8-2016

Applications of the GST- Affinity Tag in the Purification and Characterization of Proteins

Wibke Beatrice Kachel
University of Arkansas, Fayetteville

Applications of the GST- Affinity Tag in the Purification and Characterization of Proteins

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemistry

By

Wibke Beatrice Kachel
University of Regensburg
Bachelor of Science in Biochemistry, 2010
University of Arkansas
Bachelor of Science in Chemistry, 2011

August 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Thallapuranam Krishnaswamy Suresh Kumar, Ph.D.
Committee Chair

_____
Roger Koeppe II, Ph.D.
Committee Member

_____
Dan Davis, Ph.D.
Committee Member

_____
Bill Durham, Ph.D.
Committee Member

_____
David McNabb, Ph.D.
Committee Member

Abstract

With the latest innovations in biological sciences, large quantities of biologically active polypeptides as well as high throughput screening methods to quickly evaluate if these biomolecules potentially have therapeutic, diagnostic, or industrial purposes are required. The synthesis and purification of peptides and small proteins continue to be demanding as the production of high yields through chemical synthesis can involve large costs. On the other hand, there are only few examples of acquiring those biomolecules through cloning and expression in bacterial systems in form of recombinant fusion proteins. Glutathione S-Transferase (GST) is not only a very commonly used affinity tag to increase expression yields, but is also known to enhance the solubility of the protein of interest making it a valuable tool in the pursuit of purifying recombinant proteins. Moreover, multidimensional NMR spectroscopy is a widespread technique to reveal the 3D solution structure of proteins. Yet, obtaining structural information of peptides and small proteins can be difficult.

In this context, we have developed a rapid purification of peptides and small proteins by fusing them to GST. The method developed is advantageous over the other reported methods due to its easy one-step purification yielding large amounts of fusion protein. Subsequently, the fusion protein is cleaved enzymatically under mild conditions, and the cleavage products are separated using an efficient heat treatment process. Our results show, the peptide and small protein conformations are not disturbed by the heat treatment. Therefore, our method can be a valuable alternative for the production of various clinically significant small proteins and peptides.

Furthermore, we have optimized a method, which allows collecting structural information on protein/ peptide(s) of interest by employing the GST-tagged target protein during the

acquisition of NMR data. Our results demonstrate that the affinity tag GST does not affect the quality of NMR data of its fused partner but that the loss of signals in the ^{1}H-^{15}N HSQC spectrum corresponding to the affinity tag is due to the decrease in the T2 relaxation rate upon dimerization as well as the flexibility within the fusion protein caused by the linker located between GST and the target protein.

Acknowledgements

I would like to express my very great appreciation to my advisor, Dr. Kumar, for his patient guidance, and valuable and constructive suggestions during the planning and development of this research work that helped me grow as a researcher. I would also like to thank my committee for their assistance in keeping my progress on schedule and their advice during my committee meetings, and outside of lab.

My special thanks are extended to the Kumar lab group. I would like to thank Srinivas Jayanthi for all the help, useful critics, enthusiastic encouragement, and friendship. Special thanks should also be given to Jacqueline Morris, Rory Henderson, and my other lab mates for all the help along the way.

Assistance in the Pichia work provided by Dr. Pinto was greatly appreciated.

I would like to acknowledge and thank my parents, especially my mom, for their support and love that gave me the strength to get through grad school.

I want to thank my friends that I have met during grad school. I would like to express my deep gratitude to Erik Guzman, who became family to me, for looking out for me, being the very finest roommate, and inspiring me to be the best I can be. Many special thanks are extended to Kati Street for her friendship, spiritual support, and positive attitude that encouraged me throughout grad school. I cherish all the early mornings we got up to practice yoga. I would like to offer my special thanks to Matthias Knust for his friendship, brilliant analytical thinking that inspired lots of discussions, and help.

Finally, I am particularly grateful for all the love, never-ending support and encouragement, and advice along the way given by George Sakhel.

Table of Contents

Abbreviations

GST, Glutathione S-Transferase; **His-tag,** polyhistidine tag; **MBP-tag**, Maltose-binding protein-tag; **NMR**, nuclear magnetic resonance spectroscopy; **ELISA**, enzyme-linked immunosorbent assay; **G site**, specific binding site for GSH and analogues on GST; **H site**, binding site for hydrophobic substrates of GST; **SjGST26**, GST derived from *Schistosoma japonicum*, 26kDa; **GSH**, reduced glutathione; **IPTG**, Isopropyl β-D-1-thiogalactopyranoside; **PMSF**, phenylmethysulfonyl fluoride; **CNBr**, cyanogen bromide; **DTT**, dithiothreitol; **AEBSF**, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride; **Tm**, melting temperature; **CD2**, chromo-domain 2 of chloroplast signal recognition particle 43; **CD3**, chromo-domain 3 of chloroplast signal recognition particle 43; **CD2CD3**, chromo-domain 2 and 3 of chloroplast signal recognition particle 43; **NBT/BCIP,** nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate; **HSQC,** heteronuclear single quantum coherence spectroscopy; **MALDI-TOF,** matrix-assisted desorption/ionization time-of-flight; **Rg**, radius of gyration; **q**, scattering angle in $Å^{-1}$; **I(q)**, scattering intensity (SAXS); **GnRH**, Gonadotropin- releasing hormone; **KS**, Kallmann Syndrome; **FGF,** fibroblast growth factor; **FGFR,** fibroblast growth factor receptor; **SP**, signal peptide; **CR**, Cysteine-rich domain; **WAP**, whey acidic like-protein domain; **FnIII.1-4**, fibronectin type III domains 1-4; **HR**, histidine-rich domain; **ECM**, extracellular matrix; **SDS**, sodium dodecylsulfate; **YPDS-plates**, yeast extract peptone dextrose medium with sorbitol; **BMGY**, buffered glycerol-complex medium; **BMMY**, buffered methanol-complex medium; **LB**, Luria-Bertani; **PBS**, phosphate buffered saline; **Rd**, Rubredoxin; **AOX**, alcohol oxidase

1. Introduction

1.1. Protein purification

The study of proteins and their function is crucial to the understanding of both cells and organisms. These biomolecules are essential for many cellular processes, i.e. they can act as catalysts, structural elements, and are involved in signaling cascades, among many other responsibilities. Therefore, protein purification plays a vital part in the determination and characterization of the target molecule's structure, function, and interaction mechanism. Results are used for industrial or pharmaceutical applications, such as the generation of antibodies that are capable of determining the location of the protein in vivo which can give significant support to interesting hypotheses and disprove incorrect theories. Decades of biochemical research have equipped investigators with a variety of ways to isolate a protein from a complex mixture with the objective of obtaining pure protein in its native conformation. The initial material, which is derived from tissue or cell cultures, can be separated into fractions by taking advantage of the different physical or biochemical properties of the protein of interest, for example by centrifugation in terms of size, by precipitation with salt, or binding to ionic or affinity columns. These methods assist in the removal of contaminating material as well as in the enrichment of the fraction with the protein of interest. Most of the time, affinity chromatography is the preferred technique because it can simplify the purification process due to its high specificity to the target molecule. Overall, the goal of any purification is to maximize enrichment while minimizing loss of activity. For that reason, it is much easier if a rich source is identified. Modern cloning technologies have made an avenue for artificial sources of proteins. They are referred to as

recombinant proteins and gives about 10% or more of the total protein in the extract, which represents a tremendous experimental advantage.

