions<sup>143,151</sup>. Our results suggest that Gly131 and Gly132 in hAGT also play an important role in triggering protein unfolding and degradation upon alkyl transfer.

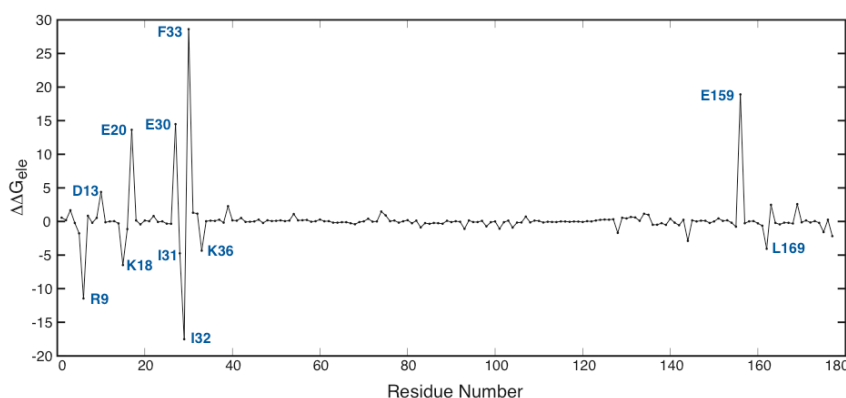
### 2.3.3 Molecular Dynamics Simulations:

In order to understand further the impact of specific mutations introduced in SNAP-tag on protein stability and reactivity, we used Molecular Dynamics (MD) methods to study both, the benzylated and non-benzylated systems. For this, we used a method using free energy perturbation in conjugation with thermodynamic integration (FEP/TI) to address specific stabilizing versus destabilizing interactions by determining the contribution of a mutation to the overall change in free energy for each residue. In particular, we focused on the specific mutations at positions (i) 157/159, (ii) 32/33 and (iii) 150 through 154 (see Figure 6 for the various sites of mutagenesis). The first mutations investigated were Asn157Gly and Ser159Glu. To study the effect of these mutations, we alchemically transformed hAGT into <sup>GE</sup>AGT using FEP/TI simulations. We found that this transformation has a destabilizing effect ( $\Delta G$  of 18 kcal mol<sup>-1</sup>), which is in line with the experimentally observed drop in stability in <sup>GE</sup>AGT versus hAGT. A per residue decomposition was used to indicate the specific molecular interactions that were most affected as a result of the transformation.



**Figure 12:** Two equilibrated MD snapshots from a FEP/TI simulation for the free enzyme SNAP-tag in which the residues 32 and 33 have been perturbed. The point mutation at this position represents one of several mutations that take place in the directed evolution step of <sup>M</sup>AGT to SNAP-tag. (A) In SNAP-tag, Glu159 is oriented toward the alpha helix and is within hydrogen bonding distance of Ser135 whereas in <sup>M</sup>AGT (B), Glu159 is oriented toward the N-terminal beta sheet and forms a salt bridge with Lys32. All distances were averaged during the last 20 ns of the MD trajectory.

The results of the decomposition show that the destabilization is mainly due to a loss of hydrogen-bonding interactions between the alpha helix (residues 127-136) and the loop (residues 157 and 159). While, overall, the mutation is destabilizing, the Ser159Glu mutation in <sup>GE</sup>AGT introduces a salt bridge interaction between Glu159 and Lys32 in the N-terminal beta sheet. Because of this interaction, Glu159 in our <sup>GE</sup>AGT model is oriented towards the beta sheet, as illustrated in Figure 12B. It is noteworthy that this conformation would prevent hydrogen bonding of the glutamate carboxylate with BG. However, Lys32 is mutated to Ile during the directed evolution from <sup>M</sup>AGT to SNAP-tag. FEP/TI simulations of this transformation show that this mutation induces a destabilization in SNAP-tag (by a  $\Delta G$  of 29 kcal mol<sup>-1</sup>). The per-residue decomposition of the  $\Delta G$  indicates that the mutation Lys32Ile strongly affects Glu159, due to the loss of the salt bridge interaction (see Figure 12A and Figure 13 for more details on the per-residue decomposition). Further, the mutation

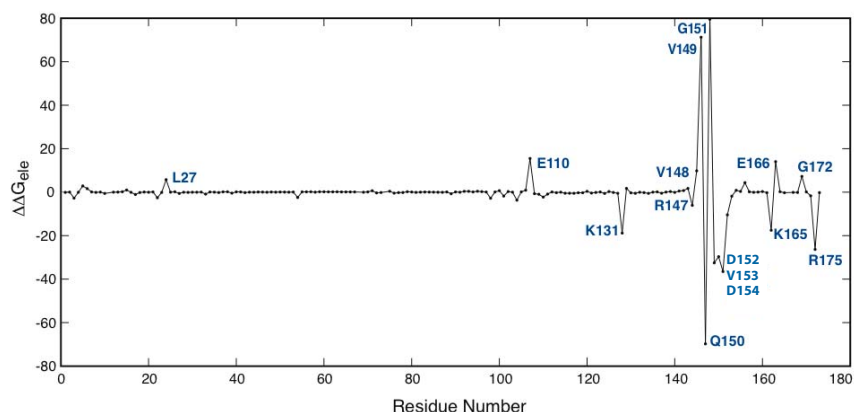


**Figure 13:** Per-residue decomposition of the alchemical transformation of MAGT to SNAP-tag focussing on position 32-33. The effect on Glu159 is strong.

Lys32Ile strongly affects the neighboring (charged) residues within the N-terminal beta sheet. Overall, the picture that emerges from our studies is that the mutation Ser159Glu in <sup>GE</sup>AGT increases the reactivity of SNAP-tag due to hydrogen bonding between Glu159 with BG. However, interactions with Lys32 keep Glu159 at least partially in an inactive conformation. The mutation Lys32Ile then increases the reactivity of SNAP-tag by disfavoring the inactive conformation of Glu159.

Analogous studies of the transformation of residues 150 to 154 demonstrate that SNAP-tag is strongly stabilized over <sup>M</sup>AGT (a  $\Delta G$  of 87 kcal mol<sup>-1</sup>), compensating the destabilizing effect of the mutation Lys32Ile, in agreement with the

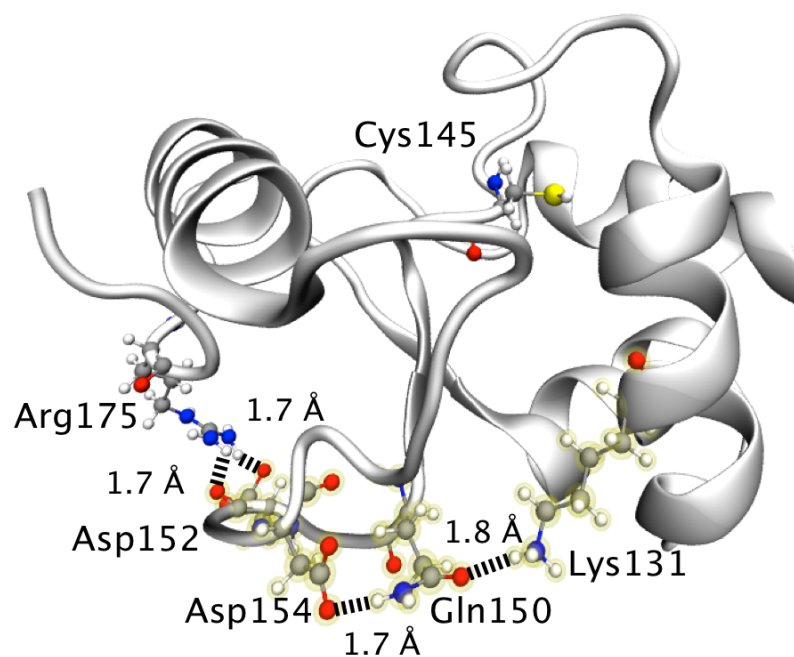
experimentally determined melting temperatures. We performed a per-residue decomposition to evaluate the residues that are the highest contributors to the stabilization of SNAP-tag (Figure 14).



**Figure 14:** Alchemical transformation of MAGT to SNAP-tag for residues 150-154. Mutations introduced in SNAP-tag contribute to the formation of a hydrogen bond network on the protein surface and clearly contribute to the gain in protein stability.

We observed that the mutations Asn150Gln, Asn152Asp and Ala154Asp contribute the most to the free energy difference because they create a highly structured hydrogen-bonding network. The network extends from the alpha helix (Lys131 hydrogen bonding to Gln150) to the tip of the loop (Gln150 hydrogen bonding to Asp154) to the C terminus (Asp152 hydrogen bonding to Arg175), as shown in Figure 15. The interactions within this hydrogen-bonding network are also observed during the MD trajectories and are preserved for 50 ns for both free and benzylated SNAP-tag structures.

We could show that the evolution of hAGT to SNAP-tag led not only to increased protein activity but also to higher stability, especially of the alkylated protein. Whereas wild-type hAGT is rapidly degraded in cells after alkyl transfer, the high stability of benzylated SNAP-tag prevents proteolytic degradation. During the last rounds of protein evolution the selection for more reactive mutants yielded mutants with increasing stability. Mutations introduced in the recognition helix (residues 127-136) and the adjacent loop-part (150-154) were major contributors to this improvement.

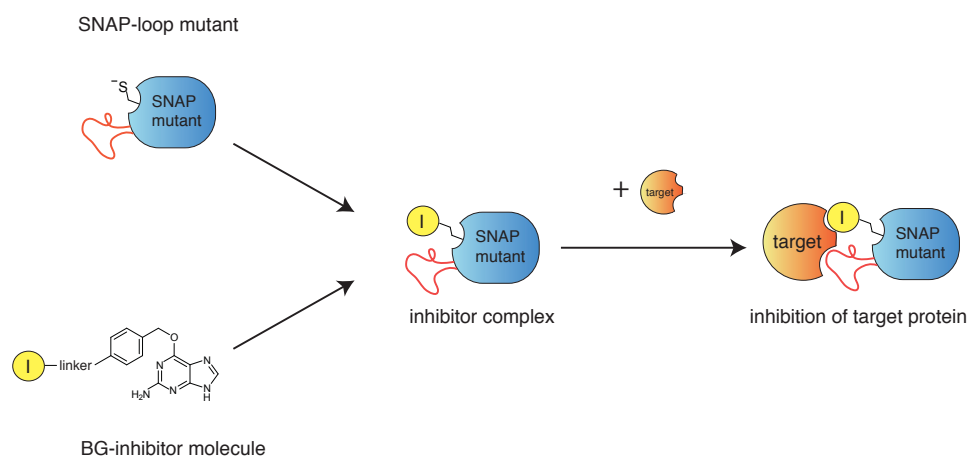


**Figure 15:** Displayed is a snap shot of the free protein structure of SNAP-tag from an equilibrated MD trajectory in which hydrogen bonding distances were averaged over the last 20 ns of the trajectory. An extended hydrogen-bonding network forms as a result of several point mutations, involving residues 131, 150, 152 and 154. The hydrogen bonding network extends from the C-terminal alpha helix (Lys131) to the loop region (Gln150, Asp152, and Asp154) and finally to the C-terminal Arg175. Such a network strongly reinforces the active site by providing stability to the alpha helix-loop domain.

### 3 Engineering of SNAP-tag-based inhibitors of protein function

#### 3.1 Introduction:

This chapter focuses on the development of a SNAP-based inhibitor of protein function. This inhibitor is a complex of two components, (i) a SNAP-mutant with binding affinity to the target protein and (ii) a small molecule inhibitor linked to BG (Figure 16). SNAP-mutant labeling with the BG-inhibitor molecule assembles the inhibitor complex, which binds to the target protein. The binding of the loop to the target protein surface leads to an increase in the effective concentration of the small molecule inhibitor at the active site of the target protein and to its inhibition.



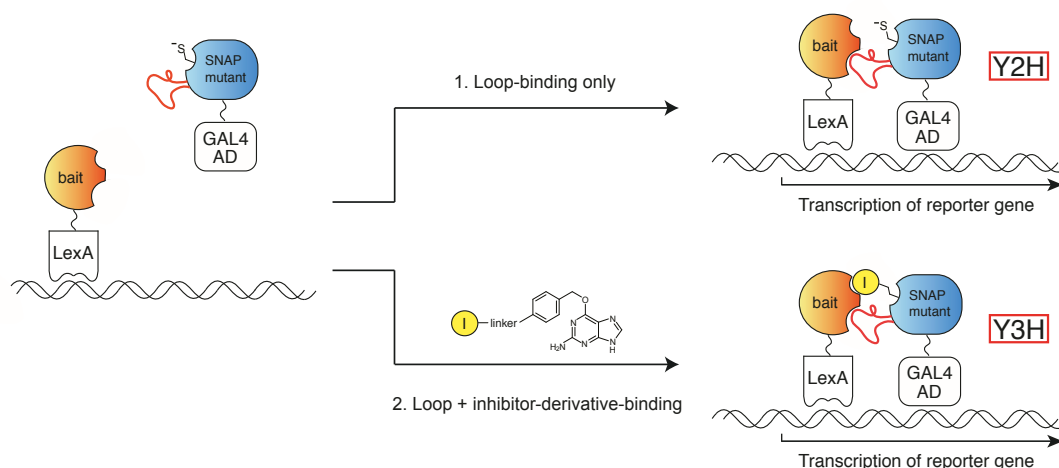
**Figure 16:** Assembly of the inhibitor complex. A loop-mutant of SNAP-tag reacts covalently with a BG-inhibitor derivative. The specific binding of the loop increases the effective concentration of the inhibitor molecule close to the active site of the target protein leading to an efficient inhibition of protein function.

For the generation of a SNAP-mutant that specifically binds to the target protein, libraries of SNAP-tag mutants were constructed. SNAP-tag was modified by the insertion of a stretch of randomized amino acids at three different positions of the protein (see also Figure 19). The created protein libraries were screened for binding



affinity towards two target proteins, *E.coli* dihydrofolate reductase (eDHFR) and human polo-like kinase 4 (Plk4).

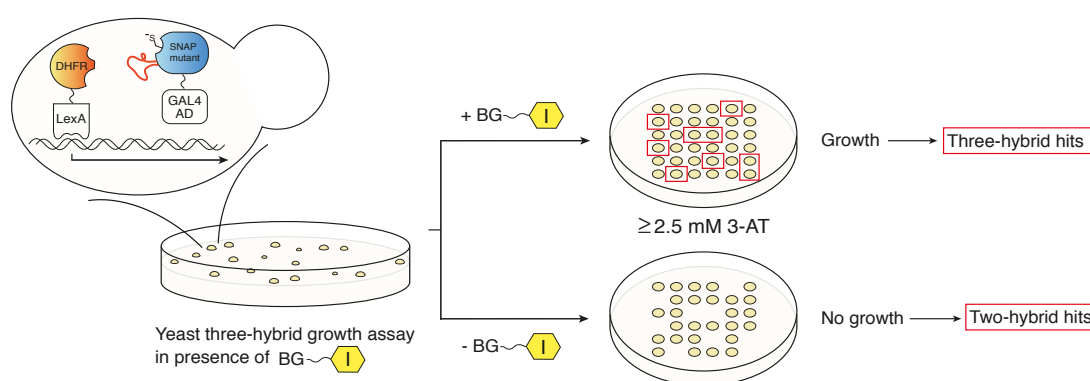
In order to identify affinity binders, a yeast two-hybrid (Y2H) and a yeast three-hybrid (Y3H) approach were used, the latter based on a recently developed method in our laboratory<sup>152</sup> (Figure 17). The loop mutant libraries were expressed as fusion proteins to the activation domain GAL4. The bait-proteins eDHFR and Plk4 were fused to the DNA-binding domain LexA. For Y2H screening the yeast cells were incubated on selection plates not containing any BG-molecule. SNAP-mutants having affinity to the bait protein trigger reporter gene transcription and induce cell growth (Figure 17, 1).



**Figure 17:** SNAP loop-library screening approach with and without BG-inhibitor. In the classical Y2H screening (1) a loop-mutant with affinity towards the bait-protein can trigger reporter gene transcription. In parallel, a Y3H screen was performed (2). Therefore, a non-specific and rather low affinity small molecule inhibitor was derivatized with a benzylguanine-moiety. Reporter gene transcription depends on the binding of both parts, the loop-part, which ensures a specific binding to the bait-protein and the small molecule inhibitor-part, whose effective concentration at the active site of the bait protein would be strongly increased.

Beside the detection of a direct bait and prey interaction, the Y3H screening was carried out including a BG-inhibitor molecule to detect small-molecule dependent interactions. Therefore, a non-specific and rather low affinity small molecule inhibitor was derivatized with a BG-moiety. In the ideal case, reporter gene transcription would be dependent on the binding of both parts, the loop-part which ensures a specific binding to the bait-protein and the small molecule inhibitor-part, whose effective concentration would be strongly increased upon loop-binding (Figure 17, 2).

Small-molecule dependency was tested by respotting of the yeast colonies on selective plates with and without BG-inhibitor molecule (Figure 18). Colonies that would grow on both plates express a mutant whose interaction does not depend on the small molecule inhibitor and would be considered as Y2H hits. If growth would depend on the BG-inhibitor molecule the cells should only grow on medium that contains the small molecule. This respotting step has been previously used for the identification of drug-targets from a genomic library where the dependency of yeast growth on the presence of the small molecule was successful for drug-target identification<sup>152</sup>.

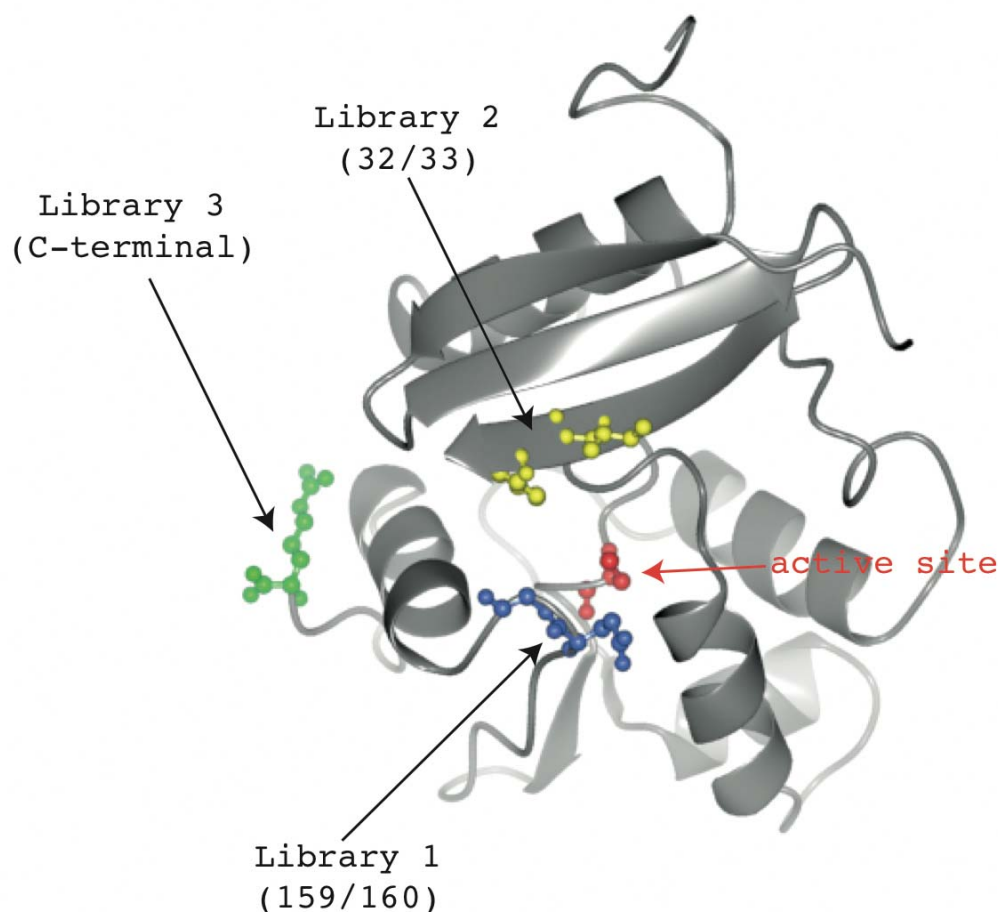


**Figure 18:** Determination of small-molecule dependency by colony respotting. Yeast colonies were resplotted on fresh agar plates, only one supplemented with 10  $\mu$ M of BG-inhibitor. In the case of a Y3H interaction, this should result in yeast growth only on the plate containing the BG-inhibitor but no growth should be observed in the absence of the molecule. Figure taken and adapted from C. Chidley et al.<sup>153</sup>

### 3.2 Construction of SNAP-tag loop libraries

This paragraph describes the construction of the SNAP-tag loop-libraries. In order to generate SNAP-tag mutants with binding affinity towards a target protein, SNAP-tag was elongated at 3 different positions (Figure 19). Besides the use of a C-terminal SNAP-tag library that was elongated by a stretch of six random amino acids at the end of the protein the introduction of loop-structures was envisioned. Earlier work of our group<sup>154</sup> had shown that <sup>M</sup>AGT could be successfully modified by loop insertion at position 32/33 and 159/160 and a high quantity of correctly folded protein was obtained. We therefore decided to conserve these positions for SNAP-tag

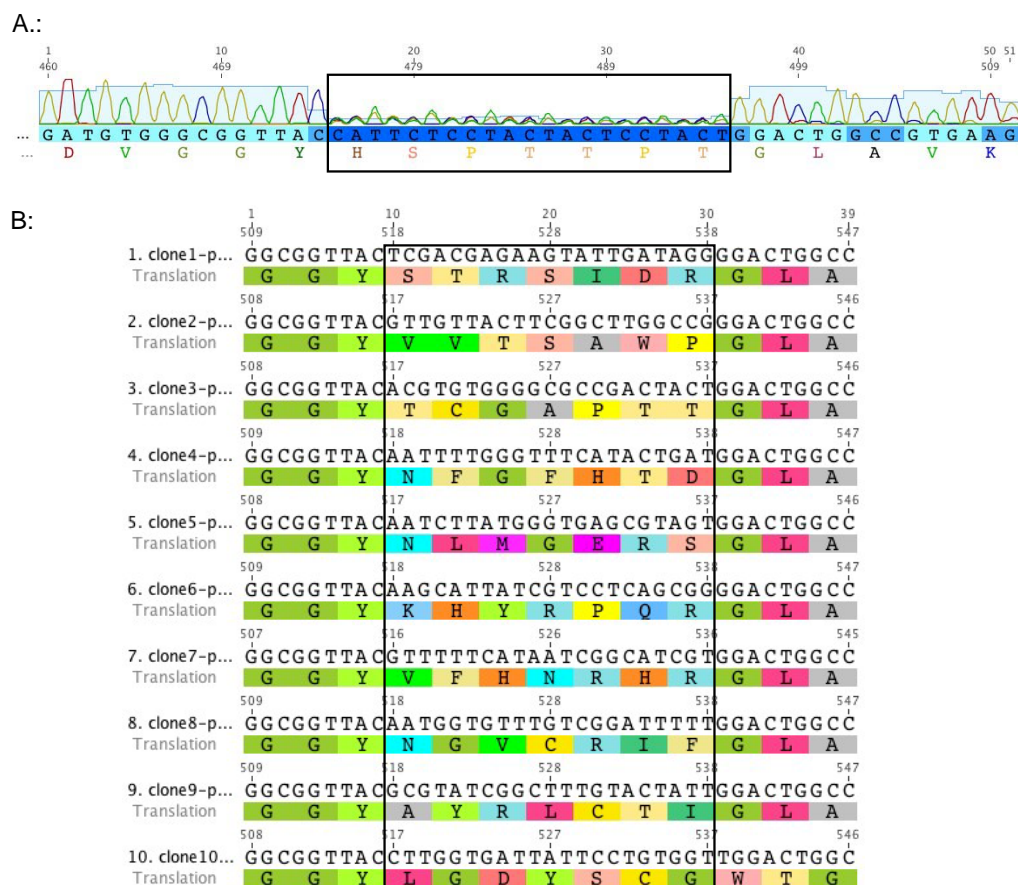
modification and inserted a stretch of seven randomized amino acids including Glu159 and Gly160 for library 1 and eight randomized amino acids between Ile32 and Phe33 for library 2. The third library carried the C-terminal elongation of six random amino acids.