1.2. Affinity and solubility tags

With the advances in biological sciences, there is a prevalent demand for large quantities of biologically active polypeptides[1]. Previously, in 2004 to 2010, the market for drugs derived from proteins boosted considerably and it is expected for it to continue to grow[2]. Since a large variety of proteins could potentially have therapeutic, diagnostic, or industrial purposes, researchers in the fields of proteomics, genomics, and bioinformatics are in the need to assess those prospective candidates quickly and efficiently[3]. In this context, recombinant proteins have been the preferred way of production.

Advantages and disadvantages of affinity and solubility tags

The biggest advantage of using affinity tags, which "can be defined as an exogenous amino acid sequence with a high affinity for a specific biological or chemical ligand"[2], in combination with the recombinant target protein is the ability to purify essentially any protein without having any prior knowledge of its biochemical properties[2]. In addition, the introduction of an affinity tag can have a positive affect on difficult-to-express protein or peptide of interest. It has been proven that in the presence of the carrier protein, the yield increased because the tag protected the protein of interest from degradation. In case of Rajan et al, the active N-terminal domain of the mouse tissue inhibitor of metalloproteinases-1 did only show stable expression when a polyhistidine -tag was present at the N-terminus[4]. However, even though one affinity tag works well for one protein that does not mean it also gives good yields for another. Sun et al were rather unsuccessful expression their target protein, human vasostatin 120-180, fused to

GST. Nevertheless, with the N-terminal His-tag, the yield of pure protein could be improved by 3-fold[5].

Initially established to enable detection and purification, affinity tags also show other advantages. As insolubility is a major bottleneck for high throughput applications it was discovered that the Thioredoxin-tag (Trx), the Maltose-binding protein-tag (MBP-tag) or the N-utilizing substance A protein-tag (NusA) influence the solubility of some to be overexpressed polypeptides, mammalian proteins, and green fluorescent protein respectively [6,7,8]. When performing an overexpression, molecular crowding is possible and can be unfavorable for the goal of correctly folded target proteins, as it might result in the formation of so called inclusion bodies. To be considered a solubility-enhancing tag, the protein obviously has to be very soluble itself and it is thought that in turn it is able to extend that property to the fusion partner. Nevertheless, this characteristic is not the only factor that promotes the solubility of the target protein and it is still unclear how exactly solubility-enhancing proteins work. Theories such as the tag being a "chaperone-magnet" or even function as a chaperone itself have been speculated [9,10]. Furthermore, even if the fusion protein is mainly found in inclusion bodies, it has been shown that the presence of a His-tag assisted in the refolding after the target protein had been purified under denaturing conditions on NTA-resin. Due to the now physical separation of the target protein on the column, the refolding procedure was successfully performed[11]. Another advantage of purifying the protein of interest under denaturing conditions can be the decreased accessibility of the protein to proteolytic degradation. In case of the urokinase-type plasminogen activator, Tang et al were able to yield 25% more active pure His-tagged recombinant protein[12].

An additional advantage of fusion proteins is their use for detection, as it was shown that for example GST can easily be recognized by an enzyme assay and therefore was able to

increase the sensitivity of binding assays[13,14]. Another useful application of fusion proteins is the increased stability for structural analysis. Even though large affinity tags were thought to be disadvantageous in the attempt to form crystals of the fusion protein because of the experience of multi-domain proteins being less likely to form well-ordered and diffracting crystals as the protease cleavage site between the tag and the protein of interest acts as a flexible linker, multiple structures of MBP-fused proteins have been solved[15]. Nevertheless, in order to avoid conformational heterogeneity, which impedes crystal growth, it has been advised to use a rigid rather than a flexible linker between tag and the target protein[16].

On the other hand, introducing an affinity tag has also been reported to result in negative effects on the target protein. In theory, any tag can influence the native structure, fold and/or activity of the protein that it is attached to. In case of the trimeric cytokine tumor necrosis factor alpha (TNF), its cytotoxicity on the L-929 cell line was decreased when the N-terminal His-tag was present. As soon as the affinity tag was removed, TNF was fully active again[17]. It should be mentioned that the already flexible N-terminus of TNF can causes a steric obstruction and the additional amino acids due to the His-tag increase said hindrance resulting in the dramatic loss of biological activity. Moreover, in 2005 Chant et al showed that the His-tag caused a conformational change of the gene regulatory protein AreA. As their urea denaturation and binding studies showed, the His-tagged protein underwent a conformational change decreasing its capacity to bind DNA[18]. Fortunately, this effect is reversible when the affinity tag is removed. Likewise, the location of His-tag was proven to affect the binding properties of a tumor-associated single chain Fv construct. When located at the C-terminus of the protein of interest, the tag interfered with the binding site which resulted in a lower binding capability of the target protein[19]. Finally, extreme overexpression of the target protein due to the more stable fusion

protein can render to being toxic for the host or a so called “metabolic burden” has also been observed in recombinant bacteria due to selective culturing conditions[20].

Even though affinity tags decrease the amount of time and resources necessary for a purification protocol, as one does not have to come up with an individualized procedure and resources for each target protein, it has to be mentioned that each choice of tag and isolation method requires to be well thought through and optimized. In the decision-making, factors such as binding capacity and buffer systems play a vital role. For one, it is desired to keep the bed volume of the resin low and to reuse the matrix multiple times. Moreover, the buffers should ideally be applicable for a wide range of proteins and downstream characterization experiments. Nevertheless, the choice of an appropriate fusion partner depends on the protein of interest as well as its applications later on. Affinity tags provide purification templates, but each construct of fusion protein might still demand detailed adjustments in order to gain the highest possible quality and quantity of protein of interest. Already established protocols are to be used as a guide rather than a definitive procedure since every protein behaves differently.

Overview of expression hosts used for fusion proteins

In order to perform studies to characterize the function, stability and structure of the protein of interest, it must be folded correctly and soluble. There are many different expression hosts available for protein overexpression, among them prokaryotes such as *Escherichia coli* (*E.coli*), or eukaryotes, i.e. yeast, insect, and mammalian cell lines. Deciding which expression host along with which affinity or solubility tag is suitable can be quite overwhelming, especially when there is only little known about the gene and the expressed protein. Most of the time generating more than a handful of constructs in various expression hosts is very labor intensive

and time consuming with regards to the screening and therefore rather impractical. For this reason, a more intuitive approach is usually made in which *E.coli* is the preferred system to start out with unless existing literature already established the need for a more complex expression system due to the need for post-translational modifications of the target protein. Subsequently, the different affinity tag constructs are analyzed and expression parameters re-optimized in order to identify the optimal growing conditions.

E.coli leads as an expression host for the production of recombinant proteins due to its simple, efficient, and economical approach[21]. However, being a prokaryote it is missing any posttranslational modification pathways, such as glycosylation or phosphorylation that might be essential for the stability or function of the protein of interest. It has also been observed that proteins that are produced in very high yields in *E.coli* tend to aggregate[22]. Consequently, nearly 50% are found in inclusion bodies when expressed in E.coli[23,24]. In addition, large proteins can be challenging to produce due to cytotoxicity and metabolic burden[25].

Yeast, specifically *Pichia pastoris* and *Saccharomyces cerevisiae*, offer a potent alternative for the secretion of recombinant proteins[25]. This expression system has been shown to obtain large quantities of expressed material. According to Braun et al, *S.cerevisiae* is valuable because of its reasonable cost along with its production in a timely manner[26]. Yet, its post-translational modifications are similar but not the same as in mammalian cell lines. An example displays the degree of glycosylation, in which yeast utilizes high mannose compared to other eukaryotes[22].

Being able to express recombinant proteins with their correct fold, including their post-translational modifications, is a strong benefit of mammalian expression hosts compared to *E.coli*. Nevertheless, establishing a stable cell line via viral infection and chromosomal

incorporation can be quite time consuming, expensive, and hence requires a lot of effort[27]. A new method represents the transient system, which is dependent on DNA transfer into the mammalian cells and a production of up to 10mg/L have been reported. Its only drawback is that transfected cells cannot be propagated, so each batch of expression requires high purity plasmid DNA[27].

Another alternative are viral expression systems that have been developed using eukaryotic cells as hosts. In case of insect cells, overexpression of recombinant proteins is usually accomplished with the help of the baculovirus system, well-known for its high yields and its correctly executed post-translational modifications[22]. Another prominent examples is vaccinia virus, a member of the poxvirus family[28]. Janknecht et al used this expression system to successfully obtain His-tagged human serum response factor (SRF). The fusion protein was purified and shown to be biologically active which requires SRF being both glycosylated and phosphorylated[29]. The downside of using vaccinia virus is the low yield during expression when compared to *E.coli*.

Poly-histidine tag (His)

In about 60% of fusion proteins the preferred affinity tag is the poly-histidine-tag, especially when structure determinations via X-ray crystallography is planned to be performed[30]. It usually consists of 5-15 histidine residues located at the N-terminus of the protein of interest, but can be fused to the C-terminus as well. This construct is quite feasible for structural and activity studies, as it usually does not interfere with the 3D structure, fold or activity of the target protein due to its small size, simple and robust structure, and low immunogenicity[13]. There are several examples of proteins and peptides that were isolated with the help of the His-tag that are part of clinical studies[31]. Another advantage of the His-tag is its low toxicity towards the

expression host[32]. It is purified because of the high affinity of the imidazole side chain of histidine to transition metal ions (Ni^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+}), with nickel or cobalt most commonly used (immobilized metal-ion affinity chromatography, IMAC) under both native and denaturing conditions[33,34]. Specifically Ni (II)-nitrilatriacetic acid (Ni^{2+}-NTA) was established by Hochuli et al in 1987[35]. Janknecht et al summarizes that it is possible to remove the target protein by means of decreasing the pH, higher concentrations of chelating agent, or imidazole. The elution of the fusion protein is preferably achieved by applying a gradient of imidazole (20-500 mM) at physiological pH and ionic strength. Lowering the pH can denature the protein and chelating agents might lead to inactivity of the target protein when looking to isolate metal-containing proteins[29]. In order to perform structural characterization studies subsequent to the purification, imidazole has to be removed via dialysis. In their studies, Hefti et al mention that imidazole is not recommended to be present during NMR and X-ray crystallography experiments because it frequently leads to the aggregation of the protein[36]. Consequently, they prefer to perform an on-column cleavage to retrieve their protein of interest instead of using imidazole for elution. Nevertheless, this affinity tag might not be suitable if the host already contains many proteins that are rich in histidines as those biomolecules could be present as impurities in the purified sample. However, using additional elution gradients at lower imidazole concentrations in the beginning stages of the purification process aid in the removal of such contaminants.

Streptavidin binding tags (Strep-tag)

Another example for affinity tags are the Streptavidin binding tags, which have been successfully used in bacterial, plant, yeast, and mammalian expression hosts[37,38,39,40]. The original octapeptide WRHPQFGG was constructed according to its affinity to the streptavidin

core, which is the shortened version of the tetrameric bacterial protein[41]. Streptavidin itself is isolated from *Streptomyces avidinii* and is noteworthy because of its strong affinity and specificity to bind biotin that is unique for any other type of non-covalent interaction[42]. However, Barrette-Ng et al mention in their studies the decreased binding capacity of the Strep-Tag if the carboxy-group is "protected" by the fusion partner as it is critical for the salt-bridge that forms between the tag and streptavidin. Therefore, they recommend to use it only as a C-terminal affinity tag[43].

In order to improve the existing tag, Strep-tag II (WSHPQFEK) was designed which has an increased affinity for Strep-Tactin, a derivative of streptavidin with higher peptide binding capacity[13,32]. Strep-tag II is advantageous because of its higher endurance against cellular proteases[44]. Moreover, the streptavidin-binding peptide (SBP) was constructed. This 38-residue peptide binds even more strongly to streptavidin than Strep II and the original octapeptide[43]. Once bound to the matrix any form of Strep-Tag can be eluted at physiological conditions with a biotin analogue[40,45]. However, it is not suggested to use this affinity tag for purification methods under denaturing conditions[44]. The reason why researches choose this tag is the fact that it is small in size in comparison to the His-tag and therefore should not interfere with the structure, fold, stability, or biological activity. Additionally, Strep II does not stimulate protein aggregation[32]. More important though is the lack of metal ions in the purification process, which can be relevant in studies of metalloproteins or downstream applications such as NMR[2,45]. This tag is also a valuable tool if the target protein is used with the intention to form functional complexes and purify them in one step rather than a tandem affinity purification[40]. Finally, this affinity tag is used often for detection purposes and employed in Western Blots or ELISAs[45].

S-tag

The S-tag is the truncated version of the S-peptide and is comprised of 15 amino acids that specifically interact with the S-protein (residues 21-124)[13]. Both the N-terminal S-peptide and protein originate from pancreatic ribonucleaseA (RNaseA) which catalyzes the cleavage of RNA[32]. Subtilisin cleaves RNaseA between residue 20 and 21. The two pieces can be reorganized, resulting in ribonuclease S (RNaseS), which comprises the S-tag and S-protein, and is comparable to RNaseA's activity[46]. The residue composition of the S-tag contains both positively and negatively charged residues, as well as uncharged and polar amino acids resulting in an overall neutral charge and only little structure of the peptide. Studies have shown that the tag can be located on either termini of the protein of interest as well as within the target[47]. However, the binding of the two fragments is very reliant on pH, temperature, and ionic strength, so that its elution conditions are mostly too severe for the protein of interest (3 M NaSCN, 3 M $MgCl_2$, or 0.2 M citrate pH 2)[32]. If the protein of interest is needed under native conditions, it is suggested to perform proteolytic digestion of the fusion protein while it is still bound to the S-protein-matrix. Nevertheless, this tag is mostly used for detection purposes using either sensitive-homogeneous assays or Western Blot. It has been reported that already 20 fmol can be made visible in solution or on Western blots[47]. The commercially available colorimetric based assays is able to support fast screening of soluble S-tagged proteins even before purification[25]. It is especially useful for high throughput applications as one can just use the lysate for the assay.