**Figure 19:** Sites for loop-modification on SNAP-tag. Additional randomized amino acids were introduced at position 159/160 for library 1, at position 32/33 for library 2 and a C-terminal elongation for library 3.

The randomized loop sequences were based on NNK codon usage (N for all four bases, K for thymine and guanine) to reduce the amino acid codon number and the amount of possible stop codon integration from three to one (TAG). After library construction, ten individual mutants, as well as one aliquot of a mixed library preparation were sequenced. Figure 20 shows the sequencing results for a pool of library 1 (A) plus ten randomly chosen loop-mutants (B). SNAP-tag was correctly modified at position 159/160 and loop sequences were randomized. No stop codons

were found and only mutant 10 showed a frame shift, most probably due to an error in primer synthesis.



**Figure 20:** Diversity control of loop library 1. Seven random amino acids have been introduced replacing residue 159/160 of SNAP-tag. A: Chromatogram of the library pool highlighting the randomized region. B: Sequencing results of ten randomly chosen plasmids confirming the diversity of the library.

The test sequencing of library 2, in which eight randomized amino acids were introduced at position 32/33 showed that four out of ten mutants carried a stop codon and one sequence was isolated twice (Figure 21). Due to these findings the diversity of full-length SNAP-mutants in library 2 was expected to be inferior to the diversity of library 1.



Library sizes in yeast and coverage of <i>E.coli</i> stocks:			
bait proteins	Library 1	Library 2	Library 3
eDHFR <sup>Gly54</sup>	3.0x10 <sup>6</sup>	1.5x10 <sup>6</sup>	3.0x10 <sup>6</sup>
eDHFR <sup>Val31Gly54</sup>	1.0x10 <sup>6</sup>	3.9x10 <sup>6</sup>	1.8x10 <sup>6</sup>
Plk4-KD	4.5x10 <sup>6</sup>	4.0x10 <sup>6</sup>	5.3x10 <sup>6</sup>
Plk4 full length	1.5x10 <sup>6</sup>	3.0x10 <sup>6</sup>	3.0x10 <sup>6</sup>
% coverage of complexity ( <i>E.coli</i> )	16.7	22.7	20.0

**Table 2:** Library sizes in yeast for all bait-proteins used in the screening. 15 µg of DNA originating from the *E.coli* stock were used for large-scale yeast transformation and the transformation efficiencies determined. In the yeast screening 1x10<sup>7</sup> different mutants were plated on each screening plate, which covered about 20% of the initial library complexities in *E.coli*.

### 3.3 Optimization of screening conditions

The optimization of the screening conditions comprised the reduction of non-specific background growth to an acceptable level, the confirmation of bait-protein expression in yeast and the determination of auto-activation for each bait protein. Further, the labeling efficiency of BG-derivatized inhibitor molecules was quantified in yeast cells.

#### 3.3.1 Bait-proteins

The bacterial enzyme dihydrofolate reductase (eDHFR) and the human protein polo-like kinase 4 (Plk4) were chosen as target proteins. eDHFR was well suited for a proof-of-concept experiment of the yeast two-hybrid- and three-hybrid approach, as point mutations introduced in eDHFR were shown to modify the binding affinity towards the small molecule inhibitor methotrexate (Mtx) from sub-nanomolar (wild-type protein) to around micromolar for the weakest mutant (Val31Gly54 mutation)<sup>155,156</sup>. The moderate binding affinity to Mtx should be beneficial for the detection of loop-dependent binding.

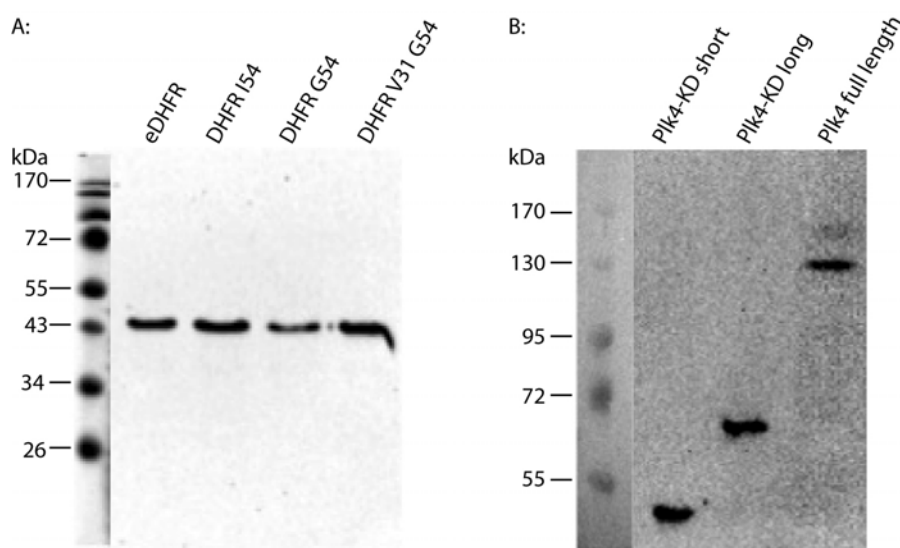
Plk4, the second bait-protein, is important for proper centriole duplication, as silencing results in disorganized mitotic spindles and apoptosis<sup>157</sup>. The kinase seemed a particularly relevant choice for our approach, as no specific small molecule inhibitors were available at that time. Besides binding specificity, SNAP-based

protein inhibition offered important advantages over existing methods. The loop-mediated binding targets the native form of Plk4. Compared to kinase depletion using siRNA, which takes approximately one to two days, SNAP-mediated inhibition depends on the protein labeling and therefore has a higher temporal resolution of several minutes to hours which would allow for the inhibition of Plk4 at distinct time points of the cell cycle.

### 3.3.2 Bait-protein expression in yeast

eDHFR and eDHFR mutants (eDHFR<sup>Ile54</sup>, eDHFR<sup>Gly54</sup> and eDHFR<sup>Val31Gly54</sup>) were expressed in yeast as C-terminal fusion proteins to the DNA-binding protein LexA.

For Plk4 two constructs of the kinase domain (residues 1-88 and 1-130 (design based on Leung *et al.*<sup>158</sup>) and the full-length protein were fused to LexA in the same way as eDHFR. The fusion proteins were expressed in yeast and analyzed by Western blot on cellular lysates using a LexA specific antibody (Figure 22). The analysis showed that all proteins were expressed as full-length constructs and could therefore be used for the screening.



**Figure 22:** Bait protein expression confirmed by Western Blot against LexA. Bands correspond to the expected size of bait proteins. A: Analysis of eDHFR and DHFR-mutants, B: Analysis of Plk4-constructs.

### 3.3.3 Bait auto-activation

The level of HIS3-reporter gene activation was measured for each bait-protein. Therefore, the individual bait-proteins fused to the DNA-binding domain LexA were

co-expressed with either GAL4 from the empty prey vector (pGAD-HA, Dualsystems) or with the fusion protein SNAP-Gal4.

No auto-activation could be detected for eDHFR and DHFR-mutants, as growth on selective medium did not yield any colonies. Even in the presence of SNAP-tag no reporter gene transcription was observed (Table 3).

Prey plasmid	Medium	Bait plasmid (Number of colonies)			
		eDHFR	DHFR <sup>Ile54</sup>	DHFR <sup>Gly54</sup>	DHFR <sup>Val31Gly54</sup>
Empty	Non selective	150	250	290	136
	selective	0	0	0	0
SNAP-fusion	Non selective	240	190	177	190
	selective	0	0	0	0

**Table 3:** Bait auto-activation of eDHFR and mutants. Incubation on non-selective media gave several hundred colonies. The same number of cells plated on selective media did not yield any single colony, confirming the absence of reporter gene activation by eDHFR and mutants.

However, all Plk4 constructs showed some auto-activating properties, especially in the presence of SNAP-tag (Table 4). The long version of Plk4 kinase domain (Plk4-KD<sup>long</sup>) was strongly activating and was excluded from the screening.



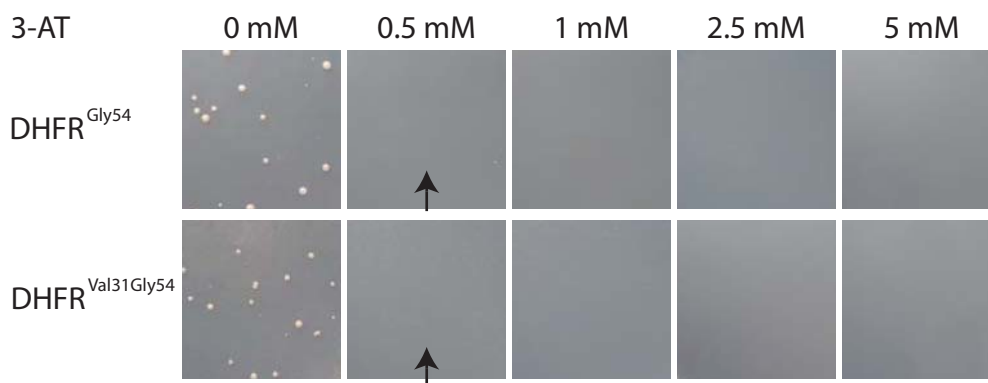
Prey plasmid	Medium	Bait plasmid (Number of colonies)			
		Plk4-KD <sup>short</sup>	Plk4-KD <sup>long</sup>	Plk4	p53 (positive ctrl.)
Empty	Non selective	130	190	80	136
	selective	1	1	6	0
SNAP-fusion	Non selective	136	108	38	
	selective	28	60	2	
LargeT-fusion	Non selective				260
	selective				95

**Table 4:** Auto-activation of transporter gene transcription of Plk4-constructs and the interacting pair p53-largeT as positive control. The co-expression with Gal4 (empty plasmid) under selective conditions gave between one and six colonies. This number increased further when SNAP-Gal4 was co-expressed with the bait-protein.

### 3.3.4 Pilot screens:

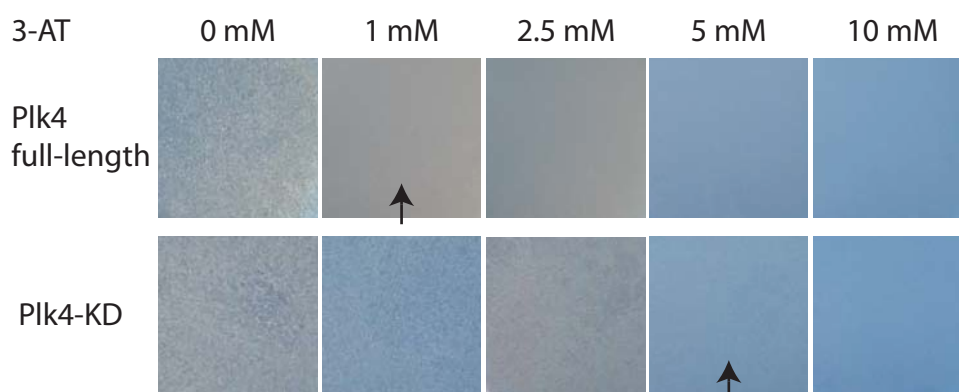
Pilot screens of all bait constructs with SNAP-tag-Gal4 were performed to determine the screening conditions for the lowest possible background growth.

Besides changes in the fusion-protein construct two other parameters can be modified to reduce background growth. One is the total amount of cells plated on solid agar, the second the concentration of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of imidazolglycerol-phosphate dehydratase, the product of the HIS3 reporter-gene. In order to determine the optimal 3-AT concentration for each bait protein, yeast cells were incubated with increasing amounts of 3-AT (from 0 to 5 mM for DHFR and 0 to 10 mM for Plk4). Yeast cells expressing DHFR<sup>Gly54</sup>LexA or DHFR<sup>Ile31Gly54</sup>LexA plus SNAP-Gal4 showed only little background growth (Figure 23). According to these results, a concentration of 0.5 mM 3-AT was sufficient to limit background growth.



**Figure 23:** Pilot screen for DHFR<sup>Gly54</sup> or DHFR<sup>Val31Gly54</sup> expressing yeast cells. A concentration of 0.5 mM 3-AT is sufficient to fully inhibit background growth (indicated by the black arrows).

For Plk4-constructs, where a stronger auto-activation had been observed previously, the amount of 3-AT had to be increased to 1 mM for Plk4 full-length and up to 5 mM 3-AT for Plk4 kinase-domain<sup>short</sup> to limit background growth to an acceptable level (Figure 24).



**Figure 24:** Pilot screen of Plk4 full-length and Plk4 kinase-domain. Both show higher background growth than DHFR and screening was performed in the presence of 3-AT. Arrows indicate 3-AT concentrations for each bait-construct respectively.

### 3.3.5 Growth dependency of eDHFH and DHFR-mutants

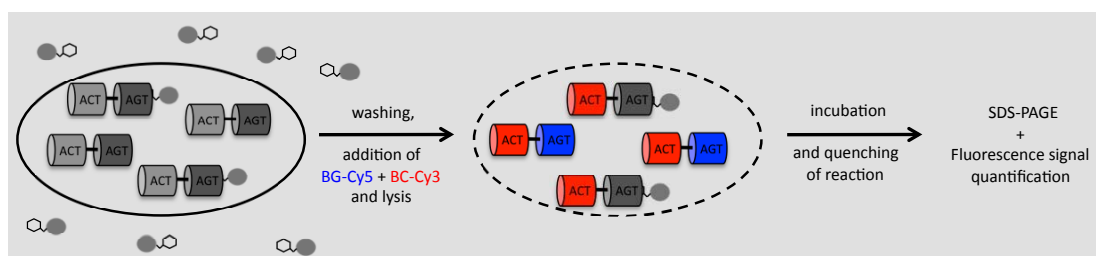
The binding of Mtx to eDHFR is a strong interaction in the low nanomolar range. The engineered mutants of eDHFR showed a decreased binding affinity for the inhibitor in



### 3.3.6 Small molecule labeling efficiency in yeast

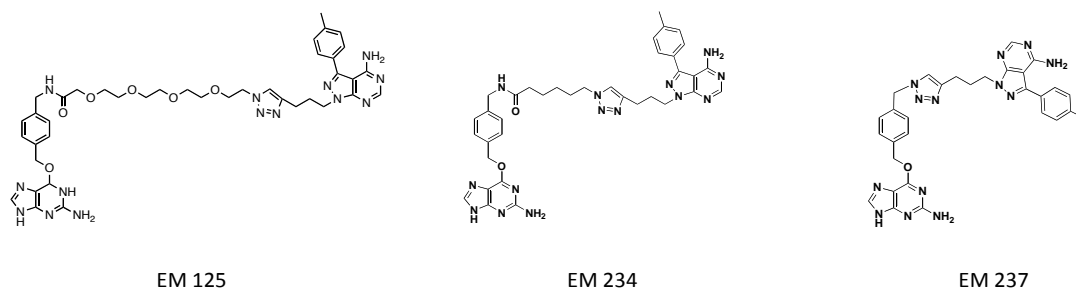
The small molecule labeling efficiency inside yeast cells is an important parameter for the Y3H screening.

The assessment of the intracellular labeling efficiency was based on the expression of a CLIP-SNAP fusion protein in yeast cells (Figure 26). CLIP-tag is an orthogonal self-labeling protein tag to SNAP-tag, which covalently reacts with benzylcytosine (BC) derivatives<sup>30</sup>. The method involves the following steps: (i) incubation of intact yeast cells expressing CLIP-SNAP with a BG-derivative of interest to (partially) label SNAP-tag either in liquid medium or on solid agar plates; (ii) cell lysis in the presence of BG-Cy5 and BC-Cy3 leading to the complete labeling of CLIP-tag and the remaining SNAP-tag in the cell lysate; (iii) SDS-PAGE and in-gel fluorescence analysis to calculate the labeling efficiency of BG-inhibitor from the ratio of Cy5/Cy3 fluorescence intensities of the CLIP-SNAP fusion protein. The calculation requires the ratio of fluorescence intensities of a control sample not incubated with BG-inhibitor to set the 0% labeling reference. CLIP-tag serves as an internal reference to account for variations in fusion protein quantities over samples.



**Figure 26:** Overview of the assay to determine intracellular SNAP-tag labeling efficiency by SNAP-CLIP fusion protein labeling. ACT-AGT dimers are expressed in yeast cells and incubated with a BG-substrate, whose labeling efficiency should be assessed. At the end of incubation, cells are washed and a mixture of BG-Cy5, BG-Cy3 is added to the cells. After lysis and further incubation to reach complete labelling, the reaction is quenched by the addition of a large excess of free BG. Sample buffer is added and the labeling efficiency is determined via SDS-PAGE analysis and quantification of fluorescence signal intensity.

The permeability of BG-Mtx was determined previously<sup>153</sup>. Here, the labeling efficiency of PP1 was assessed. PP1 is a kinase inhibitor molecule, which was linked to BG via three different linker moieties (Figure 27). The molecules were provided by Dr. Eveline Müller<sup>153</sup>. The labeling efficiencies in liquid and on solid medium measured with this method should provide a good estimate of the labeling efficiencies in Y3H screenings.



**Figure 27:** BG-PP1 derivatives with variations in linker length. These molecules were tested for their labeling efficiencies inside yeast cells.

The results clearly showed that the labeling efficiency strongly depends on the linker that was used to attach BG to Mtx (Table 5). An alkyl chain linker was preferred over polyethyleneglycol (PEG) units (compare EM234 versus EM125). At a concentration of 50  $\mu\text{M}$  BG-PP1 labeling efficiencies were slightly higher than for a concentration of 10  $\mu\text{M}$ . On solid medium, labeling efficiencies were superior to those in liquid culture, which can be explained by the prolonged incubation times necessary for yeast growth on solid media. According to these results it had to be considered that only a part of the mutants would be present in their labeled state during Y3H screening.

compound	10 $\mu\text{M}$		50 $\mu\text{M}$	
	liquid	plate	liquid	plate
EM 125	40 %	45 %	60 %	---
EM 234	75 %	83 %	95 %	---
EM 237	93 %	85 %	99 %	---
BG	---	97 %	99 %	99 %

**Table 5:** Summary of intracellular labeling efficiencies of the three tested molecules, EM125, EM123 and EM127 at two different concentrations (10 and 50  $\mu\text{M}$ ) in liquid and on solid media.

### 3.3.7 Test screening

According to the results obtained from the optimization procedures, a BG-Mtx concentration of 1  $\mu\text{M}$  together with eDHFR<sup>Gly54</sup> and eDHFR<sup>Ile31Gly54</sup> was chosen as starting conditions for the screen. Fixing the BG-Mtx concentration close to the dissociation constants should increase the sensitivity for SNAP-tag loop binding, as the eDHFR-Mtx interaction should not be sufficiently strong to trigger yeast growth.

Table 6 summarizes the results from each screening plate. Strong background growth was observed and several plates could not be evaluated. Sequencing of randomly chosen colonies showed that a high number of mutants carried a stop-codon and interactions were not reproducible in fresh yeast cells (data not shown).

	library	Number of colonies on screening plate	
		BG-Mtx (1 $\mu\text{M}$ )	control (DMSO only)
DHFR <sup>Gly54</sup>	1	178	134
	2	30	35
	C-terminal	strong background	strong background
DHFR <sup>Val31Gly54</sup>	1	strong background	strong background
	2	23	28
	C-terminal	strong background	strong background

**Table 6:** Screening output of test screening against eDHFR at a 3-AT concentration of 0.5 mM. For the C-terminal library and DHFR<sup>Val31Gly54</sup> with library 1 a strong background growth was observed. For all other screening plates between 23 and 178 colonies were detected. Further analysis identified them as false-positive interactions.

In order to improve the screening output two modifications were tested. First, the concentration of BG-Mtx on the screening plates was increased to 10  $\mu\text{M}$  to improve the intracellular labeling of the mutants. Second, selection stringency was increased using a higher 3-AT concentration of 2.5 mM.

The augmentation of BG-Mtx concentration did not improve the screening output. Around 20 colonies were detected on the screening plates but further analysis identified all of them as false-positive interactions (data not shown). The increase of 3-AT concentration improved the screening output, as mutants with bait-specific interaction were identified. However, no dependence on BG-Mtx was detected in colony respotting.

### 3.4 Screening against eDHFR

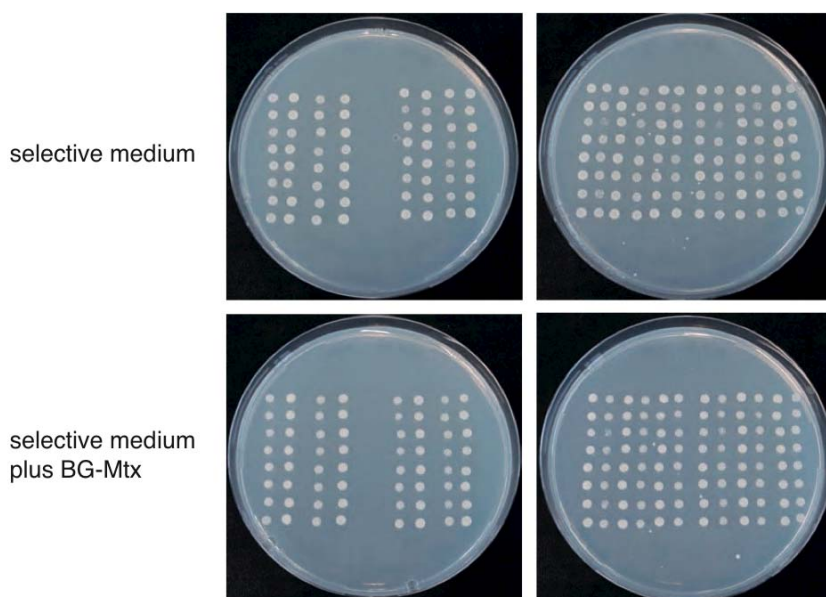
According to the results obtained in the test screening of eDHFR, higher concentrations of 3-AT were needed for sufficient background suppression. Therefore, the screening of all SNAP-libraries was done using the optimized conditions.

As a consequence, the overall number of colonies decreased to not more than 30 colonies per plate (Table 7). The highest number of colonies was detected for the C-terminal library, for loop libraries 1 and 2 the colony number further decreased.

	Number of colonies on screening plate	
	library	1 $\mu$ M BG-Mtx
DHFR <sup>Gly54</sup>	1	8
	2	0
	C-terminal	32
DHFR <sup>Val31Gly54</sup>	1	2
	2	7
	C-terminal	18

**Table 7:** Screening output under improved screening conditions. The stringency of screening was increased by a 3-AT concentration of 2.5 mM. As a consequence, the number of colonies drastically decreased compared to the first screening attempt. At the same time the quality of the detected colonies improved strongly as interactions were reproducible in fresh yeast cells.

The colonies were tested for BG-Mtx dependency and for their strength of reporter gene activation. No dependency on BG-Mtx could be observed as all colonies grew on selective medium not containing any BG-Mtx (Figure 28).



**Figure 28:** Respotting of colonies identified in eDHFR screening on selective medium with and without BG-Mtx. All identified hits were directly interacting with the bait-protein and no molecule-dependent interactions could be detected.

For the assessment of reporter gene activation cells were spotted on selective medium containing increasing amounts of 3-AT. The number of colonies decreased with increasing 3-AT concentration and at 2.5 mM 3-AT only eleven out of 67 hits showed good colony growth. Retransformation of the eleven mutants in fresh yeast cells resulted in a reproducible colony growth (Table 8). Sequencing analysis identified one loop mutant originating from library 1 and five mutants of the C-terminal library.

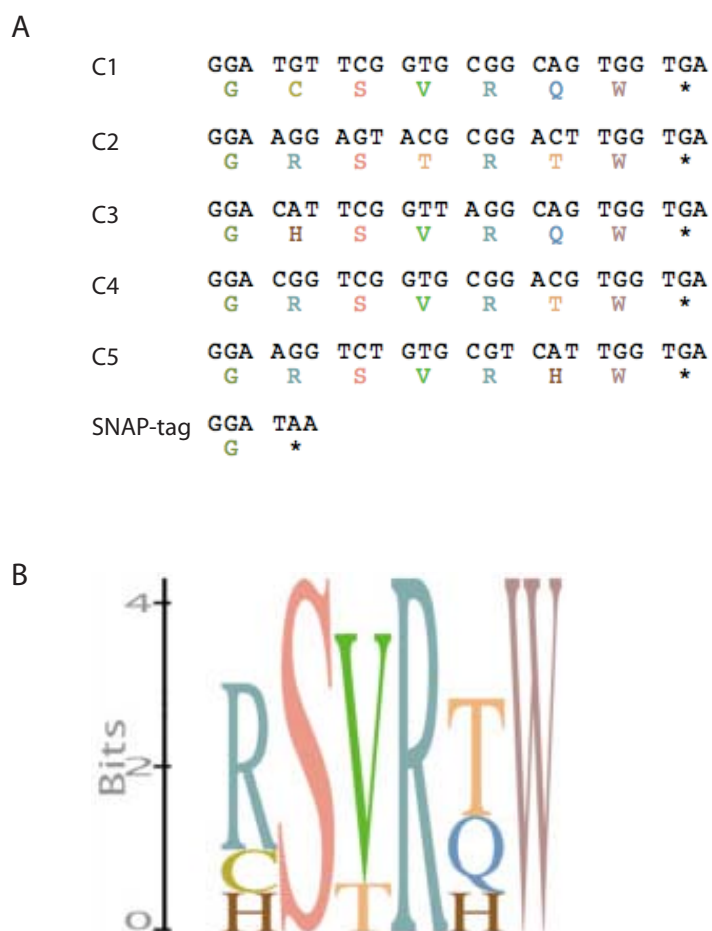
library:	sequence:	#	growth at 2.5 mM 3-AT:
159/160:	RIKARTV	2	good
C-term.:	CSVRQW	1	good
	RSTRTW	1	good
	HSVRQW	1	good
	RSVRTW	4	good
	RSVRHW	2	good

**Table 8:** Increase of 3-AT concentration improved output of library screening against eDHFR. One loop-mutant and five C-terminal mutants were isolated that all showed good growth at 2.5 mM 3-AT. The loop mutant was identified twice from the screening, the C-terminal mutants were identified between one and four-times.

The sequence analysis of the C-terminal mutants further revealed that several amino acids in the C-terminal peptide-tail were conserved (Figure 29). All mutants carried a



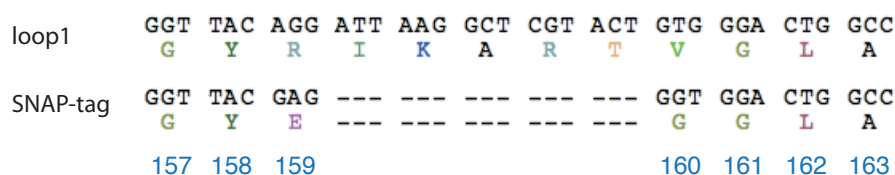
serine residue at position 2, an arginine at position 4 and a tryptophan at position 6. On the third position a valine was preferred.



**Figure 29: A:** Sequence alignment of SNAP-tag with C-terminal mutants C1-C5. Colonies isolated showed a conserved motive. Most of the amino acids use different codons, B. “Sequence logo” for isolated hits C1 to C5.

It is generally a good indicator for the screening process if the isolated sequences contain the same amino acids at conserved positions of a peptide. The fact that these amino acids were coded by different DNA-triplets only strengthened this point.

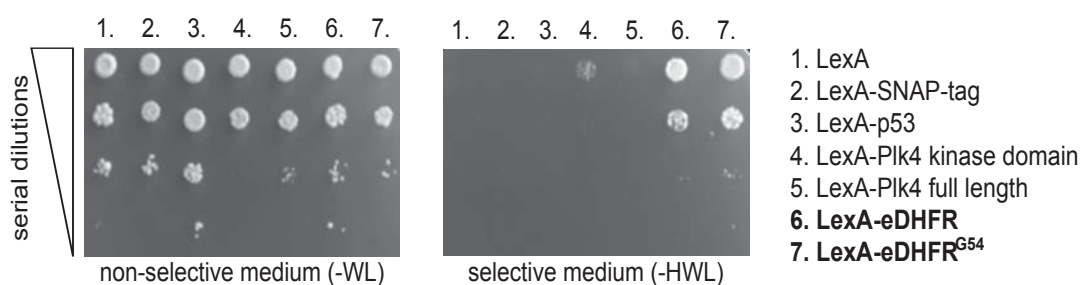
Figure 30 illustrates the sequence alignment of the isolated loop-mutant named Loop1 with SNAP-tag for the residues 157 to 163.



**Figure 30:** Sequence alignment of SNAP-tag and isolated mutant Loop1 that was isolated from the screening at higher 3-AT concentrations.

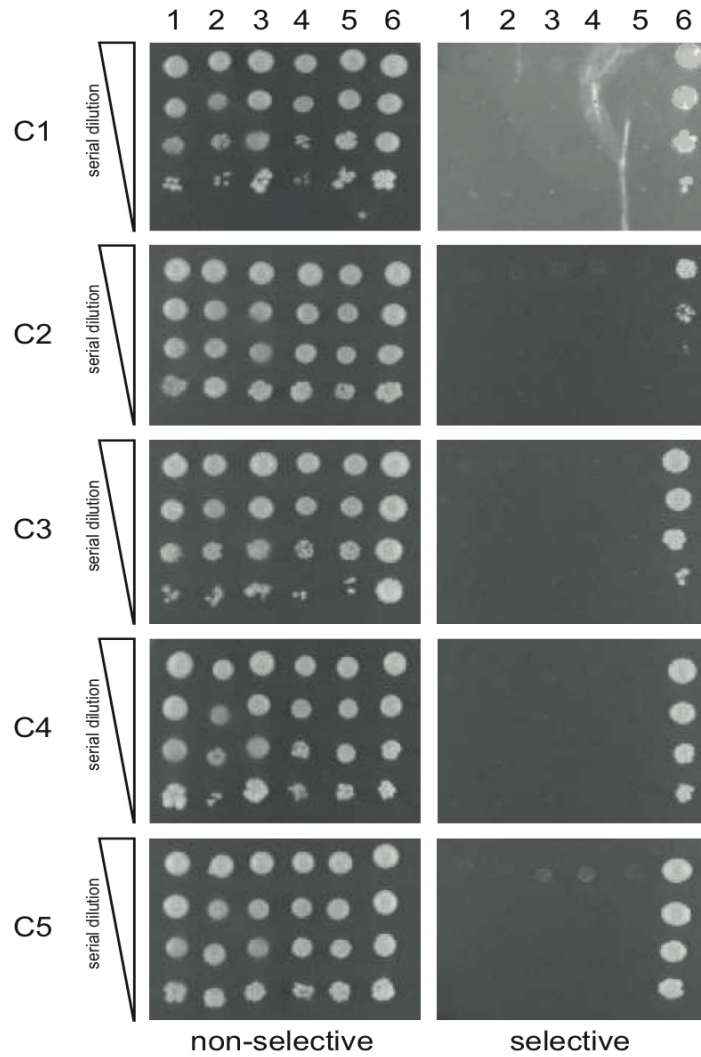
### 3.4.1 Bait specificity of eDHFR hits in yeast

The isolated hits were tested for their specificity by including several negative controls, which did not lead to colony growth (Figure 31). In the first experiment only cells expressing Loop1 together with either eDHFR or eDHFR<sup>G54</sup> were able to form colonies on selective medium. This result confirmed the specific bait-prey interaction in yeast between eDHFR (-mutants) and Loop1.



**Figure 31:** Bait-dependency test for Loop1. Serial dilutions of yeast cells (top to bottom) expressing Loop1 and different bait proteins (1-7) were spotted on non-selective (left) and selective medium (right). eDHFR-dependent growth of Loop1 was confirmed. Co-expression of LexA, the eDHFR-fusion protein used in the screening, did not trigger yeast growth.

Analysis of the C-terminal SNAP-mutants showed consistent results (Figure 32). Colonies expressing the SNAP-mutants C1 to C5 were co-transformed with the same bait-proteins than Loop1. Only co-expression of eDHFR and C1 to C5 resulted in yeast growth on selective medium.



**Figure 32:** Bait-specificity study for C-terminal mutants C1-C5. Bait-proteins from left to right: 1. LexA, 2. LexA-SNAP-tag, 3. LexA-p53, 4. LexA-Plk4 kinase domain, 5. Plk4 full length, 6. LexA-eDHFR. Only co-expression of C-terminal mutants and LexA-eDHFR triggered yeast growth.

### 3.5 Screening against Plk4

Polo-like kinase 4 (Plk4) is one of the four members of the Polo-like family and plays an essential role in centriole duplication<sup>159,160</sup>. It is essential for embryonic development<sup>161</sup>. Overexpression is leading to multiple centriole formation in the cell. Compared to the other Polo-like kinases Plk4 is the structurally most different member. It carries only one Polo-box domain instead of two and has an additional crypto Polo-box domain. Both of these domains could be shown to be important for anchoring Plk4 to the centrosome. Furthermore, the homodimerization of the Polo-box domain regulates Plk4 kinase activity<sup>162</sup>.

Plk4 stability is governed by three PEST sequences, one within the amino terminus and two within the carboxy terminus of Plk4. These regions are rich in proline (P), glutamate (E), serine (S) and threonine (T) residues and are associated with short intracellular half-lives of proteins<sup>163,164</sup>. Autophosphorylation plays an important role in regulating the stability of the kinase as trans phosphorylation of the Plk4 dimer complex targets the kinase for degradation<sup>165</sup>.

Plk4 phosphorylates in a context-dependent manner where the residues surrounding the phosphorylation site play an important role for its ability of substrate phosphorylation<sup>158</sup>. Studies on the phosphorylation motive could identify some common elements important for substrate recognition<sup>158,166,167</sup>. There is evidence that Plk4 might also play a role in other processes than centriole duplication. As many questions remain to be answered more work on this kinase is needed. The identification of SNAP-loop mutants binding to Plk4 might therefore allow new insights on substrate binding interactions and the identification of possible phosphorylation motives.

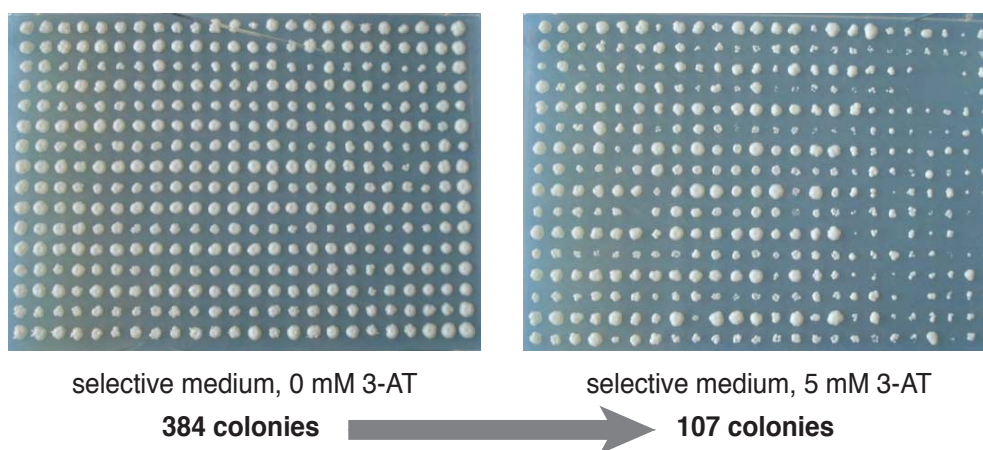
The screening of the loop-libraries against Plk4 was performed in a similar setup than the eDHFR screening. The same concentration of 3-AT was used and the screening was performed in three different settings: First, without the addition of a small molecule (Y2H-setting), second with 10  $\mu$ M of the small molecule EM125 and third with the addition of 5  $\mu$ M of the small molecule EM234 (for structures see Figure 27).

In total 384 colonies could be isolated from the screening and in general, more colonies were detected on selection plates containing a small molecule than on plates without molecule (Table 9).

Plk4 length	full	library	Number of colonies		
			no molecule	10 $\mu$ M BG-EM125	5 $\mu$ M BG-EM234
		1	40	57	50
		2	8	96	96
		3	31	39	46

**Table 9:** Colonies obtained from Plk4 screening at 2.5 mM 3-AT concentration. More colonies were obtained on plates that were supplemented with BG-inhibitor than on plates not containing any molecule.

However, the respotting on plates with and without BG-inhibitor molecules did not reveal any colonies, whose interaction would depend on the presence of the small molecule. All colonies grew on selective medium and only the titration with 3-AT reduced the colony number (Figure 33). Out of 384 isolated colonies 107 grew at a 3-AT concentration of 5 mM and were further considered as potential hits from which plasmids were isolated.

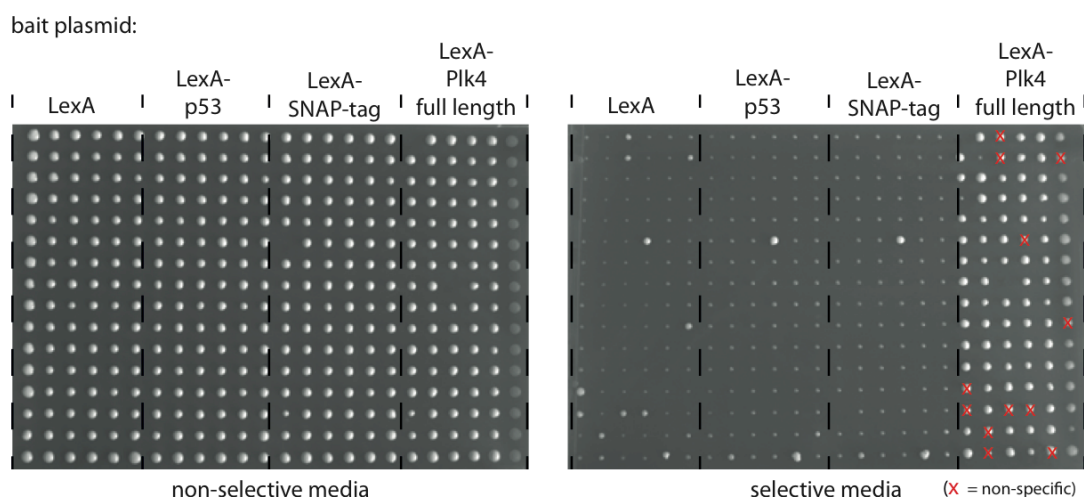


**Figure 33:** 3-AT titration reduced colony number. Only colonies that grew at 5 mM 3-AT were considered as potential hits and further analyzed.

### 3.5.1 Bait-specificity of Plk4 hits in yeast

The isolated hits were tested for their bait specificity by including three negative controls (Figure 34). As several hundred transformations had to be carried out a yeast-transformation robot was used (kindly provided by Prof. Bart Deplancke at EPF Lausanne). The output of the bait-specificity analysis showed that the majority of colonies were dependent on Plk4. A few colonies were positive for the protein LexA and only four yeast colonies seemed to be generally activated as they grew under all tested conditions. This behavior probably arose through yeast mutations, which circumvented the selection via HIS3 reporter gene activation. Only colonies that grew in a bait-specific manner were considered for further analysis and sequencing.

Sequencing identified 15 different SNAP-mutants, ten out of library 1 (loop position 159/160) and five out of library 2 (loop position 32/33) (Table 10).



**Figure 34:** Bait-specificity test of mutants identified in the screening against Plk4. Different bait-plasmids were co-transformed, only one carrying the Plk4 gene. Growth on non-selective media (left picture) showed that yeast manipulation did not affect its general viability. Incubation under selective conditions (right picture) identified several false-positive interactions that were not specific for the bait-protein Plk4, but the majority of identified colonies grew in a bait-dependent manner.

library:	ID:	sequence:	#
32/33:	L1	FLAIIRIT	1
	L2	AWVVKRY	1
	L3	RWGDCFVY	1
	L4	CSVVAWI	1
	L5	DVRVLYWL	1
159/160:	L6	LNLIYVL	2
	L7	GFGCGRVTFLIILSW	41
	L8	VRGIWMWAVTILGPFST	11
	L9	ILLCRSV	1
	L10	LRMARQI	4
	L11	LRFTHKY	1
	L12	VWAVTIFVRSWL	1
	L13	STIARQM	1
	L14	GFGCGRVTLLLIIRW	2
	L15	RLLFSSLSFL	1

**Table 10:** Output of Plk4 screening with 2.5 mM 3-AT on screening plates.

In general, a higher number of sequences were identified for loop library 1. Mutants L7, L8, L12 and L14 had the particularity to carry especially long randomized sequences. This had not been observed during the DHFR screening despite the fact that the same library stocks had been used.

Several sequences were sufficiently similar to be considered as one common sequence motive (Figure 35). In some cases inversions of amino acids were found, but overall L2, L4, L8 and L12 were considered as one motive (Trp-Ala-Val-(Thr-Ile)). L10, L13 form a second core motive (Met-Ala-Arg-Gln-Ile) and L7, L14 build the third motive block.

clone:		amount:	library:
L12		1x	1
L8	G I W M W A V T I F V R S L W	11x	1
L4	I W A V V S C	1x	2
L2	A W V V K R Y	1x	2
L10	L R M A R Q I	4x	1
L13	M Q R A I T S	1x	1
L14	G F G C G R V T L L L I R R W	2x	1
L7	G F G C G R V T F L I I L S W	41x	1

**Figure 35:** Conserved motives identified in Plk4 screening. Three major sequence blocks could be identified as putative binding sequences. Some sequences (L7 and L8) were isolated in high quantity, whereas others were isolated only once. The first motive (L2, L4, L8 and L12) was found in both libraries 1 and 2.

Interestingly, the closer examination of mutant L7 and L14 showed a very high sequence similarity with a part of the Plk4 kinase domain itself (Figure 36). Five or seven amino acids out of 15 aligned with the residues of Plk4.

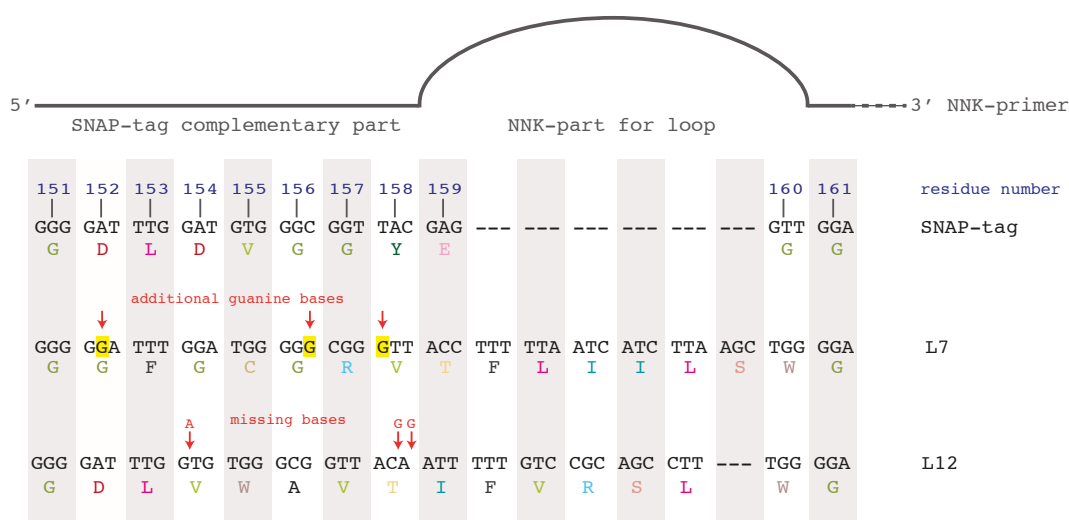
Data published by Sillibourne *et al.* on the autophosphorylation of Plk4 reported the phosphorylation of the serine residue upfront to this sequence<sup>167</sup> (Figure 36, marked in red). Additionally, the authors reported the preference for a peptide length of 12-13 amino acids as phosphorylation substrate, which might explain why mutants with longer peptide loops had been identified in our screening. It is possible that the peptides of L7 and L14 were recognized by Plk4 and potentially serve as phosphorylation substrate, as threonine and serine residues were present. However this hypothesis needs further experimental confirmation.