Thioredoxin A tag (TrxA)

Thioredoxin, along with Glutathione S-transferase, the Maltose binding protein and NusA are regarded as tags that are able to aid solubility of the fusion protein[6,10,21]. Thioredoxin A is an 11.6 kDa *E.coli* oxido-reductase that is able to function as a reducing agent through the flexible oxidation of dithiol in its active center and thio-disulfide exchange reactions[48]. The thioredoxin system, which is comprised by TrxA and NADPH-thioredoxin reductase, is involved in many biochemical procedures, such as providing hydrogen for the ribonucleotide reductase, which in turn is necessary for the enzymatic synthesis of deoxyribonucleotides[49]. TrxA is thought to have evolved from a common ancestor and can be found in both prokaryotes and eukaryotes. However, TrxA from *E.coli* is the most studied and best characterized protein of them. Originally purified in 1964 it has been acknowledged for its high solubility and also displays high thermal stability, which has been shown to be transferred to the fusion proteins as well and consequently reduced the amount misfolded cytoplasmic aggregates[25,50]. Other theories propose that TrxA acts as a chaperone on the fusion partner, guiding it to its proper tertiary fold. Due to its structure in which both N- and C-terminus of TrxA are exposed, it can be attached to either amino- or carboxyl-terminus of the protein of interest[51]. In terms of purification methods one can either use an additional affinity tag in order to isolate the fusion protein or take advantage of TrxA's thermal stability by incubating it at 80 °C for 10 minutes[50]. Moreover, La Vallie et al described TrxA's unique feature of being secreted from the *E.coli* cytosol upon osmotic shock. Additionally, they inserted peptide sequences in the active loop region of TrxA and that way obtained high yields of these small biomolecules of interest[52]. Similar to other tags, TrxA needs to be removed prior to structural characterization studies, as it would interfere with the target protein's solution structure due to its size.

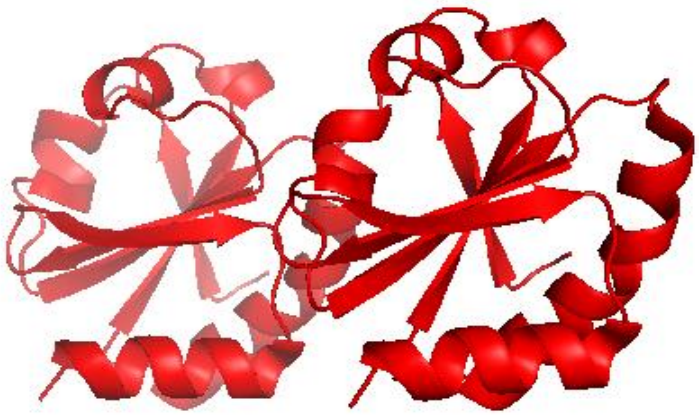


Figure 1 Pymol illustration of the Trx-tag (PDB: 2TRX)

Glutathione S-Transferase tag (GST)

Literature shows that up until the late eighties, researchers had to depend on purification methods under denaturing conditions in *E.coli*[1]. With the introduction of the pGEX vectors, scientists were now able to express and purify large quantities under mild conditions[21,53]. Fusion proteins, especially with Glutathione S-Transferase (GST) as the carrier protein, have been shown to express well in yeast and mammalian cell lines[54,55]. GST is considered to not only be an affinity tag, but also a solubility tag[2,10,21]. Due to its simple way of isolating its fusion proteins it is a commonly used carrier protein[53]. GST binds with high affinity to glutathione, which is coupled to a Sepharose matrix. The interaction is reversible and is eluted competitively with high concentrations of reduced glutathione[53]. Undoubtedly, the GST-tag was considered the most extensively used affinity tag. The drawbacks of this tag are the slow binding kinetics of the tag to immobilized glutathione in case of scaling up the purification process and consequently it results to be rather time consuming[25]. In addition, when utilizing baculovirus-mediated insect cell expression, Hunt et al observed that GST host proteins were present as impurities in the purified protein sample. In some other cases, the elution process that is performed under reducing

conditions might be problematic. Furthermore, it is known that GST is a homodimer[56,57], which could possibly also lead to oligomerization of the fusion protein. Nevertheless, when compared to the MBP-tag and His-tag, Dyson et al showed that for 32 different target proteins (17-110 kDa), GST-fusion proteins yielded the highest amount of soluble protein[7].

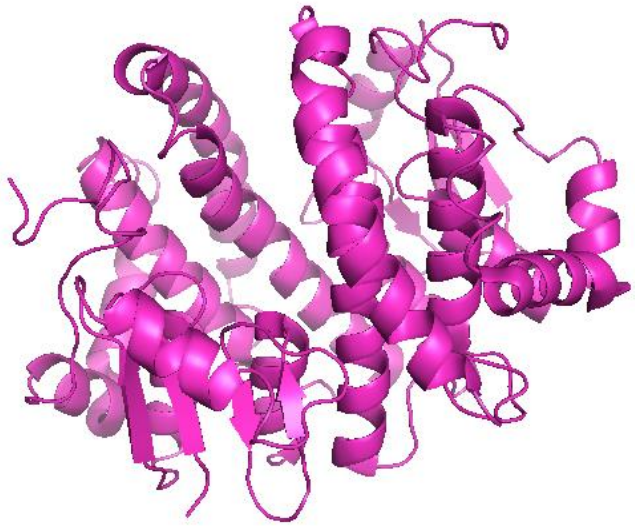


Figure 2 Pymol illustration of the GST-tag (PDB: 1Y6E)

Maltose binding protein tag (MBP)

The Maltose-binding protein (MBP) has a size of 42 kDa and is derived from the *malE* gene in *E.coli*[58]. It enables maltose to be transported across the cytoplasmic membrane[13]. It is purified by binding the tagged protein of interest to immobilized amylose and can readily be eluted under non-denaturing conditions at neutral pH using 10 mM maltose[44]. Nevertheless, it is more utilized due to its ability to improve solubility and folding[8] as it has been shown that its specificity and binding capacity are low[13]. Still, it has been successfully used for the expression of many eukaryotic proteins in *E.coli*[7,10]. Due to its size and immunogenicity, the MBP should be removed for further downstream characterization or clinical applications respectively[32]. One of

Kapust and Waugh's examples when comparing several tags for their ability to increase solubility of the target protein was TEV. When expressed as His-TEV, most of the protein was found not only inactive but also in the insoluble fraction. If produced as fusion protein with MBP-His-TEV with a TEV recognition site between the MBP- and His-tag, His-TEV seemed to be stabilized as it was found in the soluble fraction[10]. In an additional experiment, when MBP was co-expressed with His-TEV but not actually fused together, His-TEV was found in the insoluble fraction. This suggests that MBP was necessary to ensure proper folding of the target protein. There is no evidence of how MBP aids in the folding of its passenger protein. Kapust and Waugh propose a chaperone-like model in which MBP guides the not properly folded fusion partner towards its active shape through hydrophobic interactions[10]. Those hydrophobic interactions of MBP with its fusion partner are also favorable because they might inhibit aggregation of the not properly folded protein of interest[9]. However, this contact might also be the reason why certain fusion proteins do not bind as efficiently to the resin during purification[59].