Plk4	S	L	G	C	M	F	Y	T	L	L	I	G	R	P	P
L7	G	F	G	C	G	R	V	T	F	L	I	I	L	S	W
L14	G	F	G	C	G	R	V	T	L	L	L	I	R	R	W

**Figure 36:** Sequence alignment of Plk4 kinase domain (residues 201-215) and loop peptide sequences of L7 and L14. The serine highlighted in red is known to be autophosphorylated. Loop peptides contain either threonine or additional serine (*italic*), which might be phosphorylated by Plk4.



The especially long randomized sequences that were isolated in the Plk4 screening arose through errors in loop primer synthesis. Two examples are given in Figure 37. In the case of mutant L7 the presence of three additional guanine-bases resulted in the introduction of additional mutations from amino acid 152 on. For mutant L12 three bases were missing. The consequence of this was that additional mutations were introduced from position 154 on. However, after amino acid 160 all sequences aligned to the original SNAP-tag sequence coding for the full-length protein.



**Figure 37:** Errors in primer synthesis were the reason for the existence of SNAP-mutants with more than eight mutations. Mutant L7 is lacking three guanine bases upfront to the NNK-part of the primer and amino acids were mutated from residue 152 on. Mutant L12 has three missing bases and mutations start at amino acid 154. From residue 161 on all mutants are correctly aligning to the SNAP-tag sequence.

In SNAP-tag residues 152 and 154 are implicated in a hydrogen-bond network that is stabilizing the protein. The consequences of the introduced mutations are difficult to predict but they may have an influence on protein reactivity and structure. However, additional experiments on protein labeling and stability are needed to draw further conclusions.

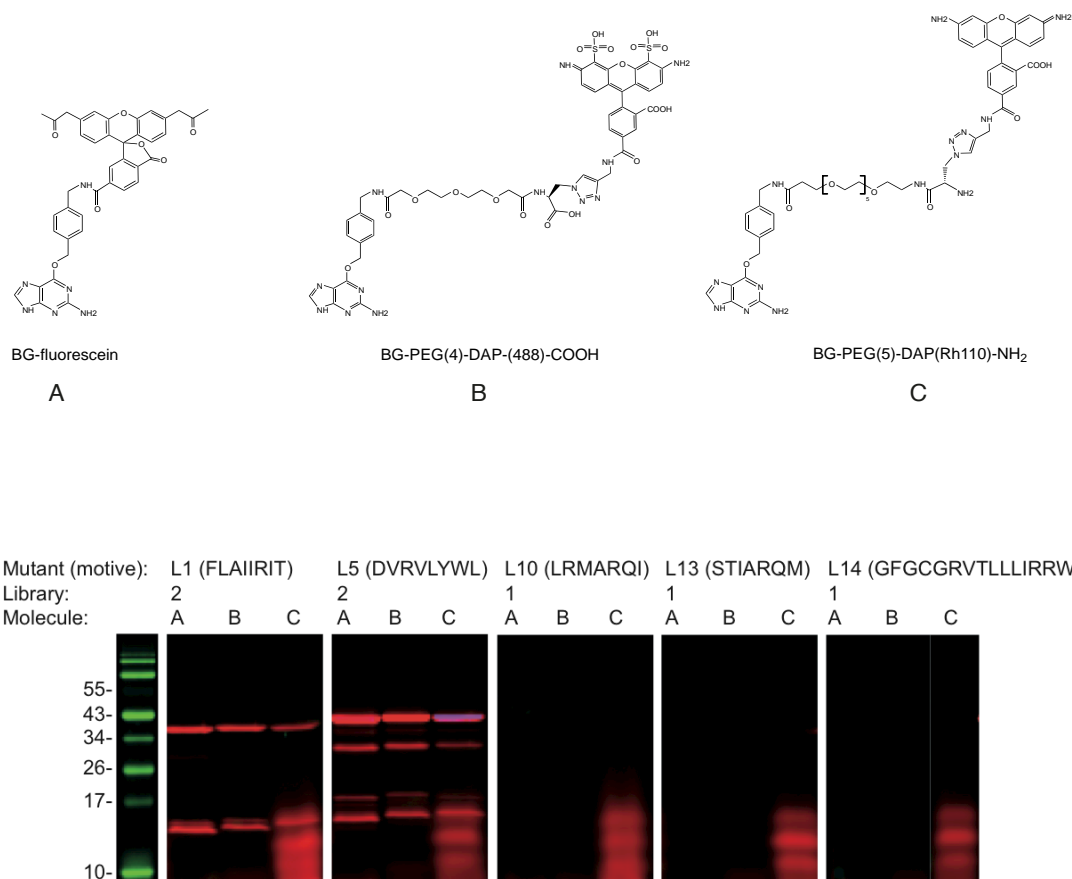
This primer “defect” represented a minor population of the library pool, as it was neither observed in the test sequencing during library construction, nor in the eDHFR-screening process. Nevertheless, these mutants were isolated frequently in the Plk4 screening and show certain sequence conservation in the randomized part.

### 3.6 *In vitro* analysis of confirmed hits

The majority of isolated mutants showed a bait-dependent growth in yeast. The next step consisted in the confirmation of protein-protein interaction *in vitro* and in the further characterization of the mutants regarding protein reactivity, stability and inhibitory activity.

#### 3.6.1 Labeling of SNAP-mutants with BG-substrates

Loop-mutants identified during the Plk4-screening were expressed in *E.coli* cells as GST-fusion proteins and the lysate was incubated with three BG-substrates that differed in size, fluorophore and linker architecture (Figure 38).



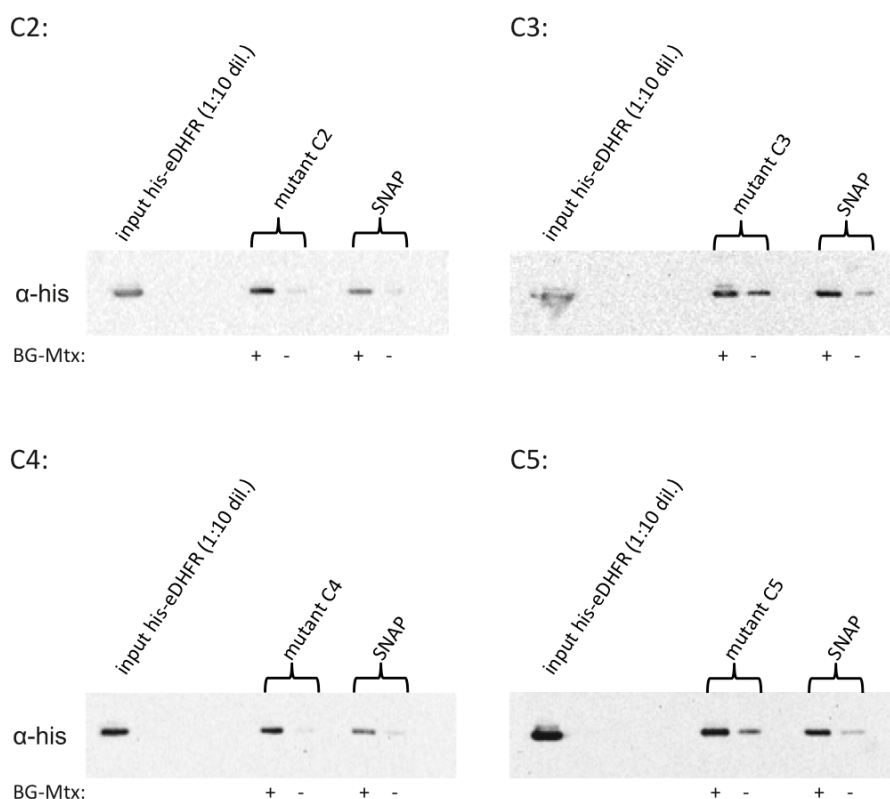
**Figure 38:** BG-substrate labeling of SNAP-loop mutants isolated during the Plk4 screening. The mutants were incubated with different BG-substrates (A, B or C). Labeling was observed for mutants from library 2 only (L2 and L6). This indicates that position 32/33 is generally better suited for loop-insertion and yields reactive mutants, whereas modifications at position 159/160 (mutants L10, L13 and L14) have a strong impact on the reactivity.

Protein labeling was assessed by fluorescence-in-gel analysis. From kinetic studies on SNAP-tag we knew that BG-fluorescein (molecule A) has good kinetics and is showing efficient labeling. Modification of BG-molecules with PEG-units was shown to reduce labeling efficiency for SNAP-tag and molecules B and C were chosen to assess this influence on the loop-mutants.

Our analysis showed that SNAP-tag modification in library 1 had a negative effect on protein reactivity. None of the mutants having a loop-insertion at position 159/160 were reacting with any BG-molecule (L10, L13 and L14). However, for library 2 in which the loop was placed between residues 32/33 (L1 and L5), protein labeling could be observed for all tested molecules, even though band intensities were weaker for compound B and C at equal protein concentrations.

### **3.6.2 GST pull-down with C-terminal SNAP-tag mutants**

The interaction with eDHFR was analysed *in vitro* by pull-down experiments. eDHFR was detected by Western blotting using a his-tag antibody (Figure 39). The experiment confirmed the interaction of the C-terminal mutants with eDHFR only in the presence of BG-Mtx. Band intensities of the pull-down samples without the BG-derivative were not significantly stronger than the background. This indicated that the potential interactions were not strong enough to be detected by a pull-down.



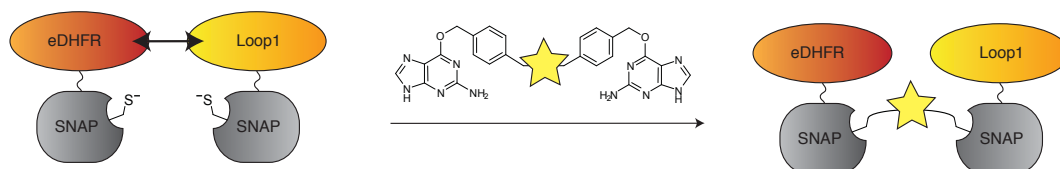
**Figure 39:** Pull-down experiment using his-tagged eDHFR with C-terminal mutants and SNAP-tag before and after labeling with BG-methotrexate. Band intensities of the input of eDHFR correspond to a 1:10 dilution.

### 3.6.3 Further characterization of mutant Loop1

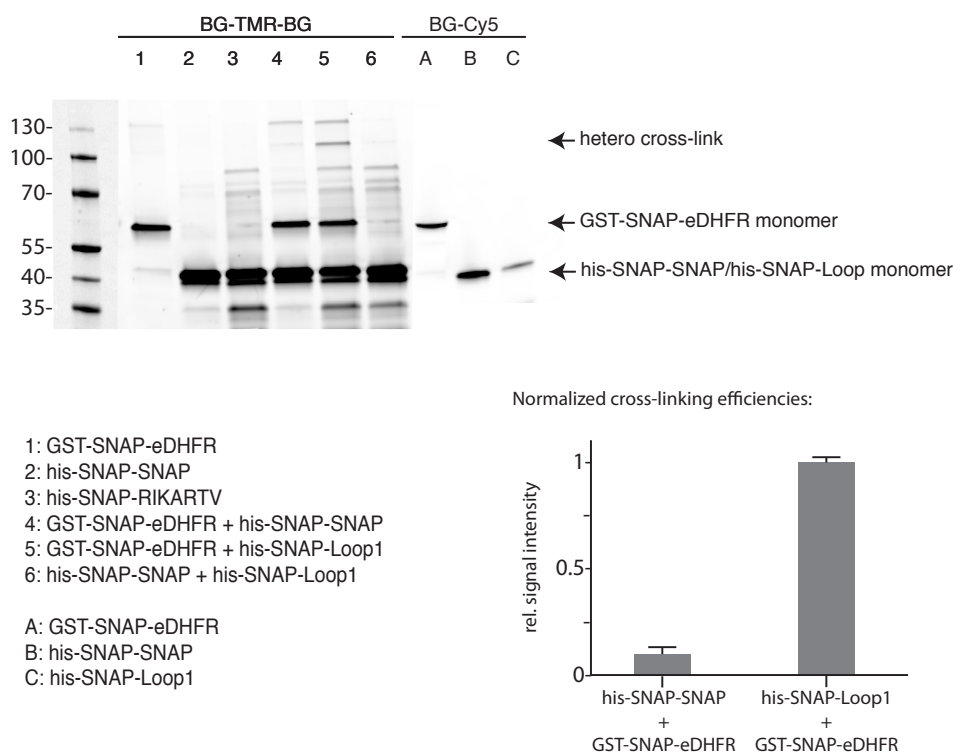
The mutant Loop1 was the only loop mutant isolated during the screening against eDHFR. As the bait-dependency test in yeast confirmed the interaction with eDHFR, it was analyzed by additional *in vitro* experiments. We decided to use a cross-linking approach to determine the interaction between Loop1 and eDHFR. Additionally, Loop1 was analyzed for BG-substrate reactivity, protein stability and inhibitor activity using an eDHFR-activity assay.

**Cross-linking of eDHFR and Loop1.** Protein binding of Loop1 to eDHFR was determined by SNAP-based cross-linking<sup>30</sup> (Figure 40). The working principle is the physical cross-linking of two interacting proteins mediated by SNAP-tag labeling using a BG-cross-linker. In the case of protein interaction the formation of a heterodimer can be detected on a SDS-PAGE.

For the protein pair eDHFR-Loop1, GST-tagged eDHFR and his-tagged Loop1 were expressed as fusion proteins to SNAP-tag. After incubation with BG-cross-linker the samples were analyzed for their in-gel-fluorescence intensity on a SDS-PAGE (Figure 41).



**Figure 40:** Working principle of SNAP-based cross-linking using a BG-cross-linker and two SNAP-tag fusion proteins. If eDHFR interacts with Loop1 the formation of the heterodimer can be resolved on a SDS-PAGE. As control a SNAP-SNAP fusion protein was included in the experiment.

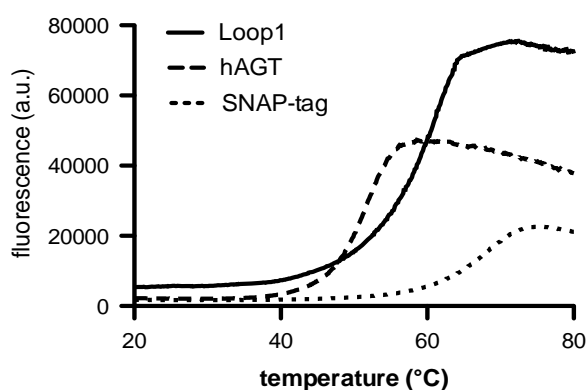


**Figure 41:** Loop1-eDHFR cross-linking experiment. Proteins were incubated with BG-TMR-TMR-BG (LR223) cross-linker and analyzed by SDS-PAGE. A band of the corresponding size of the hetero-cross-linking product of eDHFR and Loop1 was detected. Quantification showed that signal intensity was significantly higher than the control (cross-link of SNAP-eDHFR) confirming the interaction between Loop1 and eDHFR *in vitro*.

A cross-linking band for Loop1 and eDHFR fusion proteins could be observed and the presence of both proteins (eDHFR and Loop1) was confirmed by Western blot

analysis (data not shown). Quantification of band intensities and comparison with the control experiment (unspecific cross-link between SNAP-tag and eDHFR) confirmed the interaction between Loop1 and eDHFR, as signal intensities were significantly higher for Loop1 than for SNAP-tag.

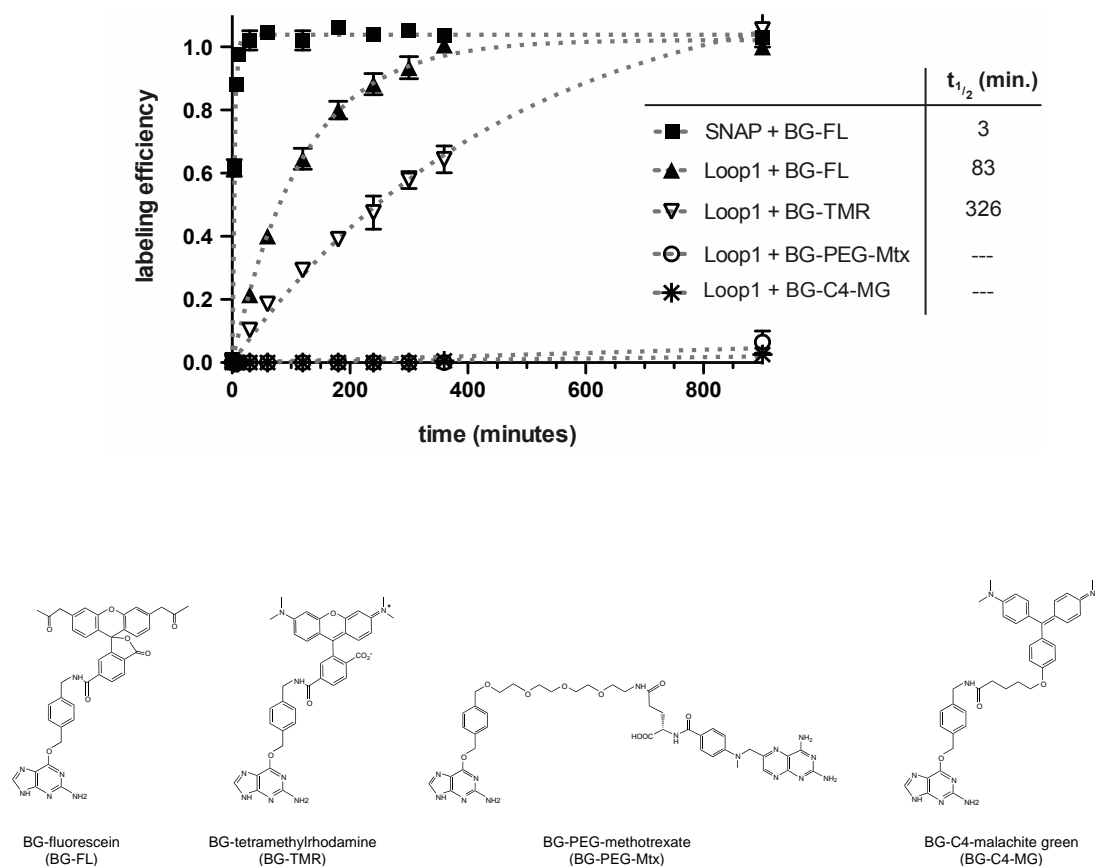
**Protein stability determination of Loop1.** The insertion of additional amino acids in a given protein structure can have a strong effect on the protein folding and on protein stability. Therefore, the protein melting-temperatures of Loop1 in comparison to hAGT and SNAP-tag were determined to assess the degree of destabilization (Figure 42). As expected, SNAP-tag turned out to be the most stable protein, with a melting point temperature of 69°C. The modifications introduced in Loop1 reduced the melting point by 7 degrees to 62°C. Loop insertion did decrease protein stability but compared to hAGT (melting point of 55°C) Loop1 is still more stable.



**Figure 42:** Loop1 is a stable protein. Melting-point analysis of hAGT, SNAP-tag and Loop1 showed that SNAP-tag possesses the highest stability, with a  $T_M$  of 69°C. Loop1 has a melting point of 62°C, which is seven degrees higher than the wild-type protein hAGT.

**Reactivity of Loop1 with BG-substrates.** The positions of loop-insertion in SNAP-tag were based on previous work done in our group<sup>154</sup>. Even though important structural units were left unchanged, the loop-insertion at position 159/160 of SNAP-tag had an influence on protein reactivity with BG-substrates. Loop1 was incubated

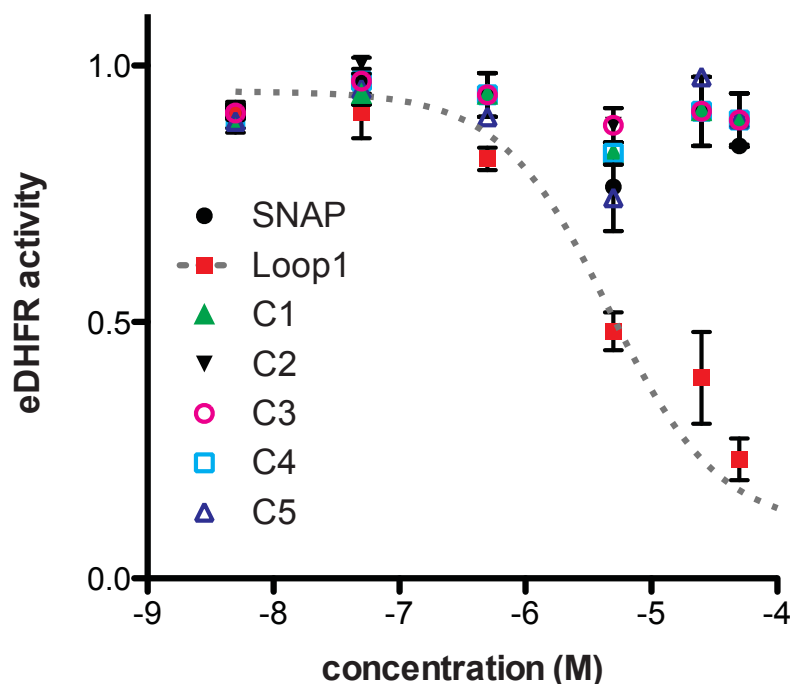
with four different BG-substrates and the labeling efficiency was recorded over time. The reactivity of Loop1 was impaired for all substrates (Figure 43).