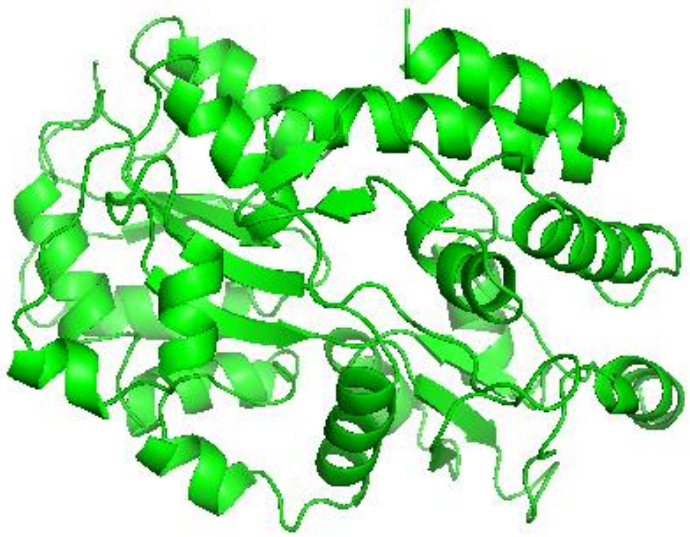


Figure 3 Pymol illustration of the MBP-tag (PDB: 1ANF)

N-utilization substance A protein tag (NusA)

N-utilization substance A is known to be one of the transcription termination factors and stimulates the RNA polymerase to take breaks from DNA transcription in *E.coli*[60]. Being considered a solubility tag, it has to be utilized in conjunction with an affinity tag. Even though there is only very little known about its ability to promote proper folding, this 55 kDa hydrophilic protein is a valuable tool in the expression of aggregation-prone proteins, as it is assumed that it reduces the translation speed allowing more time for the folding process to take place[25,61]. Another theory of how NusA facilitates higher yield of target protein is the assumption that expression levels are reliant on the stability of their mRNA[62]. Mah et al's hypothesis entails that NusA supports the RNA stem-loop and is also able to directly bind to the alpha subunit of the RNA polymerase. Still, the actual role that NusA is playing has not been discovered yet[63]. Additionally, Nallamsetty et al showed that NusA has the ability to enhance the solubility of the protein of interest by 30—50%[8]. According to Nallamsetty's and Waugh's studies, NusA and MBP displayed similar abilities to promote solubility or folding of the fusion partner and both carrier proteins are understood to more likely play a passive role in the folding of their fusion partner[8]. Consequently, the folding performance, which was estimated by the fusion proteins presence in the soluble fraction, is thought to depend on the passenger protein rather than the carrier protein. Nallamsetty's findings were confirmed by Marblestone et al's studies, in which three different proteins were fused to several affinity and/or solubility tags. Among them TrxA, GST, MBP, SUMO and NusA were utilized, resulting in a considerably increased detectable yield of protein when attached to SUMO or NusA[62].

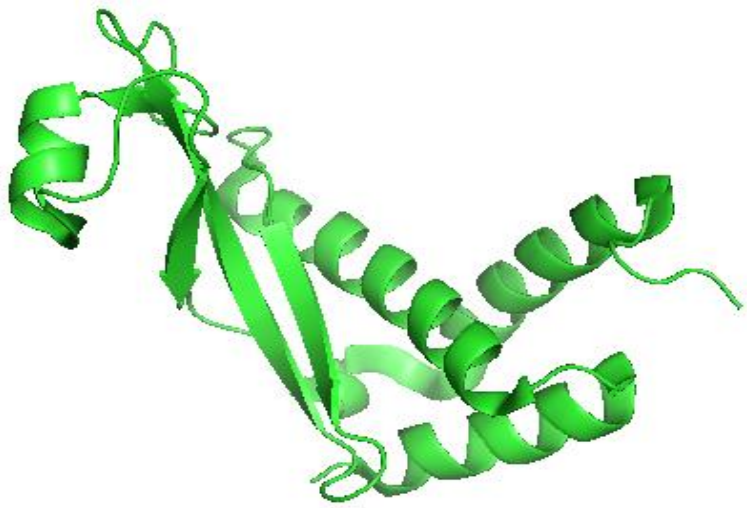


Figure 4 Pymol illustration of the amino-terminal domain of NusA (PDB 2KWP)

Small ubiquitin-related modifying protein tag (SUMO)

The small ubiquitin-like modifying protein (SUMO) is a 100 residue eukaryotic protein, derived from *Saccharomyces cerevisiae*. It aids in the post-translational modification that are important for many cellular processes, among them protein activation, protein stability, and the cell cycle[64,65,66]. The 11 kDa SUMO is added at the N-terminus of a target protein in order to increase expression levels for prokaryotic expressions as it is possible that its own resistance to proteases protects to protein of interest from degradation from the N-end[13]. Another interesting aspect of its ability to shield the target protein from degradation is by removing it from the protease rich cytosol to the nucleus. In case of Kishi et al's studies, pancreatic duodenal homeobox-1 (Pdx1) could only be localized in the nucleus when it was the sumoylated[67]. SUMO has also been shown to be helpful in promoting folding and therefore increased the stability of the fused protein of interest[62,68,69]. SUMO's structure comprises a hydrophobic core and a hydrophilic surface, which is very comparable to the 76 residue protein ubiquitin, which is known to be the fastest protein to fold[70] and to act in a similar fashion as a detergent on

otherwise insoluble target proteins[24]. Yet, an affinity tag in series to SUMO is necessary to purify the fusion protein. Literature also suggests utilizing wild type SUMO only in the bacterial expression system as *E.coli* is lacking highly conserved SUMO proteases that are only present and highly conserved in eukaryotes[2,13,44]. These proteases, such as yeast SUMO protease-1 Ulp1, recognize the conformation of the ubiquitin partner at a Gly-Gly motif rather than a specific amino acid sequence and are able to cleave under a wide range of conditions, such as temperature, pH, and ionic strength[24]. Additionally, Ulp-1 is favorable due to its low ratios of protease that are required for the cleavage (1:5,000 molar ratio), which might make this protease promising for large-scale expressions. Butt et al also report about the new generation of SUMO proteases that only require 1: 100,000 molar ratio of protease to protein of interest[24]. Besides, LifeSensors, Inc. has designed a solubility-tag based on SUMO (SUMOstar) and a protease accordingly that can be utilized in any eukaryotic expression system[44].