**Figure 43:** Labeling efficiency of Loop1 with different BG-substrates. SNAP-tag labeling with BG-fluorescein (■; BG-FL) was taken for comparison. The insertion of Loop1-motive at position 159/160 had an impact on Loop1 labeling efficiency, which was decreased 28-fold for BG-FL. Reactivity with BG-TMR was 4-fold slower than reaction with BG-FL. The reaction with BG-substrates containing PEG-units as linker-moieties did not take place. Steric hindrance might be a possible explanation for this phenomenon.

BG-fluorescein (BG-FL) reacted 28-fold slower than SNAP-tag. Reaction with BG-TMR was around four-fold slower than reaction with BG-FL and incubation with BG-PEG-Mtx or BG-C4-malachite green (BG-C4-MG) did not yield any labeled protein. This result shows that the introduction of additional amino acids at position 159/160 of SNAP-tag had a major impact on reactivity. This result limited the scope of application for Loop1 as it was not reacting with BG-Mtx and therefore not possible to assemble the inhibitory complex intended for eDHFR inhibition. However, the inhibitory activity of Loop1 itself against eDHFR was tested in an activity assay.

**eDHFR activity assay.** eDHFR activity was determined by monitoring the reduction of dihydrofolic acid to tetrahydrofolic acid using NADPH as electron donor. The consumption of NADPH leads to a change in absorption at 340 nm. Upon addition of increasing amounts of Loop1 the catalytic activity of eDHFR was decreased (Figure 44). The apparent inhibitory concentration ( $IC_{50,app.}$ ) was in the micromolar range ( $\approx 10 \mu M$ ), which showed that the Y2H screen had identified a weak affinity binder. However, in the control experiment using increasing amounts of SNAP-tag no inhibition of eDHFR could be observed. As the C-terminal mutants C1-C5 had some degree of sequence similarity with Loop1 they were tested for eDHFR inhibition as well. Interestingly, the addition of increasing amounts of C-terminal SNAP-mutants did not result in inhibition of eDHFR activity.



**Figure 44:** eDHFR-activity assay with SNAP-tag, Loop1 and C-terminal mutants C1 to C5. Even though sequence similarities between Loop1 and C-terminal mutants exist, only Loop1 inhibited the catalytic activity of eDHFR



## 4 Discussion

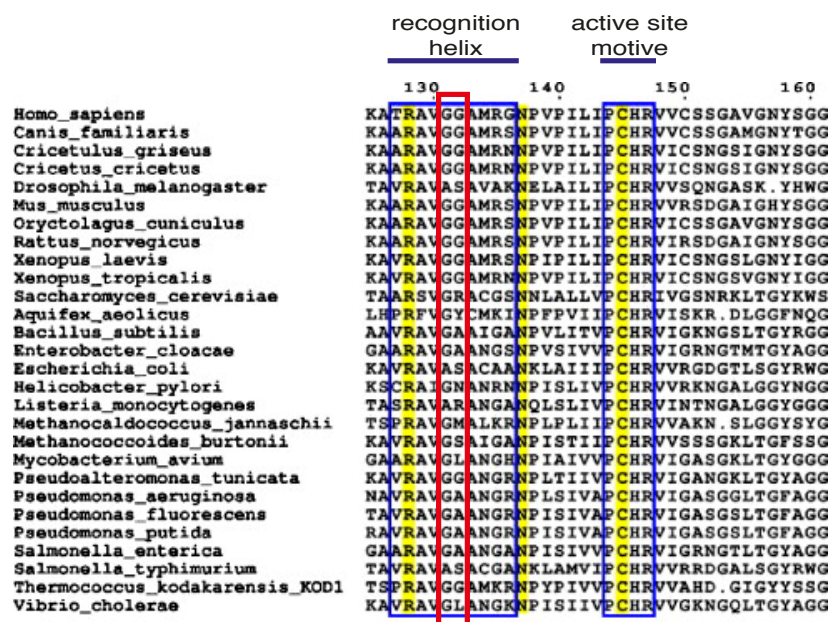
### 4.1 New insights on structural aspects of hAGT and SNAP-tag

In the first part of the current thesis we investigated the structural basis underlying the differences in protein stability between SNAP-tag and its parent protein hAGT. We could demonstrate that SNAP-tag displays not only an increased reactivity with O<sup>6</sup>-modified benzylguanine-derivatives but also had a superior stability than the parent molecule. Moreover, the combination of our crystallography and molecular modeling data provided further insight into the structural basis of the improved properties of SNAP-tag.

Generally, the evolution of proteins favors the selection of marginal stability with improvements in functionality<sup>168</sup>. It is also known from other directed evolution experiments, that activating mutations usually come at the cost of stability<sup>169 170</sup>. As most mutations are destabilizing and since evolution favors the most likely solutions over less likely ones, (directed) evolution generally favors mutants with marginal stability. Surprisingly, our study showed that the directed evolution of wild-type hAGT into SNAP-tag did not only increase the reactivity of the protein towards BG, but also affected the stability of unlabeled and, in particular labeled SNAP-tag. This gain of stability is even more remarkable when we consider that directed evolution experiments were focused on increased SNAP-tag reactivity, but not its higher stability. We believe that in the particular case of hAGT evolution, the reason for improved stability is due to the fact that higher stability also influenced the reaction rate of the protein. Furthermore, the increased *in vitro* stability translates well into prolonged protein half-life in mammalian cells, as no significant SNAP-tag degradation was detected at more than 40 hours upon labeling. This is in direct contrast to the behavior of wild-type hAGT, which becomes rapidly degraded (half-life of 3 hours) upon alkylation<sup>171</sup>. Reasons for this different behavior can be readily

explained by our analyses of the protein. The evolution from hAGT to SNAP-tag included 19 point mutations and the shortening by 25 amino acids at the protein's C-terminus. The analysis of SNAP-tag crystal structures showed as expected, that despite such drastic changes in the primary amino acid sequence, the overall protein-folding characteristic to hAGT was preserved. A detailed investigation of the single amino acid changes and the characterization of the intermediate mutants enabled us to identify important amino acid positions determining an increased stability and improved reactivity.

It is an intriguing particularity of alkylguanine-transferases in general to carry highly conserved glycine residues in the middle of an alpha helix, which plays an important role in substrate binding (the so-called recognition helix)<sup>150,151</sup>. Similarly to the human protein, AGT orthologues in other organisms share at least one or two glycine residues incorporated at the same position of the corresponding alpha helix (Figure 45). It has been argued that the role of the glycine residues at this position is to accommodate the natural substrate, an alkylated guanine flipped out of double-stranded DNA<sup>143</sup>.



**Figure 45:** Sequence alignment of AGTs from different organisms. Figure taken from Tubbs et al.<sup>172</sup> and modified.

However, changes introduced in AGT-54 at positions 131-135, including the exchange of Gly131 and Gly132 to Lys131 and Thr132, had a positive effect on helix

propensity and improved the overall protein stability. The shortened hydrogen bonds measured in the SNAP-tag crystal structures are further evidence for a local stabilization of the alpha-helical fold (refer to table of H-bond length). Additionally, there is evidence that the alpha helix 127-136 of hAGT loses its structure upon alkyltransfer, which is due to an increase in sterical restrain. Mutations introduced in SNAP-tag stabilize the alpha helix and influence the stability of the whole protein, even or especially after alkylation. Overall, our studies show that Gly131 and Gly132 are important residues that play a dual role in AGTs: accommodation of the substrate and contribution to unfolding and degradation of AGT upon alkyl transfer.

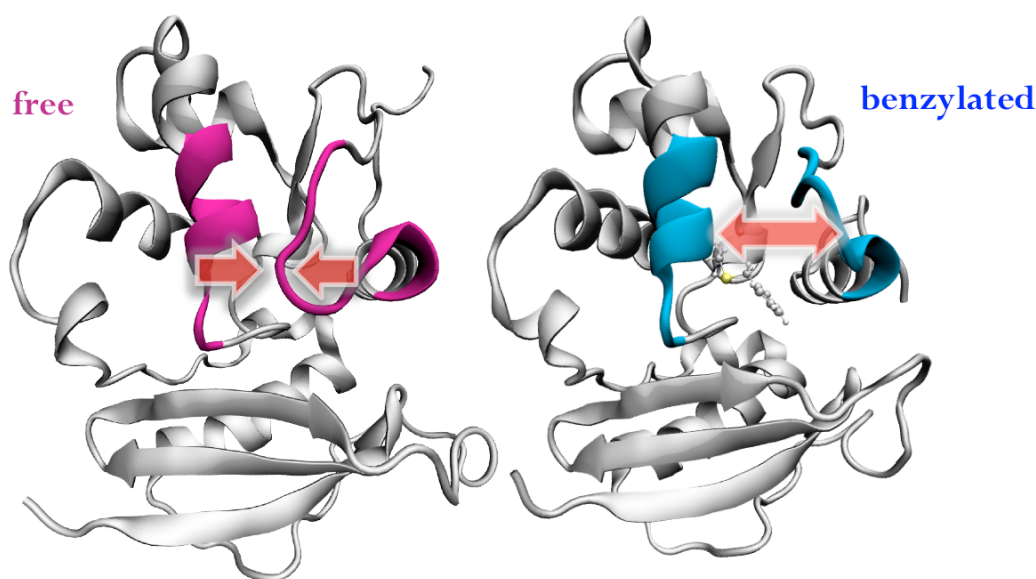
Previous studies on the improvements of thermal stability showed that random mutagenesis of structural parts with high flexibility (as protein loops) is a promising strategy to improve protein stability. In order to identify these regions, B-factor analysis of X-ray crystallographic structures has successfully been applied<sup>173</sup>. Further it was shown that mutations, which “fix” regions, can have a strongly positive effect on protein stability and were principally localized on the surface of the protein<sup>1</sup>. Despite the fact that our engineering efforts were not aiming at improvements in protein stability, we find the same mutation principles in SNAP-tag. Besides an increased reactivity, protein stability was improved by mutations in the region 150-154, which created a hydrogen-bonding network on the surface of the protein. The mutations were introduced in the last round of directed evolution from <sup>M</sup>AGT to SNAP-tag. The importance of these residues for protein stability was confirmed by measurements of melting point temperatures *in vitro* as well as by pulse-chase experiments in cell culture. The results of these experiments confirmed an increased stability of SNAP-tag as well as its alkylated form. Crystal structure analysis and our computational data show a major contribution from the mutations Asn150Gln, Asn152Asp and Ala154Asp. The stabilizing effect of these interactions manifest itself also in lower B-factors in the SNAP-tag crystal compared to hAGT.

Even though selection was focusing on reactivity, the stability of the generated mutants improved over the selection rounds. Especially in the last round of protein evolution (<sup>M</sup>AGT to SNAP-tag) the introduced mutations had a strong impact on both parameters. We believe that in the special case of alkyltransferases, the reason for the identification of more reactive mutants is due to the fact that the introduced mutations

stabilized the final product of the reaction, the alkylated protein. Improvements in protein stability might render the active site more accessible and improve reactivity.

#### 4.1.1 Conclusions and Outlook:

Our structural study on protein evolution allowed us to draw several conclusions about hAGT stability and denaturation. First, two point mutations introduced in wild-type AGT (Ser159Glu and Asn157Gly) improved reactivity but had a negative effect on protein stability (see <sup>GE</sup>AGT). Second, the exchange of Ser159 to Glu159 introduced additional interactions with the substrate, which improved reactivity. Third, computational analysis indicates that the loop (residues 157-159) and the recognition-helix (residues 127-136) in free hAGT are in close proximity and make contacts (data not shown). As soon as the alkylation takes place these interactions are lost, the domains drift apart and the recognition-helix starts to unfold (Figure 46).



**Figure 46:** Snapshots of the different states of hAGT before and after benzylation obtained by Molecular Dynamics Simulations. Important regions of the recognition helix and the concerned loop-part are highlighted in pink (free protein) and blue (benzylated protein). Red arrows indicate proximity of the two structural domains versus the increase in distance after benzylation. Figure by courtesy of E. Brunk and U. R  thlisberger, Laboratory of Computational Chemistry and Biochemistry, EPF Lausanne.

We believe that the close proximity of alpha helix and loop residues functions as a kind of trigger point that leads to hAGT denaturation upon protein methylation. This

structural motive (alpha helix-loop architecture) is also found in other alkyltransferases in various organisms. Additional work is currently undertaken to further characterize these interactions in the context of the onset of protein denaturation and helix stability.

## 4.2 Generation of novel protein binders based on the SNAP-tag scaffold

The second part of this thesis was aiming for the generation of a new SNAP-tag-based inhibitor complex against the two target proteins, eDHFR and Plk4. It was envisaged that this complex would interact with the target protein via amino acid loops and decrease its function only upon labeling with BG-inhibitor molecules. SNAP-tag was used as a scaffold into which randomized loop sequence were introduced at three different positions (residues 159/160 for library 1, residues 32/33 for library 2 and C-terminal for library 3). The utilization of two yeast-based systems, the yeast three-hybrid and two hybrid technologies, allowed for the differentiation of small-molecule dependent and independent binding interactions. It could be demonstrated that a specific protein-loop interaction could be generated by this approach. It could further be shown that inhibition of the catalytic activity of the target protein *E.coli* dihydrofolate reductase by a SNAP-loop mutant was possible. However, the identification of a small-molecule dependent interaction by Y3H has not been successful.

### 4.2.1 Lessons from library design:

Library design is one of the most critical steps in protein engineering. SNAP-tag loop-libraries were based on a previously applied protein design<sup>154</sup>. Based on insights from this work we decided to add a stretch of random amino acids, either at position 32/33 or position 159/160 of SNAP-tag.

The analysis of Loop1 and the mutants obtained from the Plk4 screening revealed that the modification at position 159/160 strongly reduced protein reactivity.

The reaction speed of Loop1 with BG-fluorescein was 28-fold lower than for SNAP-tag and labeling with substrates carrying PEG-units or other linker-moieties was not successful. For SNAP-tag it was known that the reactivity with BG-substrates having a linker-unit was reduced compared to the reactivity with BG-fluorescein or BG-

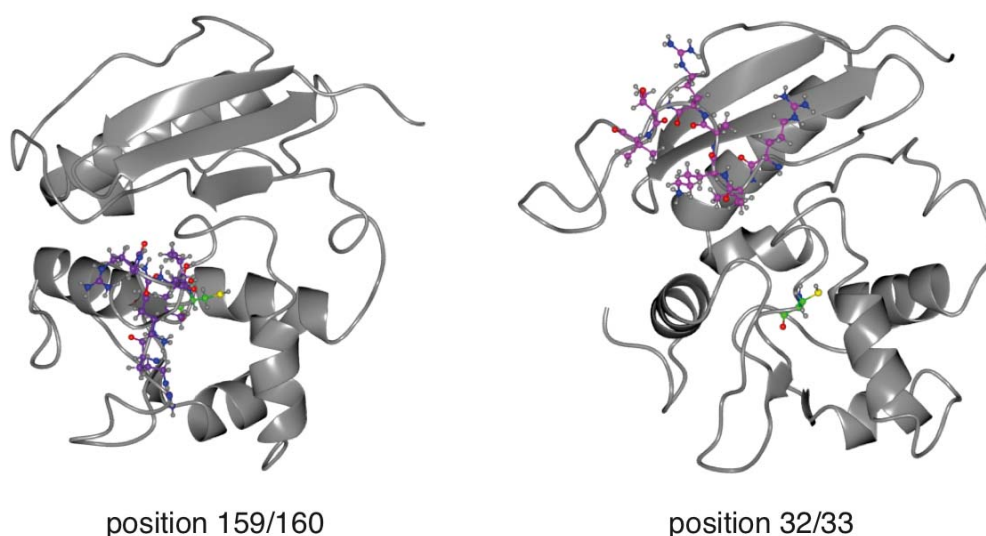
TMR. In these molecules a hydrophobic aromatic residue and a xanthene-moiety are directly attached to BG. A possible explanation for the good reactivity with these molecules is the way the directed evolution of SNAP-tag has been performed.

During the directed evolution of SNAP-tag BG-fluorescein or structurally very similar molecules have been used in combination with fluorophore-specific antibodies to isolate the most reactive mutants. As evolution favored this kind of molecules over other structures they represent the best substrates. As a consequence molecules that carry a linker-unit are less reactive. In the case of the loop-mutants the reduction in reactivity was that strong that over-night incubation with BG-molecules containing a linker-moiety did not yield a detectable amount of labeled protein.

For a better visualization of the loop positioning a computational model of Loop1 has been created (Figure 47 on the left). This model shows that the loop-part occupies the space that is in direct proximity to the active site. In hAGT this region is representing a part of the alkylguanine-binding channel. Our studies on hAGT and the intermediate mutants in the first part of this work showed that mutations in this part of the protein have an influence on protein reactivity, as mutations at position 157/159 in <sup>GE</sup>AGT improved reactivity because of the creation of new interactions with the substrate.

Compared to this, the loop insertion at position 32/33 in library 2 is much less affecting the active site and therefore protein reactivity. A model for position 32/33 is shown in Figure 47 on the right, which confirms that the active site is unoccupied by this loop modification. The results of the labeling experiment (see also chapter 3.6.1) are in agreement to this hypothesis, as they showed that mutants from library 2 reacted with all tested BG-substrates, even if they contained linker-units such as PEG.

However, SNAP-mutants from library 2 were less frequently identified as protein-binders than mutants from library 1. The loop sequences in library 1 might generally be presented in a better way for protein binding or the position of loop-insertion in this library is leading to a more structured loop, which would improve target protein binding.



**Figure 47:** Model of loop insertion in SNAP-tag for library 1 (position 159/160) and library 2 (position 32/33). The grafted peptide sequence (purple carbon atoms) is taken from Loop1 (RIKARTV) in both cases. The loop in library 1 occupies more space close to the active site cysteine (green carbon atoms), whereas the peptide insertion in library 2 seems to leave the active site unchanged. An impact on BG-reactivity for library 1 might be explained by this sterical influence.

#### 4.2.2 Conclusions and outlook:

The aim of SNAP-tag engineering was the generation of a protein-binder that would inhibit the target protein function. We envisioned that inhibition would depend on the labeling with a BG-small molecule and considered three different approaches, all based on loop-mutant binding. First, the labeling with a BG-inhibitor molecule would directly target the active site of the protein leading to inhibition. Second, the labeling with a BG-photosensitizers for Chromophor-Assisted-Light-Inactivation (CALI) would release upon radiation Reactive Oxygen Species (ROS) in close proximity to the target protein, leading to the disruption of protein function<sup>174</sup>. The use of a self-labeling protein-tag for CALI-applications has been reported recently<sup>175</sup>. The protein of interest was fused to the self-labeling protein-tag HaloTag and labeling with a suitable compound resulted in the depletion of protein function. Compared to this approach the binding of a reactive SNAP-mutant would have the advantage to target the native protein.

The third approach would allow the monitoring of the protein-binding event using fluorogenic substrates linked to BG. Changes in fluorescence signal would represent changes in the protein conformation due to protein binding. Such a biosensor would

allow to monitor the protein dynamics inside living cells and would give precise information about the spatiotemporal dynamics of the target protein. A prerequisite for such an approach is that the binding event depends on the activity state of the target protein. A similar approach has been used by Nalbant *et al.* to monitor the timing and localization of active Cdc42, a small GTPase of the Rho-subfamily<sup>176</sup>. However, the isolation of a loop mutant with good reactivity towards BG-molecules is a prerequisite and further work is needed to isolate mutants with good reactivity.

The use of the Y3H technique was not successful in the identification of small-molecule dependent interactions. However the Y2H screening identified protein-binders whose interaction could be confirmed *in vitro*. The library that gave the best hit (library 1) carried the loop modification at a position that rendered the protein (partially) inactive, as no labeling with BG-inhibitor derivatives could be achieved.

The decreased reactivity strongly limited the usefulness of the isolated mutants and was the main reason to not further pursue the experiments with the hits of loop library 1 obtained in the Plk4-screening. However, these sequences should be analyzed for phosphorylation by Plk4 to see if the screening identified substrate peptides.

For further protein engineering the positioning of the loop and its impact on reactivity have to be optimized. For the successful generation of a small-molecule dependent inhibitor of protein function it is important to choose a loop position that is in proximity to the active site and additionally does not have a negative impact on reactivity. As we have seen that the modification at position 32/33 was well accepted by the protein, one could envision the generation of a protein library carrying longer loop-sequences at this position to improve surface exposition and binding abilities. Alternatively, one could try to elongate SNAP-tag N-terminal to the recognition helix, between residues 125/126. From our experience on SNAP-tag engineering we know that point mutations in this region are accepted. Further, the positioning should ensure a good loop exposure on the protein surface and the substrate entry side should be less occupied than for library 1.

Strength of the Y2H-technology is that weak interactions or complexes with low on/off-rates can be detected. This is extremely important for the identification of biological interactions as for example during the analysis of whole proteomes. In the case of our experiment, a more direct control over the binding conditions could have



facilitated the isolation of SNAP-mutants that react with BG-inhibitor molecules and at the same time bind to the protein of interest. To improve the screening output the focus needs to be put on methods that allow the isolation of mutants that either show different binding affinities dependent on protein labeling, or that react well with diverse BG-derivatives to bring the labels close to the target protein. The latter would be important for approaches as CALI or the use of fluorogenic probes as their working principle relies on the close proximity to the target protein. Selections based on protein display, on phage or bacteria, could represent an interesting alternative to the Y2H screening, as libraries could be screened for reactivity by Fluorescence Activated Cell Sorting (FACS) prior, or in combination with binding affinity screening



## 5 Materials and Methods

### 5.1 Structural analysis of hAGT and mutants:

**SNAP-tag, hAGT and SNAP-mutant expression and purification.** hAGT, SNAP-tag and all mutants were cloned into pRSET vector (kind gift of Prof. Dr. Michael Groll) carrying an N-terminal hexa-histidine-tag (his<sub>6</sub>-tag) followed by a Prescission protease cleavage site. The plasmids were transformed into E.coli BL21 DE3 and directly used to inoculate a 20 mL pre-culture in LB containing 100µg/mL ampicillin. The pre-culture was diluted 50-fold and cells were grown at 37°C until an OD of 0.6-0.8. At this cell density protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After 16h at 18°C the cultures were harvested, the pellets were taken up in PBS supplemented with protease inhibitor cocktail (Roche) and lysed by sonication. The soluble fraction was used for Ni-NTA purification according to the manufacturers protocol (Qiagen). Protein was eluted in 50 mM K<sub>2</sub>HPO<sub>4</sub>, 150 mM imidazol, 10% glycerol, pH 8.0. The eluted protein was incubated over night with Prescission protease and 10 mM DTT at 4°C. Subsequently, it was subjected to a ResQ 6 mL column (GE Healthcare) using buffer A (20 mM TRIS, 4 mM DTT, pH 8.0) and buffer B (20 mM TRIS, 4 mM DTT, 1 M NaCl, pH 8.0). The corresponding elution fractions were pooled and concentrated. Homogenous protein was obtained by size exclusion chromatography (Superdex 200 column, GE Healthcare) using a buffer containing 20 mM TRIS, 4 mM DTT, 200 mM NaCl, pH 8.0. The main fractions of the elution peak were pooled and concentrated to 5-8 mg/mL. One aliquot of protein was directly used to assess protein stability by thermal denaturation. The other aliquots were supplemented with 30% glycerol and stored at -20°C until further use.

**Trypsination.** Protein solutions of hAGT, SNAP-tag and mutants were diluted to 1 mg/mL in reaction buffer (20 mM TRIS, 4 mM DTT, 200 mM NaCl, pH 8.0) and separated in two aliquots. One aliquot of each protein was incubated with a 2-fold molar excess of BG-fluorescein for 1h at room-temperature, to the second aliquot the corresponding amount of DMSO was added. To check if the labelling reaction was completed a small aliquot was taken and quenched with 40-fold excess of BG-Cy5. Samples were analysed on SDS-PAGE to confirm complete labelling of the proteins prior to trypsination. Per sample 50  $\mu$ L (50  $\mu$ g) protein were mixed with 50  $\mu$ L of different trypsin dilutions (trypsin stock were 1.0 mg/mL (1:1), 100  $\mu$ g/mL (1:10), 10  $\mu$ g/mL (1:100), 1.0  $\mu$ g/mL (1:1000), 0.1  $\mu$ g/mL (1:10000) and 0  $\mu$ g/mL). Samples were incubated for 1h at room temperature and quenched by the addition of 6x SDS-sample buffer and heating for 3 minutes at 95°C. Samples were analysed by SDS-PAGE followed by fluorescence-in-gel scanning on a Pharos FX Molecular Imager (Bio-Rad) and subsequent coomassie staining.

**Thermal denaturation assay.** Proteins were diluted to 25  $\mu$ M stock solutions in reaction buffer (20 mM TRIS, 4 mM DTT, 200 mM NaCl, pH 8.0) and either pre-incubated with 2 equivalents of BG or the corresponding amount of DMSO for 1 hour. A small aliquot was taken and quenched with 30-fold excess of BG-fluorescein. The samples were analyzed by SDS-PAGE to ensure complete labelling prior to the experiment. 10  $\mu$ L of the stock solutions were mixed with 10  $\mu$ L of 30x SyproOrange (Invitrogen). The measurements were performed in MicroAmp Fast 96-Well Reaction Plates (Applied Biosystems) on a 7900 HT Real Time PCR System (Applied Biosystems). Heating ranged from 20-95°C at a ramping rate of 1%. Changes in fluorescence were monitored for all wells simultaneously using a charge-coupled device (CDD) camera. Dissociation curves were analyzed by plotting the first derivative of the curves (slope) against the temperature ( $dF/dT^\circ$ ) averaged over 4°C increments. The maximal value gave the melting temperatures ( $T_M$ ) of the proteins.

**Pulse-chase experiment.** All constructs were cloned into pECFP-Nuc plasmid for mammalian cell expression using NheI/BamHI restriction sites. HEK 293 cells were cultured in EX-CELL 293 medium (Sigma) and transfected with the corresponding

plasmids using polyethylenimine (PEI). For this, cells were transferred to fresh HEK 293 medium one day prior to the transformation at a cell density of  $1.5 \times 10^6$  cells/mL. The day of the transfection cells were pelleted by centrifugation at 800 rpm for 5 minutes and resuspended in RPMI1640 + GlutaMAX (GIBCO)+ 0.1% Pluronic F68 solution (Sigma) at a density of  $2 \times 10^7$  cells/mL. 12.5  $\mu$ g plasmid per millilitre cells were added plus 75  $\mu$ L of 1  $\mu$ g/ $\mu$ L PEI. Cells were incubated at 37°C under shaking (200 rpm). After 2 hours cultures were diluted 20-fold in fresh HEK 293 medium. Cells were cultured for 24 hours prior to the experiment. Cells were labelled for 15 min with 0.65  $\mu$ M CP-TMR-star (Covalys) and blocked with 500  $\mu$ M BG. Extensive washing with 500  $\mu$ M BG-containing media for 30 minutes and 100  $\mu$ M BG-containing media for 3 hours blocked labelling and removed non-reacted dye sufficiently. Samples were taken at different time points for protein extraction and SDS-PAGE analysis followed by fluorescence-in-gel scanning (Pharos FX Molecular Imager, Bio-Rad).

**Molecular Dynamics calculations.** Computational models of SNAP-tag and wild-type hAGT were constructed, starting from crystallographic non-benzylated (pdb entries 3KZY and 1EH6, respectively) and benzylated structures (pdb entries 3L00 and 1EH8, respectively) in which the unresolved loop region (residues 36 to 44 in hAGT and residues 36 to 49 in SNAP-tag) was added manually. AMBER parm99SB charges and atom types were used to build the topologies for each of the structures with zinc parameters, where distance restraints were used for the coordination of the zinc ion to the cysteine residues in the N-terminal domain. The structures were solvated in a periodically repeated TIP3P water box with dimensions 81 X 80 X 70  $\text{\AA}^3$  (corresponding to a 15  $\text{\AA}$  solvation shell around the protein). All structures were minimized, heated to 300K under constant volume conditions with 5.0 kcal mol<sup>-1</sup> positional restraints on the protein (except the unresolved loop), followed by equilibration under constant temperature and pressure conditions, slowly releasing the restraints over 4 ns. Data were collected from production phase simulations, in which Molecular Dynamics (MD) trajectories were run for 30-50 ns using the Particle Mesh Ewald (PME) MD module in AMBER v10. During the simulations, the PME method was used with a cut off of 8.0  $\text{\AA}$  for non-bonding interactions. Constant pressure periodic boundaries conditions were maintained with a pressure relaxation time of 2

ps. The SHAKE algorithm was used to constrain all bonds that involve hydrogen atoms. The Langevin method for temperature control was used with a collision frequency of  $1 \text{ ps}^{-1}$ .

Free energy perturbation methods, in combination with thermodynamic integration (FEP/TI), were used within the Simulated Annealing with NMR-derived Energy Restraints (SANDER) module of AMBER v10. Using a dual topology paradigm, two topologies (state 0 and state 1) for each transformation were constructed by manually imposing point mutations on the equilibrated MD structures of the free enzymes, SNAP-tag and hAGT. The electrostatic and Lennard Jones terms were decoupled by performing three separate alchemical transformations: (i) decharging state 0 using 30-40 lambda points (ii) transforming the atoms of state 0 into those of state 1 using 20 lambda points and (iii) recharging state 1 using 30-40 lambda points. Convergence was tested for by extending the duration of the MD run and by increasing the number of lambda points. In each transformation, all lambda points were individually minimized, equilibrated, and data were collected during a production phase of 1 ns. The change in the potential energy as a result of the perturbation was integrated over the lambda values to obtain the  $\Delta G$  for each mutation. A per residue decomposition for the charging and de-charging steps was performed to indicate the primary points in the protein that are affected by the mutation.

## 5.2 Engineering of SNAP loop-mutants:

**Standard reagents.** The standard protocols in yeast biology were taken from the textbook “Short protocols in molecular Biology<sup>177</sup>”. If not specified elsewhere, the same source was used for the preparation of standard reagents for molecular biology experiments.

**Yeast strain.** The reporter yeast strain NMY51C3 was derived from NMY51 (Dualsystems Biotech, Switzerland) by gene disruption of PDR5 and SNQ2, two genes coding for a multi-drug resistance transporter for better uptake of small

molecule compounds. NMY51C3 genotype: *MATa his3-Δ200 trp1-901 leu2-3,112 LYS2::(lexAop)<sub>4</sub>-HIS3 ura3::(lexAop)<sub>8</sub>-lacZ (lexAop)<sub>8</sub>-ADE2 GAL4 Δpdr5::loxP Δsnq2::loxP*. Reporter genes includes two auxotrophic (HIS3 and ADE2) and one colorimetric (lacZ).

### 5.2.1 Library construction and screening set-up:

**SNAP-tag libraries.** SNAP-tag loop libraries were produced by PCR using degenerated primers. Loop insertion at amino acid position 159/160 (library 1) of SNAP-tag was realized with the following scheme: A first PCR reaction was performed using BM01/BM04, a second, separate PCR reaction with BM02/BM03. After purification of the PCR products by gel-extraction an assembly-PCR of the two PCR-products that had partially overlapping fragments was performed using primers BM03/BM04. For loop insertion at amino acid position 32/33 (library 2) of SNAP-tag the same scheme was used with different primers pairs for the first (BM04/BM05) and second (BM03/BM06) PCR-reaction. Assembly-PCR was performed using the same primers as above (BM03/BM04). The full-length products were digested with restriction endonuclease SfiI for 5h at 50 °C, purified by gel-extraction and ligated into pGAD-HA following an optimized protocol using three equivalents of insert to one equivalent linearized vector (1μg vector per ligation reaction) at 21°C for 4h. Large-scale transformation of 10 μg library DNA into XL1-blue was performed by stepwise electroporation of 0.5 μg DNA per 100μl cells. Transformation mixes were pooled, incubated at 37°C for 1 hour and plated on twenty 15 cm agar plates. After incubation ON 5 mL LB medium, 25% glycerol were added on each plate, the cells were scraped off and collected in an 50 mL Erlenmeyer. Transformation efficiency was determined for each library and was determined to be  $3 \times 10^6$  colonies per μg DNA for library 1 and  $2.2 \times 10^6$  colonies per μg DNA for library 2 and a total complexity of  $3 \times 10^7$  different clones for library 1 and  $2.2 \times 10^7$  different clones for library 2. Cells were aliquoted and stored at -80°C for subsequent large scale DNA preparation (MaxiPrep, Qiagen) according to the manufacturers protocol.

**Bait vectors.** Constructs for eDHFR, eDHFR mutant 1, eDHFR mutant 2 and eDHFR mutant 3 expression in yeast were amplified by PCR using the primer pair BM09/10 and cloned into the vector pLexA using the SacI/SalI restriction sites. The mutants differ in the following point mutations from wild-type eDHFR: mutant 1 has point mutation of L54I, mutant 2 carries the mutation L54G and mutant 3 has the mutations of L54G plus F31V. All those mutations influence the binding affinities to the substrate and NADPH. Constructs of Plk4 full length (Plk4-fl), a short version of the kinase domain (Plk4-KD-short, 795 bp) and a longer version of the kinase domain (Plk4-KD-long, 1170 bp) were cloned into pLexA-N vector using the primers BM11/BM12 for full length construct, BM11/BM13 for the short fragment of the kinase domain and BM11/BM14 for the long kinase fragment and SacI/SalI restriction sites.

**Bait expression control in yeast.** After transformation of the corresponding plasmids and two control plasmids (empty pLexA, pLexAp53) into the reporter yeast strain NMY51C3 the cells were streaked out on CM-W agar plates and incubated at 30°C for 2 days. A single colony of the strain was re-suspended in CM-W liquid medium and grown overnight at 30°C and intensive shaking. The next morning, the cultures were diluted to an OD<sub>600</sub> of 0.4 and was shaken at 30°C for an extra 5-6 hours. The optical density OD<sub>600</sub> of the culture was determined and the equivalent of 1 mL of an OD<sub>600</sub>=1 culture (= 1 OD<sub>600</sub>) was pelleted in a 1.5 mL tube. The cell pellets were washed in 1mL water, resuspended in 50 µL ice-cold 1.85 M NaOH supplemented with 7.4 % beta-mercaptoethanol. After incubation on ice for 10 minutes 50 µL of 50% trichloroacetic acid were added to each tube and incubated again for 10 minutes on ice. Samples were centrifuged at maximal speed for 2 minutes, the supernatant was discarded and the pellets resuspended in 10 µL of 1M Tris base. 40 µL of 2x SDS-sample buffer were added and samples were boiled at 95°C for 5 minutes. After centrifugation at maximal speed for 2 minutes 20 µL of the supernatant were loaded on a 15% SDS-PAGE gel.

The proteins were then transferred onto a PVDF membrane (Immobilion-P, Millipore) using a semi-dry transfer system according to the membrane manufacturer's instructions (Millipore Immobilon-P transfer membrane user guide). After transfer,



the membrane was blocked for 1h in TBST (20 mM Tris-HCl, 500 mM NaCl, 0.005% v/v Tween-20, pH 7.5) + 5% skim milk powder. After washing 2 times with TBST for 5 minutes, the membrane was incubated with anti-LexA mouse antibody (stock at 200 µg/µL, Dualsystems Biotech) diluted 1:5000 in TBST + 5% skim milk powder for 45 minutes at room temperature then for 11h at 4°C. After washing 2 times with TBST, the membrane was incubated with an anti-mouse antibody-HRP conjugate (A4416, Sigma) diluted 1:2000 in TBST + 5% skim milk powder for 1.25 h at RT. The detection of the immunoblot was performed using ECL Plus Western Blotting Detection Reagents (GE Healthcare) and a Kodak Image Station 440CF. The quantification of the chemi-luminescence signal was performed using the Kodak 1D software.

**Quantification of labeling efficiency of SNAP-tag with BG-derivatives in yeast cells.** Yeast cells were transformed with the plasmid pLexA-CLIP-SNAP and grown in CM-WL medium. Transformed yeast cells were grown to an OD600 of 0.8-2.0. For liquid labeling  $1 \times 10^8$  cells were harvested by centrifugation at 1,500x g for 5 minutes and then washed in 1 volume PBS pH 6.5. The cell pellets were next re-suspended in 100 µL PBS and BG-derivatives were added to a final concentration of 10 and 50 µM. In order to quantify the full BG-Cy5 labeling, a control sample was included in which no BG-derivative was added (DMSO only). The suspension was then shaken in a benchtop thermomixer (Eppendorf) at 1,200 rpm at 30°C for 3 hours. The cells were harvested, washed in PBS and taken up in 120 µL labeling mix (50 mM HEPES pH 7.2, 1mM DTT, 2.5 µM BG-Cy5 and 2.5 µM BC-Cy3 (Covalys). A volume of 200 µL of acid-washed glass beads(Sigma) was added to the suspension and yeast cells were disrupted by full-speed vortexing for 5 x 1 minute with 1 minute incubation on ice between each cycle. The disrupted cells were incubated for 30 minutes at RT for quantitative labeling of SNAP-tag and CLIP-tag. After centrifugation at 16,000x g for 5 minutes, 50 µL of the supernatant were added to 10 µL 6x SDS sample buffer. After boiling at 95°C for 5 minutes the samples were loaded on a 15% SDS-PAGE gel.

NMY51 C3 cells expressing the ACT-AGT fusion protein were grown in liquid medium (CM-WL) ON and plated out on 1% agarose plates containing different BG-derivatives in 10 µM concentration to compare labeling efficiency in liquid medium

and on agar plates. As a control, a plate not containing any BG-derivative (DMSO only) was included. Cells were incubated for 4 days at 30°C, resuspended in 1-2 mL PBS and  $1 \times 10^8$  cells per sample were washed and resuspended in 50 mM HEPES pH 7.2, 1 mM DTT. Cells were lysed by glass-beads in the presence of 200  $\mu$ M BG-Cy5 and BC-Cy3 (Covalys). After centrifugation at 16,000x g for 5 minutes, 50  $\mu$ L of the supernatant were added to 10  $\mu$ L 6xSDS sample buffer. After boiling at 95°C for 5 minutes the samples were loaded on a 15% SDS-PAGE gel.

In-gel fluorescence scanning of dyes BC-Cy3 and BG-Cy5 was performed on a PharosFX Molecular Imager (BioRad) and intensities of fluorescence were quantified using Quantity One software (BioRad). The labeling efficiency of SNAP-tag with the different BG-derivatives in living yeast cells was then calculated based on the signal intensity of the control labeling samples (DMSO only), which was set to 100% of labeling.

**Optimization of screening conditions.** The screening conditions were optimized according to the different levels of auto-activation of the different baits proteins. The yeast reporter strain was transformed with pLexA-N-bait constructs together with pGAD-HA-SNAP and plated on selective (CM-HWL) and non-selective (CM-WL) agar plates. After 5 days of incubation at 30°C the plates were analysed. Additionally, large-scale transformations according to the final screening conditions were performed following the protocol of Gietz et al<sup>178</sup>. As screening conditions 2.5 mM 3-aminotriazole were chosen to suppress any unspecific reporter gene activation due to the bait-proteins and quantity of yeast cells plated on screening plates.

**High efficiency transformation of yeast with SNAP-tag loop libraries.** The yeast strain NMY51C3 was pre-transformed with the SNAP-tag bait plasmid (pLexA-SNAP) according to standard procedures. The bait expressing yeast strain was then transformed with the SNAP-tag libraries achieving at least  $1 \times 10^6$  transformants per library. All three libraries were transformed using a high efficiency protocol from R.D. Gietz et al<sup>178</sup>. Transformations using the 10x TRAFO scale were performed each with 5  $\mu$ g plasmid DNA. The TRAFO protocol was followed except for the heat

shock time which was increased to 40 minutes at 42°C. The transformed cells were homogenously spread (4 mm diameter glass beads) on 5 to 10 large 150 mm Petri dishes containing CM-WL growth medium. The transformation efficiency for each library was determined by plating dilutions of the transformed cells onto CM-WL agar plates. After 2-3 days at 30°C, the yeast cells were re-suspended in 1xTE and pooled. The pooled yeast cells were washed twice with 1 pellet volume 1xTE and re-suspended in 1 pellet volume freezing solution (65% glycerol, 100 mM MgSO<sub>4</sub>, 25 mM Tris-HCl pH 8.0). The 50% slurry was aliquoted and stored at -80°C. All manipulations with yeast cells from the beginning of the transformation procedure to the preparation of aliquots of the transformed yeast cells were performed under sterile conditions in a laminar flow hood.

### 5.2.2 Yeast screening:

**Yeast growth media.** Standard media was prepared as described in “Short Protocols in Molecular Biology<sup>177</sup>”. The preparation of specific growth media for Y3H applications is described hereunder. All media were supplemented with 0.1 g/l adenine (NMY51-type yeast are ade2 mutant).

**Preparation of non-selective growth medium (CM -LW).** Non-selective medium for Y3H was composed of 1.4 g/l dropout mix -HLWU (Sigma), 0.1 g/l uracil, 0.05 g/l histidine, 0.1 g/l adenine, 1.7 g/l yeast nitrogen base, 5 g/l ammonium sulphate, 20 g/l D-glucose (added after autoclaving), 15 g/l agar and adjusted to pH 5.6 with NaOH.

**Preparation of selective growth medium (CM -HLW).** Selective medium for Y3H was composed of 1.4 g/l dropout mix -HLWU (Sigma), 0.1 g/l uracil, 0.1 g/l adenine, 1.7 g/l yeast nitrogen base, 5 g/l ammonium sulphate, 20 g/l D-glucose (added after autoclaving), 15 g/l agar and adjusted to pH 5.6 with NaOH.

**Preparation of selective growth medium for Y3H positive selection (CM -HLW + 2.5 mM 3-AT + 10  $\mu$ M BG derivative).** Selective medium for Y3H was composed of 1.4 g/l dropout mix -HLWU (Sigma), 0.1 g/l uracil, 0.1 g/l adenine, 1.7 g/l yeast nitrogen base, 5 g/l ammonium sulphate, 20 g/l D-glucose (added after autoclaving), 10 g/l agarose and adjusted to pH 6.5 with NaOH. Furthermore, the medium was adjusted after autoclaving ( $T < 50^\circ\text{C}$ ) to 2.5 mM 3-AT from a 1 M stock in H<sub>2</sub>O and to 10  $\mu$ M BG drug derivative from a 10-20 mM stock in DMSO.

**Y2H and Y3H selection of affinity mutants.** Each selection (1 library versus one bait protein) was performed on 1 large petri dish (150 mm diameter) containing CM-HWL + 2.5 mM 3-amino-1,2,4-triazole (3-AT) + 1% agarose, pH 6.5. The screening was performed as Y2H screening without the incorporation of a BG-small molecule and as Y3H screening in which the agar plates were supplemented with 10  $\mu$ M of BG-small molecule. Yeast cell stocks containing the corresponding bait protein together with one library were thawed and approximately  $1 \times 10^7$  colony forming units (cfu) were diluted in 1xTE buffer and homogeneously spread (4 mm diameter glass beads) on the selection plates. After 5-7 days at  $30^\circ\text{C}$ , all colonies larger than 1 mm in diameter were picked using plastic disposable 200  $\mu$ L pipette tips and resuspended in 100  $\mu$ L 1xTE in 96-well plates. Using a multi-channel pipette the re-suspended colonies were transferred onto large non-selective agar plates (CM-WL). After 3 to 4 days at  $30^\circ\text{C}$ , the arrayed colonies were transferred back to 100  $\mu$ L 1xTE in 96-well plates and re-plated using a Singer RoToR HDA robot. The respotting on selective (CM-HWL) media with or without BG-small molecule derivative (10  $\mu$ M) assessed the small molecule dependency of the interaction.

**Identification of screening hits.** All yeast spots showing growth after re-spotting were picked from the corresponding non-selective plates by manual colony picking using disposable 200  $\mu$ L pipette tips and re-suspended in 100  $\mu$ L 1xTE in 96-well plates. Using the RoToR HDA robot the re-suspended colonies were transferred onto CM-HWL agar plates containing increasing amounts of 3-AT (0  $\mu$ M, 1  $\mu$ M, 2.5  $\mu$ M, 5  $\mu$ M and 10  $\mu$ M). After 4 days of growth at  $30^\circ\text{C}$  the colonies were selected according to their resistance towards 3-AT. Colonies growing at 2.5  $\mu$ M or higher 3-AT concentrations were considered as putative hits and inoculated in 1.5 mL CM-WL

liquid medium in 96-well deep well plates. After growing until saturation, the cultures were used for plasmid DNA-preparation. Yeast cells were centrifuged at 1,000x g for 10 minutes, the pellets resuspended in 200  $\mu$ L zymolyase solution (1.2 M sorbitol, 100 mM sodium phosphate pH 7.4, zymolyase 20T 2 mg/mL (Seikagaku Biobusiness) and incubated at 37°C for 90 minutes with shaking. The plasmid DNA was then isolated by MiniPrep (Quiagen) or by using a DNA isolation kit in 96-well format (NucleoSpin M-96 Plus Plasmid, Macherey-Nagel). The instructions from the manufacturer were followed starting from the addition of the alkaline lysis solution to the yeast suspension pre-treated with zymolyase. The isolated plasmids were transformed in chemical competent E.coli XL1-blue cells to increase the plasmid yield and to ensure, that only one plasmid per colony is propagated. The plasmids were again isolated using a DNA-isolation kit (Qiagen or Macherey-Nagel) and sequenced.