Figure 5 Pymol illustration of SUMO (PDB: 1A5R)

Table 1 Common affinity and solubility tags for recombinant proteins

Tag	Size (aa, kDa)	Comments
polyHis-tag	5-15, 0.7-2	Most commonly used affinity tag
Streptag II	8, 1	Does not stimulate protein aggregation
S-tag	15, 1.7	Mostly used for detection purposes
Small ubiquitin-like modifier (SUMO)	100, 11	Increases stability of the fusion protein
Thioreroxin (Trx)	109, 11.6	Purification methods via thermal stability or osmotic shock
Glutathion S-transferase (GST)	201, 26	Increases solubility and yield, yet slow binding kinetics[13]
Maltose binding protein (MBP)	396, 42	Enhances solubility[8]
N-utilization substance A (NusA)	495, 54	Enhances solubility[8]

1.3. Glutathione S-Transferase (GST)

Glutathione S-Transferase embodies an important contributor in the phase II detoxification of endogenous and xenobiotic alkylating agents, among them environmental toxins or therapeutic drugs. Armstrong et al has described GST as one of the most important enzymes in the elimination of harmful electrophilic compounds, which is found in animals, plants, and many microorganisms[71]. While mainly cytochrome P450 monooxygenases oxidize xenobiotics in phase I of the breakdown of foreign and toxic compounds, GST among other key players is responsible for the catalysis of the conjugation reaction of electrophilic harmful substances to the reduced cellular tripeptide glutathione[72,73,74]. In addition, the GST enzymes protect against hydroperoxides that are byproducts during chemical and oxidative stress[75].

When first studying this family of enzymes, it was uncertain from a biological point of view why GST forms a dimer to be fully active. Studies performed by Dirr and Reinemer demonstrate that being a dimer is beneficial for the thermostability, the fold, and overall tertiary structure of the protein as it was shown that the dissociation and unfolding reaction are carefully interconnected[76]. One would assume that oligomers would display a stable intermediate state in the unfolding process, but additional experiments done by Erhardt and Dirr confirm the absence of thermodynamically stable intermediates such as a folded monomer. In contrast their findings suggest a two-state transition from folded dimer to unfolded monomers[77]. Both hydrophobic and hydrophilic interactions have been found to stabilize the interface between the monomers. Each subunit in the protein dimer, meaning each GST, contains its own catalytic center and consists of two components. The N-terminal alpha/beta domain 1 is smaller and contains most of the residues that make up the G site, which is the specific binding site for GSH or analogues. Due to the specificity to GSH these residues in the binding site are highly conserved[78]. Alpha domain 2 is larger and contains the H site, which binds the hydrophobic substrate that can display a great structural variance[73]. It has been shown that in the presence of foreign compounds the expression of GST was increased considerably[79] suggesting that the more GST is present, the better the cell is prepared for a broad spectrum of toxic chemicals. Up until the mid nineties, already more than 100 chemicals, some of them both substrates as well as inducers, have been identified that stimulate GST expression[73]. Furthermore, the vast variety of substrates, all of them displaying structural differences, that GST is able to metabolize is impressive[80] which is probably the reason for the presence of numerous GST isozymes in most species[81]. In fact, in humans GST can make up 4% of the cytosolic proteins in the liver[78]. Nevertheless, species, strain, age, sex, and organ seem to impact the induction of the GST activity[73]. An important characteristic of

most GST isozymes is that they are only active when they form dimers[1]. Moreover, glutathione seems to be restricted to aerobic organisms, which is why GST is not anticipated to occur in any anaerobic organisms[80].

Nomenclature of GST

When first categorized, the different GSTs were sorted dependent on their substrate specificity and the molecular weight of the monomer[71], but due to the overlap in the usage of substrates among the isozymes this approach did not have a solid foundation and was neglected soon[74]. Another method to organize the different isozymes was based on the composition of subunits to form the functional dimer. Mannervik et al showed that GST is able to arrange as either homo- or heterodimers, leading to an Arabic numeral annotation[82]. Up until now, reports of heterodimers indicate though that they are comprised of subunits from the same class[78] implying that there are explicit structural requirements for the subunit interactions. Since there is not enough evidence to date that the isozymes of different mammalian species match, the nomenclature is performed according to the same principle but independently. In cases of human GSTs, Greek letters were originally chosen for the categorization and the Arabic numerals have not been implemented yet.

There are three major families of GSTs: cytosolic, mitochondrial and microsomal[75]. Cytosolic and membrane-bound GSTs are the most studied and found in all eukaryotic organisms and also in bacteria[32,83]. The cytosolic enzymes, which are found in higher organisms such as humans, do not share a common "ancestor gene" but are the product of five different gene families, hence designated class alpha, mu, pi, sigma and theta[73]. Studies in the species rat, mouse, and human by Mannervik et al showed that the isoenzymes of cytosolic GSTs within a

group have similar structural characteristic, as they all form dimers, as well as related catalytic features and amino acid sequence similarities. The major representatives of human GST are the class alpha, mu and pi and were described by Mannervik et al as the basic, neutral, and acidic type respectively[81].

Class alpha GST

In accordance with Mannervik's classification, various GSTs belonging to class alpha exhibit an isoelectric point at a pH larger than 7.8[84]. In humans, the genes coding for GSTs of this class are found on a cluster mapped to chromosome 6[85]. Additionally, studies showed that they are the most abundantly expressed glutathione S-transferases in the liver. In contrast to the other GST classes, alpha GSTs exhibit a blocked N-terminal amino group[86]. This acylated serine residue is a usual modification found in proteins. Furthermore, it was shown that alpha isozymes share 55% sequence identity[73] and exhibit glutathione peroxidase activity[81]. They process bilirubin and some anti-cancer drugs in the liver in order to defend the cells from reactive oxygen species and the products of peroxidation. Mutation studies on an alpha class GST performed by Board and Mannervik suggest that the C-terminus is responsible for the substrate specificity[87]. Once the residues located at the C-terminus were deleted or mutated, GST-2 lost most of its activity towards its substrate cumene hydroperoxide. In contrast to the majority of GSTs that are found in the cytoplasm, some mouse and human alpha GSTs have been discovered interacting with membranes and mitochondria[88].

Class mu GST

Class mu GSTs have been shown to share 65% sequence identity[73]. Furthermore, being classified as neutral GSTs their isoelectric point was confirmed at pH 6.6[84]. With a dimer size of 53,000 Da, mu GST proteins have a larger molecular weight than alpha or pi proteins (51,000 Da and 47,000 Da respectively). Interestingly, only 60% of tested adults, but no fetal tissue exhibit this class of GSTs[84]. This indicates that this group of proteins is stimulated later in life, maybe due to repeated contact to xenobiotics as class mu members have been found to interact mostly with epoxides[84]. Armstrong et al revealed that different residues in the H-site are responsible whether the enzyme is active predominantly regarding epoxides or halogenated benzenes[89].

As an example, GST that is used for the affinity tag, is a 26 kDa protein found in the parasitic worm *Schistosoma japonicum*[53] and belongs to the mammalian class mu based on sequence homology[90]. The crystal structure exposes the C-terminus as a relatively free structured domain at the surface of the dimerized protein[91]. It is known that the N-terminal domain binds to glutathione, which is the reason that the GST-tag is always at the N-terminus of the fusion protein: the N-terminus of GST is still able to bind to the resin while the C-terminus is connected to the protein of interest[91]. Furthermore, it has been reported that GST exists as a homodimer[21]. According to Kaplan's studies, whose results also show that GST is purified as a homodimer under non-reducing conditions, also demonstrates substantial amounts of 160 kDa and larger aggregates that are still catalytically active[56].