**Confirmation of the yeast hits.** A confirmation of the Y2H interaction using retransformed yeast strains was performed for all sequenced plasmids coding for a mutant of SNAP-tag. Plasmids were transformed into yeast already carrying the corresponding pLexA-bait plasmid using a standard yeast transformation protocol. A single yeast colony for each transformed yeast strain was diluted into 100  $\mu$ L 1x TE in 96-well plates. The yeast suspensions were serially diluted 10-fold 4 times. 2  $\mu$ L of the suspensions and dilutions were spotted each on CM-WL, CM-HWL and CM-HWL + 2.5 mM 3-AT agar plates using a RoToR HDA robot. The plates were incubated at 30°C for 3-5 days prior to growth analysis. Yeast colonies that grew on selective plates were considered as reconfirmed potential hits in yeast and submitted to further testing.

**Bait dependency test of potential hits in yeast.** The specificity of Y2H interactions was tested by transformation of different pLexA-bait plasmids in yeast that was already expressing one pGal4-SNAP-mutant at a time. The transformations were performed according to standard protocol. A single yeast colony for each transformed yeast strain was diluted into 100  $\mu$ L 1x TE in 96-well plates. The yeast suspensions were serially diluted (10-fold). 2  $\mu$ L of the suspensions and dilutions were spotted each on CM-WL, CM-HWL and CM-HWL + 2.5 mM 3-AT agar plates using a

RoToR HDA robot or manually. The plates were incubated at 30°C for 3-5 days prior to growth analysis. Yeast colonies that grew only in the presence of the right bait protein on selective plates were considered bait specific potential hits in yeast and submitted to further testing.

### 5.2.3 Hit-validation:

**DHFR activity assay.** eDHFR was expressed and purified from *E.coli*. The construct for N-terminal (his)<sub>6</sub>-tagged eDHFR was created by PCR using the primers BM17/BM18. The PCR product was ligated in pET-15b vector (Novagen) via BamHI sites. Activity of eDHFR was measured based on NADPH consumption causing a decrease of absorption at 340 nm. 5 nM eDHFR were added to a solution containing 80  $\mu$ M NADPH and 200  $\mu$ M dihydrofolic acid (DFH<sub>2</sub>) (both Sigma) in reaction buffer (50 mM Tris-HCl pH 6.8, 0.01% Triton-X-100, 10 mM beta-mercaptoethanol, 0.5 mg/mL BSA). The assay was performed in a 96-well micro test plates (Ratiolab) and measured on a micro plate spectrophotometer (SpectraMAX, Molecular Devices). For inhibition studies serial dilutions of control (SNAP-tag) and inhibitor (Loop1 mutant or Methotrexate) were pre-incubated with eDHFR for 30 minutes prior to the addition of NADPH and DFH<sub>2</sub>. Samples were measured every 15 seconds for 1 hour and evaluated by plotting the initial slopes against the corresponding inhibitor concentration.

**Cross-linking experiment.** Proteins to be tested were expressed as SNAP-tag-fusion proteins either in *E.coli* cells or HEK293 cells and further purified via his-tag, GST-tag or FLAG-tag according to the manufacturers protocols. Proteins were dialyzed in 20 mM Tris, pH 6.8, 50 mM NaCl, 1 mM beta-mercaptoethanol supplemented with 0.1% BSA. For the experiment 200 nM protein A were mixed with 200 nM protein B in dialysis buffer to a total volume of 100  $\mu$ L and incubated for half an hour at room temperature. 1  $\mu$ M of BG-TMR-BG cross-linker (LR 223) or CP-TMR-CP (LR222) were added and mixes were incubated for 1h at 37°C. 6x-SDS-sample buffer was added and reactions were heated for 3 minutes at 95°C. After cooling down of the samples, 30  $\mu$ L were loaded on a 5-15% gradient Tris-buffered gel and SDS-PAGE

analysis was performed. Fluorescence intensities were recorded on a PharosFX Molecular Imager (BioRad) and bands confirmed by Western blotting against GST-, and his-tag following the manufacturers recommendations.

**Loop mutant labeling efficiency.** The identified loop mutants from the Plk-4 screening were analyzed for their labeling efficiency in *E.coli* BL21 DE3 cells. Expression vectors were prepared using Gateway recombination cloning (Invitrogen) with primer pair BM19/20, pDONR221 and destination vector pDEST15. Cells were lysed by sonication (2-times 5 minutes, 50 % duty, 80% power), supplemented with protease inhibitor cocktail (Roche) and incubated with 5  $\mu$ M of BG-dye for 30 minutes at room temperature. Three different BG-dyes were tested: BG-fluorescein (New England Biolabs), BG-PEG<sub>(4)</sub>-DAP<sub>(488)</sub>-COOH (Dr. Luc Raymond, Johnsson laboratory) and BG-PEG<sub>(7)</sub>-DAP<sub>(Rh110)</sub>-NH<sub>2</sub> (Dr. Luc Raymond, Johnsson laboratory). 50  $\mu$ L of the lysate were mixed with 10  $\mu$ L of 6x SDS-loading dye and samples were boiled for 3 minutes at 95°C. 15  $\mu$ L of each sample were loaded on a 15% SDS-PAGE gel and analyzed by in-gel fluorescence scanning on a PharosFX Molecular Imager (BioRad).

**GST-pulldown.** GST-constructs of identified SNAP-mutants were made by cloning into pGex-2T vector using the primer pair BM07/BM15 for the loop motive RIKARTV and BM07/BM16 for the C-terminal SNAP-mutants (BamHI/EcoRI cloning sites). Proteins were expressed in *E.coli* BL21 DE3 cells using standard LB medium, induction at OD<sub>600</sub>= 0.6 with 0.5 mM IPTG and expression overnight at 16°C. Cells were lysed and cleared by centrifugation at full speed for 30 minutes. The supernatant was collected and glycerol added to a final concentration of 15% (v/v). Lysate stocks were either used directly or stored at -80°C. For the pulldown experiment 25  $\mu$ L of 20% GST-sepharose beads slurry (GE Healthcare) were added to 50  $\mu$ L GST-SNAP or GST-mutant lysate and incubated under rotation for 30 minutes at 4°C. The beads were washed three times in pulldown buffer (50 mM Tris-HCl pH 7.9, 150 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, 1 tablet inhibitor cocktail (Roche) for 10 mL buffer) and resuspended in pulldown buffer containing 10  $\mu$ M BG-methotrexate. Mix was incubated for 30 minutes at room temperature and washed

three times with buffer. Purified DHFR was added to a total amount of 5-10  $\mu\text{g}$  of protein and incubated for 1 hour at 4°C. Beads were washed three times in buffer and protein eluted by the addition of glutathione elution buffer (50 mM Tris-HCl pH 7.9, 10 mM reduced glutathione). Mix was incubated for 30 minutes at room temperature, 15  $\mu\text{L}$  supernatant collected and mixed with 15  $\mu\text{L}$  2x SDS-sample buffer. For analysis 20  $\mu\text{L}$  of sample were loaded on a SDS-PAGE and either stained by coomassie or protein was transferred to PVDF membrane (Immobilion-P, Millipore) for Western blot analysis.



### 5.3 List of Primers:

BM01:CCAGGGGGATTTGGATGTGGGCGGTTACNNKNNKNNKNNKNNKNN  
KNNKGGACTGGCCGTGAAGGAATG

BM02:GTAACCGCCACATCCAAATCCC

BM03:ACCAATTGCCTCCTCTAA

BM04:TGCACGATGCACAGTTGAAG

BM05:GAGCAGGGTCTGCACGAAATANNKNNKNNKNNKNNKNNKNN  
KGGCAAGGGGACGTCTGCAGCTGAT

BM06:TATTTTCGTGCAGACCCTGCTC

BM07:CGCGGATCCGACAAGGATTGTGAAATGAAACGC

BM08:CGCGGATCCTTATGGCTTCCCCAACCGGTGGCC

BM09:GACGAGCTCATGATCAGTCTGATTGCGG

BM10:GAGTTGGTCGACTTACCGCCGCTCCAGAATC

BM11:AGTAGAGCTCATGGCGACCTGCATCGG

BM12:ATCTGTCGACTTACTATCAATGAAAATTAGGAGTCG

BM13:ATCTGTCGACTTACTACATAAAAGGATGGTCCAATAC

BM14:CGGCGTCGACTTACTATTGAGACTGTCTATTAGAAGTGCC

BM15:GCGGAATTCTATGGCTTCCCCAACCGGTG

BM16:GCGGAATTCCGATGCCACCCCTCTAGATC

BM17:TCAGGATCCTATGATCAGTCTGATTGCGGC

BM18:TCAGGATCCTTACCGCCGCTCCAGAATCT

BM19:GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGACAAGGATTGTGA  
AATGAA

BM20:GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATCATCCCAAGCCT  
GGCTTC



## **6 Acknowledgements**

Looking back on my time spent at EPFL I want to take the opportunity and thank all the people who made it an enjoyable place to work.

First, I would like to thank my supervisor Prof. Kai Johnsson for having given me the opportunity to work in his lab. It was a very good experience that I would not want to miss. I always appreciated the freedom you gave me in conducting my projects and to always have your support when needed.

I would also like to thank the members of my thesis jury and president for having taken the time to examine my thesis.

I am grateful to Prof. U. Röthlisberger and E. Brunk for the good collaboration we had on the SNAP-structure project and for your openness to interdisciplinary questions. Furthermore, I would like to thank Liz for her support and friendship and of course for our nice winter runs we had together.

I want to thank all my colleagues at the Laboratory of Protein Engineering for the nice atmosphere, the cakes, the chocolate, the Lithuanian mushrooms that I will painfully miss and much more.

Thank you Grazvydas, Hiro, Christopher and most recently Cindy for sharing the lab with me over all these years. Even though our lab was maybe not the most talkative one, I always had your attention when needed and you were always very helpful and supportive in finding solutions to my questions. I really enjoyed working with you.

I would like to thank the former and new members of “the big lab” and the Chemistry lab. Simone, for being a good friend, such a nice running buddy and for looking at the thesis manuscript and PowerPoint slides. Miriam, for bringing some northern support and for your help in taking care of the students. Alberto for being such a relaxed

colleague and the great risotto you make. Thank you Rudolf for your good mood and the iPhone support when needed. Thanks Anastasiya for being the living proof that women run faster than (certain) men and swim in the ice-cold sea. Monica, thank you for your help with the “-80°C” organization that you know much better now than I do. I want to thank Luc for the many compounds, especially the “good TMR cross-linker” and for your help in reading the thesis manuscript. Thank you Claudia for your help and friendship and for being the person of the lab having the clearest statements of all of us. Thank you Karo for all your friendship, your help and the numerous coffees and pizza evenings.

Finally, I would like to thank my parents for always being supportive and reassuring. Your help allowed me to focus on the things I liked, to enjoy my studies and to become the person I am today.

Thank you Oma for all the times you kept your fingers crossed to wish me luck and the many hiccups I had in return.

A big thank you goes to Josy and Christian for their help and support and for cheering us up when it was needed.

And of course I want to thank you, Mathieu and there are no words big enough to explain how happy I am to have you at my side. Thank you for all your help and support not only throughout the thesis preparation. This all means a lot to me!

## 7 References:

1. Eijsink, V.G.H., Gåseidnes, S., Borchert, T.V. & van den Burg, B. Directed evolution of enzyme stability. in *Biomol Eng* Vol. 22 21-30 (2005).
2. Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. in *Annu Rev Biophys* Vol. 37 153-73 (2008).
3. Johannes, T.W. & Zhao, H. Directed evolution of enzymes and biosynthetic pathways. in *Curr Opin Microbiol* Vol. 9 261-7 (2006).
4. Bloom, J.D. & Arnold, F.H. In the light of directed evolution: pathways of adaptive protein evolution. in *Proc Natl Acad Sci USA* Vol. 106 Suppl 1 9995-10000 (2009).
5. Dougherty, M.J. & Arnold, F.H. Directed evolution: new parts and optimized function. in *Curr Opin Biotechnol* Vol. 20 486-91 (2009).
6. Shafikhani, S., Siegel, R.A., Ferrari, E. & Schellenberger, V. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. in *BioTechniques* Vol. 23 304-10 (1997).
7. Zaccolo, M., Williams, D.M., Brown, D.M. & Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. in *J Mol Biol* Vol. 255 589-603 (1996).
8. Christians, F.C., Scapozza, L., Crameri, A., Folkers, G. & Stemmer, W.P. Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. in *Nat Biotechnol* Vol. 17 259-64 (1999).
9. Coco, W.M. et al. DNA shuffling method for generating highly recombined genes and evolved enzymes. in *Nat Biotechnol* Vol. 19 354-9 (2001).

10. Cramer, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P. Improved green fluorescent protein by molecular evolution using DNA shuffling. in *Nat Biotechnol* Vol. 14 315-9 (1996).
11. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. in *Nature* Vol. 370 389-91 (1994).
12. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. in *Proc Natl Acad Sci USA* Vol. 91 10747-51 (1994).
13. Zhao, H. & Arnold, F.H. Optimization of DNA shuffling for high fidelity recombination. in *Nucleic Acids Res* Vol. 25 1307-8 (1997).
14. Black, M.E. & Loeb, L.A. Random sequence mutagenesis for the generation of active enzymes. in *Methods Mol Biol* Vol. 57 335-49 (1996).
15. Davidson, J.F., Anderson, J., Guo, H., Landis, D. & Loeb, L.S. Directed molecular evolution of proteins. in *Wiley-VCH Verlag GmbH* (2002).
16. Miyazaki, K. & Arnold, F.H. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. in *J Mol Evol* Vol. 49 716-20 (1999).
17. Giepmans, B.N.G., Adams, S.R., Ellisman, M.H. & Tsien, R.Y. The fluorescent toolbox for assessing protein location and function. in *Science* Vol. 312 217-24 (2006).
18. Subach, F.V. et al. Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. in *Nat Chem Biol* Vol. 5 118-26 (2009).
19. Terskikh, A. et al. "Fluorescent timer": protein that changes color with time. in *Science* Vol. 290 1585-8 (2000).
20. Fuchs, J. et al. A photoactivatable marker protein for pulse-chase imaging with superresolution. in *Nat Methods* Vol. 7 627-30 (2010).
21. Patterson, G.H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. in *Science* Vol. 297 1873-7 (2002).
22. Hinner, M.J. & Johnsson, K. How to obtain labeled proteins and what to do with them. in *Curr Opin Biotechnol* Vol. 21 766-76 (2010).

23. O'Hare, H.M., Johnsson, K. & Gautier, A. Chemical probes shed light on protein function. in *Curr Opin Struct Biol* Vol. 17 488-94 (2007).
24. Gaietta, G. et al. Multicolor and electron microscopic imaging of connexin trafficking. in *Science* Vol. 296 503-7 (2002).
25. Griffin, B.A., Adams, S.R. & Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. in *Science* Vol. 281 269-72 (1998).
26. Gronemeyer, T., Godin, G. & Johnsson, K. Adding value to fusion proteins through covalent labelling. in *Curr Opin Biotechnol* Vol. 16 453-8 (2005).
27. Chen, I., Howarth, M., Lin, W. & Ting, A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. in *Nat Methods* Vol. 2 99-104 (2005).
28. Yin, J., Liu, F., Li, X. & Walsh, C.T. Labeling proteins with small molecules by site-specific posttranslational modification. in *J Am Chem Soc* Vol. 126 7754-5 (2004).
29. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. in *Nat Biotechnol* Vol. 21 86-9 (2003).
30. Gautier, A. et al. An engineered protein tag for multiprotein labeling in living cells. in *Chem Biol* Vol. 15 128-36 (2008).
31. Los, G.V. et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. in *ACS Chem Biol* Vol. 3 373-82 (2008).
32. Keppler, A. et al. Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. in *Methods* Vol. 32 437-44 (2004).
33. Keppler, A., Pick, H., Arrivoli, C., Vogel, H. & Johnsson, K. Labeling of fusion proteins with synthetic fluorophores in live cells. in *Proc Natl Acad Sci USA* Vol. 101 9955-9 (2004).
34. Juillerat, A. et al. Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. in *Chem Biol* Vol. 10 313-7 (2003).

35. Juillerat, A. et al. Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. in *Chembiochem* Vol. 6 1263-9 (2005).
36. Gerson, S.L. MGMT: its role in cancer aetiology and cancer therapeutics. in *Nat Rev Cancer* Vol. 4 296-307 (2004).
37. Pegg, A.E. et al. Mechanism of inactivation of human O6-alkylguanine-DNA alkyltransferase by O6-benzylguanine. in *Biochemistry* Vol. 32 11998-2006 (1993).
38. Srivenugopal, K.S., Yuan, X.H., Friedman, H.S. & Ali-Osman, F. Ubiquitination-dependent proteolysis of O6-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O6-benzylguanine or 1,3-bis(2-chloroethyl)-1-nitrosourea. in *Biochemistry* Vol. 35 1328-34 (1996).
39. Rasimas, J.J., Dalessio, P.A., Ropson, I.J., Pegg, A.E. & Fried, M.G. Active-site alkylation destabilizes human O6-alkylguanine DNA alkyltransferase. in *Protein Sci* Vol. 13 301-5 (2004).
40. Kanugula, S., Goodtzova, K. & Pegg, A.E. Probing of conformational changes in human O6-alkylguanine-DNA alkyl transferase protein in its alkylated and DNA-bound states by limited proteolysis. in *Biochem J* Vol. 329 ( Pt 3) 545-50 (1998).
41. Oh, H.K. et al. Conformational change in human DNA repair enzyme O6-methylguanine-DNA methyltransferase upon alkylation of its active site by SN1 (indirect-acting) and SN2 (direct-acting) alkylating agents: breaking a "salt-link". in *Biochemistry* Vol. 35 12259-66 (1996).
42. Cherry, J.R. et al. Directed evolution of a fungal peroxidase. in *Nat Biotechnol* Vol. 17 379-84 (1999).
43. Ness, J.E. et al. DNA shuffling of subgenomic sequences of subtilisin. in *Nat Biotechnol* Vol. 17 893-6 (1999).
44. Zhao, H. & Arnold, F.H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. in *Protein Eng* Vol. 12 47-53 (1999).
45. Schmid, F.-X. Lessons about protein stability from in vitro selections. in *Chembiochem* Vol. 12 1501-7 (2011).



46. Martin, A., Schmid, F.X. & Sieber, V. Proside: a phage-based method for selecting thermostable proteins. in *Methods Mol Biol* Vol. 230 57-70 (2003).
47. Kristensen, P. & Winter, G. Proteolytic selection for protein folding using filamentous bacteriophages. in *Fold Des* Vol. 3 321-8 (1998).
48. Heinis, C., Alessi, P. & Neri, D. Engineering a thermostable human prolyl endopeptidase for antibody-directed enzyme prodrug therapy. in *Biochemistry* Vol. 43 6293-303 (2004).
49. Pedersen, J.S., Otzen, D.E. & Kristensen, P. Directed evolution of barnase stability using proteolytic selection. in *J Mol Biol* Vol. 323 115-23 (2002).
50. Tamakoshi, M., Nakano, Y., Kakizawa, S., Yamagishi, A. & Oshima, T. Selection of stabilized 3-isopropylmalate dehydrogenase of *Saccharomyces cerevisiae* using the host-vector system of an extreme thermophile, *Thermus thermophilus*. in *Extremophiles* Vol. 5 17-22 (2001).
51. Eijssink, V.G.H. et al. Rational engineering of enzyme stability. in *J Biotechnol* Vol. 113 105-20 (2004).
52. Arnold, F.H. How proteins adapt: lessons from directed evolution. in *Cold Spring Harb Symp Quant Biol* Vol. 74 41-6 (2009).
53. Jaenicke, R. & Böhm, G. The stability of proteins in extreme environments. in *Curr Opin Struct Biol* Vol. 8 738-48 (1998).
54. D'Amico, S., Marx, J.-C., Gerday, C. & Feller, G. Activity-stability relationships in extremophilic enzymes. in *J Biol Chem* Vol. 278 7891-6 (2003).
55. Wolf-Watz, M. et al. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. in *Nat Struct Mol Biol* Vol. 11 945-9 (2004).
56. Miyazaki, K., Wintrode, P.L., Grayling, R.A., Rubingh, D.N. & Arnold, F.H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. in *J Mol Biol* Vol. 297 1015-26 (2000).
57. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. in *Nature* Vol. 450 983-90 (2007).
58. Alber, F. et al. The molecular architecture of the nuclear pore complex. in *Nature* Vol. 450 695-701 (2007).

59. Nooren, I.M.A. & Thornton, J.M. Diversity of protein-protein interactions. in *EMBO J* Vol. 22 3486-92 (2003).
60. Kiel, C., Beltrao, P. & Serrano, L. Analyzing protein interaction networks using structural information. in *Annu Rev Biochem* Vol. 77 415-41 (2008).
61. Giot, L. et al. A protein interaction map of *Drosophila melanogaster*. in *Science* Vol. 302 1727-36 (2003).
62. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. in *Proc Natl Acad Sci USA* Vol. 98 4569-74 (2001).
63. Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. in *Science* Vol. 303 540-3 (2004).
64. Rain, J.C. et al. The protein-protein interaction map of *Helicobacter pylori*. in *Nature* Vol. 409 211-5 (2001).
65. Stelzl, U. et al. A human protein-protein interaction network: a resource for annotating the proteome. in *Cell* Vol. 122 957-68 (2005).
66. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. in *Nature* Vol. 403 623-7 (2000).
67. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. in *Curr Opin Chem Biol* Vol. 13 245-55 (2009).
68. Skerra, A. Engineered protein scaffolds for molecular recognition. in *J Mol Recognit* Vol. 13 167-87 (2000).
69. Steinmeyer, D.E. & McCormick, E.L. The art of antibody process development. in *Drug Discov Today* Vol. 13 613-8 (2008).
70. Skerra, A. Imitating the humoral immune response. in *Curr Opin Chem Biol* Vol. 7 683-93 (2003).
71. Raju, T.N. The Nobel chronicles. 1984: Niels Kai Jerne, (1911-94); César Milstein (b 1926); and Georges Jean Franz Köhler (1946-95). in *Lancet* Vol. 355 75 (2000).
72. Springer, T.A. César Milstein, the father of modern immunology. in *Nat Immunol* Vol. 3 501-3 (2002).

73. Kelley, B. Industrialization of mAb production technology: the bioprocessing industry at a crossroads. in *MAbs* Vol. 1 443-52 (2009).
74. Scolnik, P.A. mAbs: a business perspective. in *MAbs* Vol. 1 179-84 (2009).
75. Bandejas, T.M. et al. Structure of wild-type Plk-1 kinase domain in complex with a selective DARPIn. in *Acta Crystallogr D Biol Crystallogr* Vol. 64 339-53 (2008).
76. Stumpp, M.T., Binz, H.K. & Amstutz, P. DARPins: a new generation of protein therapeutics. in *Drug Discov Today* Vol. 13 695-701 (2008).
77. Theurillat, J.-P. et al. Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer. in *Mod Pathol* Vol. 23 1289-97 (2010).
78. Amstutz, P. et al. Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. in *J Biol Chem* Vol. 280 24715-22 (2005).
79. Zahnd, C., Pecorari, F., Straumann, N., Wyler, E. & Plückthun, A. Selection and characterization of Her2 binding-designed ankyrin repeat proteins. in *J Biol Chem* Vol. 281 35167-75 (2006).
80. Koide, A. & Koide, S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. in *Methods Mol Biol* Vol. 352 95-109 (2007).
81. Nixon, A.E. & Wood, C.R. Engineered protein inhibitors of proteases. in *Curr Opin Drug Discov Devel* Vol. 9 261-8 (2006).
82. Skerra, A. Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. in *FEBS J* Vol. 275 2677-83 (2008).
83. Schönfeld, D. et al. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. in *Proc Natl Acad Sci USA* Vol. 106 8198-203 (2009).
84. Weiss, G.A. & Lowman, H.B. Anticalins versus antibodies: made-to-order binding proteins for small molecules. in *Chem Biol* Vol. 7 R177-84 (2000).

85. Gilbreth, R.N., Esaki, K., Koide, A., Sidhu, S.S. & Koide, S. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. in *J Mol Biol* Vol. 381 407-18 (2008).
86. Henderson, K.A. et al. Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. in *Structure* Vol. 15 1452-66 (2007).
87. Grönwall, C. & Ståhl, S. Engineered affinity proteins--generation and applications. in *J Biotechnol* Vol. 140 254-69 (2009).
88. Barbas, C.F., Kang, A.S., Lerner, R.A. & Benkovic, S.J. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. in *Proc Natl Acad Sci USA* Vol. 88 7978-82 (1991).
89. Kretzschmar, T. & von Rüden, T. Antibody discovery: phage display. in *Curr Opin Biotechnol* Vol. 13 598-602 (2002).
90. Barbas, C.F., Burton, D.R., Scott, J.K. & Silverman, G.J. Phage Display: A Laboratory Manual. in Cold Spring Harbour Laboratory Press.
91. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. in *Science* Vol. 228 1315-7 (1985).
92. Barbas, C.F. et al. Recombinant human Fab fragments neutralize human type 1 immunodeficiency virus in vitro. in *Proc Natl Acad Sci USA* Vol. 89 9339-43 (1992).
93. Barbas, C.F. et al. In vitro evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity. in *Proc Natl Acad Sci USA* Vol. 91 3809-13 (1994).
94. Burton, D.R. et al. A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. in *Proc Natl Acad Sci USA* Vol. 88 10134-7 (1991).
95. Burton, D.R. et al. Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. in *Science* Vol. 266 1024-7 (1994).

- 
96. McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. in *Nature* Vol. 348 552-4 (1990).
  97. Wang, H.W. et al. Engineering and functional evaluation of a single-chain antibody against HIV-1 external glycoprotein gp120. in *Clin Exp Immunol* Vol. 141 72-80 (2005).
  98. Yang, W.P. et al. CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. in *J Mol Biol* Vol. 254 392-403 (1995).
  99. Binz, H.K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. in *Nat Biotechnol* Vol. 23 1257-68 (2005).
  100. Nygren, P.-A. Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. in *FEBS J* Vol. 275 2668-76 (2008).
  101. Nygren, P.-A. & Skerra, A. Binding proteins from alternative scaffolds. in *J Immunol Methods* Vol. 290 3-28 (2004).
  102. Bratkovic, T. et al. Affinity selection to papain yields potent peptide inhibitors of cathepsins L, B, H, and K. in *Biochem Biophys Res Commun* Vol. 332 897-903 (2005).
  103. Lunder, M. et al. Comparison of bacterial and phage display peptide libraries in search of target-binding motif. in *Appl Biochem Biotechnol* Vol. 127 125-31 (2005).
  104. Lunder, M., Bratkovic, T., Kreft, S. & Strukelj, B. Peptide inhibitor of pancreatic lipase selected by phage display using different elution strategies. in *J Lipid Res* Vol. 46 1512-6 (2005).
  105. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. in *Nat Biotechnol* Vol. 15 553-7 (1997).
  106. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. in *Nat Protoc* Vol. 1 755-68 (2006).

107. Georgiou, G. et al. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. in *Nat Biotechnol* Vol. 15 29-34 (1997).
108. Boder, E.T. & Wittrup, K.D. Yeast surface display for directed evolution of protein expression, affinity, and stability. in *Meth Enzymol* Vol. 328 430-44 (2000).
109. Buonpane, R.A. et al. Neutralization of staphylococcal enterotoxin B by soluble, high-affinity receptor antagonists. in *Nat Med* Vol. 13 725-9 (2007).
110. Jin, M. et al. Directed evolution to probe protein allostery and integrin I domains of 200,000-fold higher affinity. in *Proc Natl Acad Sci USA* Vol. 103 5758-63 (2006).
111. Daugherty, P.S. Protein engineering with bacterial display. in *Curr Opin Struct Biol* Vol. 17 474-80 (2007).
112. Löfblom, J., Sandberg, J., Wernérus, H. & Ståhl, S. Evaluation of staphylococcal cell surface display and flow cytometry for postselectional characterization of affinity proteins in combinatorial protein engineering applications. in *Appl Environ Microbiol* Vol. 73 6714-21 (2007).
113. Löfblom, J., Wernérus, H. & Ståhl, S. Fine affinity discrimination by normalized fluorescence activated cell sorting in staphylococcal surface display. in *FEMS Microbiol Lett* Vol. 248 189-98 (2005).
114. Harvey, B.R. et al. Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. in *Proc Natl Acad Sci USA* Vol. 101 9193-8 (2004).
115. Harvey, B.R. et al. Engineering of recombinant antibody fragments to methamphetamine by anchored periplasmic expression. in *J Immunol Methods* Vol. 308 43-52 (2006).
116. Kronqvist, N., Löfblom, J., Jonsson, A., Wernérus, H. & Ståhl, S. A novel affinity protein selection system based on staphylococcal cell surface display and flow cytometry. in *Protein Eng Des Sel* Vol. 21 247-55 (2008).
117. Rockberg, J., Löfblom, J., Hjelm, B., Uhlén, M. & Ståhl, S. Epitope mapping of antibodies using bacterial surface display. in *Nat Methods* Vol. 5 1039-45 (2008).

118. Rockberg, J., Schwenk, J.M. & Uhlén, M. Discovery of epitopes for targeting the human epidermal growth factor receptor 2 (HER2) with antibodies. in *Mol Oncol* Vol. 3 238-47 (2009).
119. Hanes, J. & Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. in *Proc Natl Acad Sci USA* Vol. 94 4937-42 (1997).
120. Mattheakis, L.C., Bhatt, R.R. & Dower, W.J. An in vitro polysome display system for identifying ligands from very large peptide libraries. in *Proc Natl Acad Sci USA* Vol. 91 9022-6 (1994).
121. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. in *Nat Methods* Vol. 4 269-79 (2007).
122. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. in *FEBS Lett* Vol. 414 405-8 (1997).
123. Roberts, R.W. Totally in vitro protein selection using mRNA-protein fusions and ribosome display. in *Curr Opin Chem Biol* Vol. 3 268-73 (1999).
124. Gold, L. mRNA display: diversity matters during in vitro selection. in *Proc Natl Acad Sci USA* Vol. 98 4825-6 (2001).
125. Chien, C.T., Bartel, P.L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. in *Proc Natl Acad Sci USA* Vol. 88 9578-82 (1991).
126. Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. in *Cell* Vol. 43 729-36 (1985).
127. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. in *Nature* Vol. 340 245-6 (1989).
128. Rual, J.-F. et al. Towards a proteome-scale map of the human protein-protein interaction network. in *Nature* Vol. 437 1173-8 (2005).
129. Koegl, M. & Uetz, P. Improving yeast two-hybrid screening systems. in *Brief Funct Genomic Proteomic* Vol. 6 302-12 (2007).

130. Brasch, M.A., Hartley, J.L. & Vidal, M. ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. in *Genome Res* Vol. 14 2001-9 (2004).
131. Brizuela, L., Richardson, A., Marsischky, G. & Labaer, J. The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. in *Arch Med Res* Vol. 33 318-24 (2002).
132. Temple, G. et al. From genome to proteome: developing expression clone resources for the human genome. in *Hum Mol Genet* Vol. 15 Spec No 1 R31-43 (2006).
133. Hudson, J.R. et al. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. in *Genome Res* Vol. 7 1169-73 (1997).
134. Lamesch, P. et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. in *Genomics* Vol. 89 307-15 (2007).
135. Parrish, J.R. et al. High-throughput cloning of *Campylobacter jejuni* ORfs by in vivo recombination in *Escherichia coli*. in *J Proteome Res* Vol. 3 582-6 (2004).
136. Lin, H., Abida, W., Sauer, R. & Cornish, V.W. Dexamethasone-methotrexate: an efficient chemical inducer of protein dimerization in vivo. in *Journal of the American ...* (2000).
137. Hussey, S.L., Muddana, S.S. & Peterson, B.R. Synthesis of a beta-estradiol-biotin chimera that potently heterodimerizes estrogen receptor and streptavidin proteins in a yeast three-hybrid system. in *J Am Chem Soc* Vol. 125 3692-3 (2003).
138. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. in *Proc Natl Acad Sci USA* Vol. 91 10340-4 (1994).
139. Dirnberger, D., Unsin, G., Schlenker, S. & Reichel, C. A small-molecule-protein interaction system with split-ubiquitin as sensor. in *Chembiochem* Vol. 7 936-42 (2006).
140. Eyckerman, S. et al. Design and application of a cytokine-receptor-based interaction trap. in *Nat Cell Biol* Vol. 3 1114-9 (2001).



141. Caligiuri, M. et al. MASPIT: three-hybrid trap for quantitative proteome fingerprinting of small molecule-protein interactions in mammalian cells. in *Chem Biol* Vol. 13 711-22 (2006).
142. Gronemeyer, T., Chidley, C., Juillerat, A., Heinis, C. & Johnsson, K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. in *Protein Eng Des Sel* Vol. 19 309-16 (2006).
143. Daniels, D.S. et al. Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. in *EMBO J* Vol. 19 1719-30 (2000).
144. Bojkowska, K. et al. Measuring in vivo protein half-life. in *Chem Biol* Vol. 18 805-15 (2011).
145. Pantoliano, M.W. et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. in *J Biomol Screen* Vol. 6 429-40 (2001).
146. Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. in *Anal Biochem* Vol. 357 289-98 (2006).
147. Pegg, A.E. et al. Use of antibodies to human O6-alkylguanine-DNA alkyltransferase to study the content of this protein in cells treated with O6-benzylguanine or N-methyl-N'-nitro-N-nitrosoguanidine. in *Carcinogenesis* Vol. 12 1679-83 (1991).
148. Javadpour, M.M., Eilers, M., Groesbeek, M. & Smith, S.O. Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. in *Biophys J* Vol. 77 1609-18 (1999).
149. Pace, C.N. & Scholtz, J.M. A helix propensity scale based on experimental studies of peptides and proteins. in *Biophys J* Vol. 75 422-7 (1998).
150. Loktionova, N.A. & Pegg, A.E. Interaction of mammalian O(6)-alkylguanine-DNA alkyltransferases with O(6)-benzylguanine. in *Biochem Pharmacol* Vol. 63 1431-42 (2002).
151. Daniels, D.S. et al. DNA binding and nucleotide flipping by the human DNA repair protein AGT. in *Nat Struct Mol Biol* Vol. 11 714-20 (2004).

152. Chidley, C., Haruki, H., Pedersen, M.G., Muller, E. & Johnsson, K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. in *Nat Chem Biol* Vol. 7 375-83 (2011).
153. Chidley, C. A Novel Yeast-Based Platform for Small Molecule Target Deconvolution Reveals Previously Unknown Binding Partners for the Approved Drugs Erlotinib, Atorvastatin and Sulfasalazine. *EPFL thesis Nr 4917* (2010).
154. Heinis, C., Schmitt, S., Kindermann, M., Godin, G. & Johnsson, K. Evolving the substrate specificity of O6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. in *ACS Chem Biol* Vol. 1 575-84 (2006).
155. Mayer, R.J., Chen, J.T., Taira, K., Fierke, C.A. & Benkovic, S.J. Importance of a hydrophobic residue in binding and catalysis by dihydrofolate reductase. in *Proc Natl Acad Sci USA* Vol. 83 7718-20 (1986).
156. Murphy, D.J. & Benkovic, S.J. Hydrophobic interactions via mutants of *Escherichia coli* dihydrofolate reductase: separation of binding and catalysis. in *Biochemistry* Vol. 28 3025-31 (1989).
157. Schmit, T.L. & Ahmad, N. Regulation of mitosis via mitotic kinases: new opportunities for cancer management. in *Mol Cancer Ther* Vol. 6 1920-31 (2007).
158. Leung, G.C., Ho, C.S., Blasutig, I.M., Murphy, J.M. & Sicheri, F. Determination of the Plk4/Sak consensus phosphorylation motif using peptide spots arrays. *FEBS Lett* 581, 77-83 (2007).
159. Bettencourt-Dias, M. et al. SAK/PLK4 is required for centriole duplication and flagella development. in *Curr Biol* Vol. 15 2199-207 (2005).
160. Habedanck, R., Stierhof, Y.-D., Wilkinson, C.J. & Nigg, E.A. The Polo kinase Plk4 functions in centriole duplication. in *Nat Cell Biol* Vol. 7 1140-6 (2005).
161. Hudson, J.W. et al. Late mitotic failure in mice lacking Sak, a polo-like kinase. in *Curr Biol* Vol. 11 441-6 (2001).
162. Leung, G.C. et al. The Sak polo-box comprises a structural domain sufficient for mitotic subcellular localization. in *Nat Struct Biol* Vol. 9 719-24 (2002).

- 
163. Fode, C., Motro, B., Yousefi, S., Heffernan, M. & Dennis, J.W. Sak, a murine protein-serine/threonine kinase that is related to the *Drosophila* polo kinase and involved in cell proliferation. in *Proc Natl Acad Sci USA* Vol. 91 6388-92 (1994).
164. Rechsteiner, M. & Rogers, S.W. PEST sequences and regulation by proteolysis. in *Trends Biochem Sci* Vol. 21 267-71 (1996).
165. Guderian, G., Westendorf, J., Uldschmid, A. & Nigg, E.A. Plk4 trans-autophosphorylation regulates centriole number by controlling betaTrCP-mediated degradation. in *J Cell Sci* Vol. 123 2163-9 (2010).
166. Johnson, E.F., Stewart, K.D., Woods, K.W., Giranda, V.L. & Luo, Y. Pharmacological and functional comparison of the polo-like kinase family: insight into inhibitor and substrate specificity. in *Biochemistry* Vol. 46 9551-63 (2007).
167. Sillibourne, J.E. et al. Autophosphorylation of polo-like kinase 4 and its role in centriole duplication. in *Mol Biol Cell* Vol. 21 547-61 (2010).
168. Taverna, D.M. & Goldstein, R.A. Why are proteins so robust to site mutations? in *J Mol Biol* Vol. 315 479-84 (2002).
169. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D.S. How protein stability and new functions trade off. in *PLoS Comput Biol* Vol. 4 e1000002 (2008).
170. Romero, P.A. & Arnold, F.H. Exploring protein fitness landscapes by directed evolution. in *Nat Rev Mol Cell Biol* Vol. 10 866-76 (2009).
171. Xu-Welliver, M. & Pegg, A.E. Degradation of the alkylated form of the DNA repair protein, O(6)-alkylguanine-DNA alkyltransferase. in *Carcinogenesis* Vol. 23 823-30 (2002).
172. Tubbs, J.L., Pegg, A.E. & Tainer, J.A. DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. in *DNA Repair (Amst)* Vol. 6 1100-15 (2007).
173. Reetz, M.T., Carballeira, J.D. & Vogel, A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. in *Angew Chem Int Ed Engl* Vol. 45 7745-51 (2006).

174. Jay, D.G. & Sakurai, T. Chromophore-assisted laser inactivation (CALI) to elucidate cellular mechanisms of cancer. in *Biochim Biophys Acta* Vol. 1424 M39-48 (1999).
175. Takemoto, K. et al. Chromophore-assisted light inactivation of HaloTag fusion proteins labeled with eosin in living cells. in *ACS Chem Biol* Vol. 6 401-6 (2011).
176. Nalbant, P., Hodgson, L., Kraynov, V., Touthkine, A. & Hahn, K.M. Activation of endogenous Cdc42 visualized in living cells. in *Science* Vol. 305 1615-9 (2004).
177. Ausubel, F.M. *Short protocols in molecular biology: a compendium of methods from Current protocols in molecular biology*, 2 v. (various pagings) (Wiley, New York 2002).
178. Gietz, R.D. & Woods, R.A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. in *Meth Enzymol* Vol. 350 87-96 (2002).

**Name :** Birgit Mollwitz  
**Date of birth:** August 30<sup>th</sup> 1981  
**Nationality:** German  
**Languages:** German (mother tongue), English and French (C1, written and spoken)

## Research Experience

---

Dec. 2007 – Current	<b>Graduate Research Assistant</b> , Laboratory of Protein Engineering, Swiss Federal Institute of Technology (EPFL) Lausanne, Switzerland. Head: Prof. Kai Johnsson. <i>Directed evolution of proteins, study of structure-function relationship, yeast two-hybrid technology.</i>
April – Oct. 2007	<b>Undergraduate Research Assistant</b> , Department of Biochemistry and Molecular Cell Biology, University Medical Centre Hamburg, Germany. Head: Prof. Ulricke Beisiegel. <i>Implementation of new imaging tool for the visualization and quantification of lipoprotein metabolism.</i>
April – Aug. 2006	<b>SOKRATES exchange student</b> , University of Neuchâtel, Switzerland. Head: Prof. Georg Suess-Fink. <i>Investigation on the role of ruthenium nanoparticles in benzene hydrogenation.</i>
July – Oct. 2005	<b>Exchange student</b> , University of Cardiff, School of Pharmacy, Wales, UK. Head: Prof. Chris McGuigan. <i>Synthesis of novel antiviral compounds.</i>
Feb. – May 2005	<b>Undergraduate Research Assistant</b> , European Molecular Biology Laboratory (EMBL), Hamburg, Germany. Head: Dr. Matthias Wilmanns. <i>Assembly, expression, purification and crystallization of constructs of human SH3-protein domains.</i>
April – Jun. 2004	<b>Intern</b> , Eppendorf AG, R&D Unit, Hamburg, Germany. <i>Testing and evaluation of different methods and materials to improve performance of medical analysis PCR-device.</i>

## Education

---

Dec. 2007 – Current	<b>Ph.D. candidate in Biotechnology and Bioengineering</b> , Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.
November 2007	<b>Diplom Chemikerin</b> , grade: very good, University of Hamburg, Germany.

## Publications

---

- 4 B. Mollwitz, E. Brunk, S. Schmitt, F. Pojer, M. Bannwarth, U. Röthlisberger, K. Johnsson. Directed evolution of the suicide protein O<sup>6</sup>-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. In *Biochemistry* Epub ahead of print (2011).
- 4 L. Reymond, G. Lukinavicius, K. Umezawa, D. Mauel, M. Brun, A. Masharina, K. Bojowska, B. Mollwitz, A. Schena, R. Griss and K. Johnsson. Visualizing Biochemical Activities in Living Cells through Chemistry. in *CHIMIA* vol. 65 (11), 868-71 (2011).
- 3 O. T. Bruns, H. Ittrich, K. Peldschus, M. G. Kaul, U. I. Tromsdorf, J. Lauterwasser, M. S. Nikolic, B. Mollwitz, M. Merkel, N. C. Bigall, S. Sapra, R. Reimer, H. Hohenberg, H. Weller, A. Eychmüller, G. Adam, U. Beisiegel and J. Heeren. Real-time magnetic resonance imaging and quantification of lipoprotein metabolism *in vivo* using nanocrystals. in *Nat. Nanotechnol.* vol 4 (3) 193-201 (2009).
- 2 U. I. Tromsdorf, N. C. Bigall, M. G. Kaul, O. T. Bruns, M. S. Nikolic, B. Mollwitz, R. A. Sperling, R. Reimer, H. Hohenberg, W. J. Parak, S. Förster, U. Beisiegel, G. Adam and H. Weller. Size and surface effects on the MRI relaxivity of manganese ferrite nanoparticle contrast agents. in *Nano Lett.* vol 7 (8) 2422-7 (2007).
- 1 G. Suess-Fink, B. Mollwitz, B. Therrien, M. Dadras, G. Laurenczy, A. Meister and G. Meister, Ruthenium nanoparticles intercalated in hectorite: A reusable hydrogenation catalyst for benzene and toluene. in *J. Cluster Sci.* vol 18 (1) 87-95 (2007).

## Congress Poster Presentations

---

Swiss Chemical Society Fall Meeting, 2011, Lausanne, Switzerland.

EMBL Chemical Biology Conference, 2010, Heidelberg, Germany.

Rigi Workshop of the Swiss Academy of Sciences, 2009, Rigi-Kulm, Switzerland.

German Society for Biochemistry and Molecular Biology Fall Meeting, 2007, Hamburg, Germany.

Mosbacher Kolloquium of German Society for Biochemistry and Molecular Biology, 2006, Mosbach, Germany.