Class pi GST

In agreement with Mannervik's description of class pi GSTs being acidic, their isoelectric point was measured to be at pH 4.8[84] and is found in placenta and erythrocytes[92]. Another

characteristic of this category is that it displays high activity and specificity for ethacrynic acid[80]. Even though this substrate is not as hydrophobic as substrates from other classes, its recognition by pi GSTs is attributed to their slightly more open H binding site. Reinemeier et al call attention to the folding motif of domain 1 that is matching the pattern of thioredoxin in bacteriophage T4 and is also very similar to thioredoxin found in *E.coli*[93]. Nevertheless, it is still uncertain whether this means these two proteins are evolutionary related. A popular representative of class pi GSTs is GST P1-1, which is the most predominant isozyme in mammalian cells[94]. Studies showed that the majority of human tumors and tumor cell lines have substantial quantities of class pi GST present due to higher expression levels, which is the reason for the particular interest for this protein[73].

1.4. GST as an affinity tag: Sj26GST

Glutathione S-transferase, derived from *Schistosoma japonicum* (Sj26GST), belongs to the class mu GSTs due to its sequence homology and has found application as affinity tag in the late eighties[56,90]. In its original organism Sj26GST is important for the parasite's detoxification pathway as it aids in the "S-conjugation between the thiol group of glutathione and an electrophilic moiety of xenobiotic toxic compounds"[71]. This parasite has only very few enzymes to assist in the cleansing process, i.e. superoxide dismutase, cytochrome P450, and catalase, leaving GST as one of its key protection mechanisms against electrophilic and oxidative damage[95]. Once the toxic molecule is attached to glutathione, the conjugates are more soluble in water which leads to the removal from the cell in order to be excreted[56]. Smith et al discovered that Sj26GST is also involved in the removal of insoluble hematin, which would otherwise accumulate in the parasite's gut. Sj26GST binds to this reduced form of the heme prosthetic

group resulting in the secretion and therefore preventing the formation of large crystals[90]. In addition, they mentioned studies using antibodies directed against Sj26GST in order to inhibit its solubilizing function and therefore induce a lethal constipation of the worm rather than having the conjugate causing blockages in the host's liver[96].

Crystal structure of Sj26GST, Dimerization and Ligand Binding

The numbering of the residues are from the latest crystal structure of Sj26GST[97], PDB code: 1Y6E, and differ slightly with regards to the other references.

McTigue et al solved the crystal structure for Sj26GST in the absence of its substrate gluthatione in 1995, while Lim et al had already elucidated the three-dimensional structure for the complex in 1994[98,99]. They show that Sj26GST, consisting of 218 residues, is comparable to other members of the GST family and that there is no significant conformational change upon substrate binding. Each subunit of the homodimer consists of two domains. Domain 1 at the N-terminus includes the residues 1-84, with residue 77-84 representing the short linker between domain 1 and 2, and shows a folding topology of bababb. The beta sheets are mainly arranged in an antiparallel order. The residues comprising domain 2 at the C-terminus are 85- 218 and form 5 a-helices with a succeeding loop section (residues 195-218)[98,99]. Furthermore, unlike in the other classes, S26jGST exhibits a so-called mu loop located in domain 1 (residues 33-42)[91]. In order to be functionally active, the dimerization of this enzyme is essential. According to McTigue et al, the dimer dimensions for Sj26GST are 57Å × 47Å × 44Å and exhibits a two-fold rotation axis. It was revealed that the dimer interface for class mu GSTs is more hydrophobic than the one of the other categories but is also displaying a "lock-and-key" type interaction characteristic for alpha, mu, and pi GSTs[93,98]. Specifically for the Sj26GST, Phe51 of one

subunit is buried in a hydrophobic pocket formed by the residues 91-94 and 129-133 of the other GST[99]. Comparing results of McTigue, Lim, and Rufer, the residues in each GST involved in the dimer interface are 50-53, 63-70, 88-109, and 129-136. The key participants are the following: Phe51, Leu64, Ala69 of subunit 1 of one GST interact with Ala89, Met93, Leu94, and Phe132 from subunit 2 of the partner GST[98]. In addition to the hydrophobic interactions, Lim et al found that a hydrophilic channel in close proximity to the hydrophobic dimer interface stabilizes the dimer. There is more flexibility to the hydrophilic interactions compared to the hydrophobic ones, but crucial residues include salt bridges between Asp 76 – Arg88 and Glu50 - Arg135, as well as Gln66 - Arg72, and Ser92 - Asp100[98]. These amino acids as well as their corresponding residues in subunit 2 are found on a-helices. Overall, the association of two GSTs constructs a 40Å long and 6-10Å wide pocket with mainly polar residues, but also leucine and methionine are located in it[99].

In addition, Lim et al, among other groups, were able to identify the residues involved in the interaction with its substrate glutathione (GSH). Several interactions are necessary to have GSH attached to domain 1. First, the gamma-Glu of GSH needs to be aligned and stabilized. This is achieved through hydrogen bonds between the carboxyl group of Glu and Gln66 - Ser67 of GST as well as a salt bridge established by the N-atom of Glu and Asp100 of domain 2. Next, hydrogen bonds forming from the GST residues Asn53 and Leu54 help in the process to orient the peptide backbone of GSH properly. Then, the carboxyl group of Gly in GSH needs to be stabilized through a hydrogen bond between the carbonyl oxygen on GSH with the indol ring of Trp7 in GST. This specific residue is crucial for the active side of Sj26GST. Mutational studies showed that the substitution with Phe lead to a decreased binding efficiency of GSH by two thirds as well as less than 2% remaining enzymatic activity of GST[100]. Last, the interaction with

the sulfhydryl group of GSH needs to be established to ensure enzymatic activation. However, the exact mechanism of creating the thiol anion is still not revealed.

Next to the active site, which is also called G-site, there is also a nonsubstrate ligand-binding site located at domain 2, also known as H-site because the residues associated with this region are mainly hydrophobic. Due to the range of hydrophobic substrates that exhibit structural variances, different amino acids of GST are involved in the interactions. Key residues, however, are Ile9, Leu12, Ser106, Tyr110, Gln203, and Gly204.

pGEX vectors

The plasmin pSj5 has been shown to synthesize Sj26, controlled by the IPTG-inducible tac promoter. Various changes of the plasmid resulted in 3 commercially available plasmids that were introduced in 1988: pGEX-1, pGEX-2T, and pGEX-3X make the expression of polypeptides fused to GST in *E.coli* possible. Literature shows that up until the late eighties, researchers had to depend on purification methods under denaturing conditions in *E.coli*[1]. With the introduction of the pGEX vectors, scientists were now able to express and purify large quantities under mild conditions[53]. Its success is shown in its more that 1,000 citations within the first 5 years[1]. The vectors contain a DNA sequence that signals for the origin of replication. Furthermore, the tac promoter[101] is an important characteristic, followed by the nucleotide sequence coding for GST (Sj26). Instead of the termination codon for GST, one can find a polylinker including the restriction enzyme recognition sites of *BamHI, SmaI,* and *EcoRI*[53]. Finally, translation will be stopped due to the termination codon TGA. In case of pGEX-2T, the polylinker is comprised and codes for the cleavage recognition sequence for the protease thrombin, while in pGEX-3X it encodes for the recognition site of factor X_a. In order to ensure

the ability to grow under selective conditions, the vector contains the β-lactamase-coding gene Ap^R. Overexpression of the protein of interest needs to be able to be controlled. For this reason, a fragment of the lac operon is introduced. It is comprised of the *lacI*^{q} allele of the *lac* repressor as well as part of *lacZ.* The introduction of the pGEX vectors has been proven a very successful and valuable tool in the production of biological active proteins, mainly due to its mild conditions necessary during the purification of the fusion protein.

1.5. Versatility of the GST-tag

The GST-tag is a highly soluble protein and is found in the cytoplasm[53]. Due to this fact and because of its large size (26 kDa) it is anticipated that it extends its solubility to its fusion partner[21]. In case of the expression of antimicrobial peptides, in more than 25% GST and Thioredoxin are the preferred fusion partners[102]. It has been reported that proteins as big as 97 kDa have been expressed with GST as its fusion partner[1]. Frangioni and Neel adjusted the purification protocol to still be able to obtain pure samples of large GST fusion proteins[103]. They confirmed that the larger the protein of interest, the more difficult the fusion protein is expressed as well as its reduced efficiency to bind to the chromatography resin. In addition, the insolubility of some fusion proteins is influenced by the presence of extremely hydrophobic or charged residues[1]. Nevertheless, the fact that most proteins fused to GST can be isolated without using denaturants or detergents is advantageous for downstream applications and eliminates the need for lengthy purification protocols. Furthermore, the GST-tag is often referred to as solubility tag as it assists in protein folding[21]. Besides, it helps avoiding intracellular digestion if fused to the target protein and preserves the recombinant protein in the soluble fraction[98, 104].

Single-step Purification of GST –fused proteins

Glutathione- agarose beads are able to bind roughly 8 mg of fusion protein per 1ml of swollen resin[53]. GST fused proteins can easily be purified from the bacterial crude lysate in a timely fashion using a single step purification under non-denaturing conditions by absorption onto immobilized glutathione, followed by competitive elution via reduced glutathione[53,81]. Procedures such as the batch-binding mode or low-pressure columns that utilize either gravity flow or a peristalitic pump make this type of purification very feasible for the laboratory[21]. The yield ranges from 15-60 mg fusion protein per 1 liter bacterial culture[1,21,53]. However, if the fusion protein is toxic to the cell, yields could potentially be much lower. The purity of this purification technique has shown to be greater than 90%. The fact that it requires only one chromatography step shows its competitiveness in comparison with other affinity tags. The economical and practical benefits of a time and resource-saving purification procedure are significant aspects for consideration. In case there is a problem of more contaminating bacterial proteins, the addition of Triton X-100 during the absorption of the fusion protein to the glutathione- agarose resin has been shown to reduce such impurities[53]. Nevertheless, Triton X-100 might impair biological assays later on; therefore minimizing the cycles of sonication could be of more advantage. Furthermore, a low yield of purified fusion protein is most likely due to insolubility. There are many factors that can influence the solubility of the overexpressed protein of interest. Increased degradation of the target protein can be overcome by adding protease inhibitors, such as 1mM EDTA (ethylenediaminetetraacetic acid) or 1mM PMSF (phenylmethylsulfonyl fluoride) to the cell lysate. In addition, it is suggested to test several strains of bacterial host as the stability of the overexpressed protein can vary greatly[53]. Another tactic to improve the yields is to alter the growth conditions. In some cases changing the

concentration of the inducing agent IPTG (Isopropyl β-D-1-thiogalactopyranoside) as well as the growth temperature has been shown to impact the amount as well as the stability of the target protein[102]. Another advantage next to the well-established purification protocols is the ability of using this tag in various expression hosts. Even though *E.coli* is probably the most common host for recombinant proteins[105], yeast[54], insect[106], and mammalian[107] cell lines have also been used to express GST-fusion proteins. By expressing in eukaryotic expression systems, researchers are able to obtain post-translational modified target proteins even when produced as fusion proteins.

1.6. Usages of GST-fused proteins

GST-fusion proteins have found usage in various biological applications. Due to the high yields and simple purification method, this construct is often used for structure determinations of the protein of interest. Even though the tag needs to be removed for NMR studies due to its size, several crystal structures of fused proteins exist. Zhan et al showed that especially when trying to crystallize certain parts of a protein, such as the regulatory domain, it is very beneficial to use the GST-fusion protein as this domain is generally very challenging to form crystals individually[108]. The structure of GST has already been fully revealed, therefore making it easy to find the conformation of the target molecule through the phase information in a molecular replacement method. Another example for the success of this approach is the elucidation of the structure of small domains or peptides as described by Lim et al, in which they were able to crystallize a peptide fused to GST. It is usually difficult to grow crystals of peptides or specific parts of a protein but when fused to GST the researchers could acquire structural information[98]. When comparing the structure of GST in the fusion proteins with individually crystallized Sj26GST, Zhan et al confirmed that they were very similar. Moreover, they discovered that the linker

between GST and the fusion partner as well as the fusion partner are in an extended conformation. In case of pGEX-1, in which a protease cleavage site is missing, the fusion partner closed back towards GST. Nevertheless, the folding back did not seem to have an effect on the structure of the fused peptide[108]. Interestingly, Lally et al grew good crystals of their peptide of interest fused to GST but could not yield valuable diffraction data. Nevertheless, they were able to use electron microscopy to confirm the intact fusion protein. More intriguingly though, their results imply that the GST part of the fusion protein dimerizes while the attached peptides was extended and possibly quite flexible[109].

GST fusion proteins are also relevant in protein- protein interactions, which involve the detection of GST fusion proteins using an enzymatic assay or immunoassay. The GST pull-down assay is probably the most prominent technique for this purpose, in which the GST fusion protein is immobilized and resembles the "bait" of the protein-protein interaction[110]. In addition, GST-fusion proteins are valuable tools in the studies of DNA-protein interactions. In a similar way to the GST pull-down assay, the DNA-binding protein is expressed and purified as a GST fusion protein. Due to the tag, the fusion protein can be immobilized and the specific conditions for DNA-protein interactions, such as transcription factors can be studied[111]. Another way to apply GST-fusion proteins is their efficient use in the production of vaccines. Yip et al composed a fusion protein, comprised of GST and the ErbB-2 peptide epitope. Mice were injected multiple times with this construct and antibodies were assayed via ELISA[112]. In comparison to other carriers, GST provoked the strongest antibody response. Besides, this approach is quite beneficial due to its straightforward and economical means of production of the GST-fusion construct.

1.7. Removal of affinity tags

The removal of the affinity tag is often viewed as the weak link of this purification strategy. Most of the time small affinity tags, i.e. His-tag or Strep II do not need to be removed because of their small size. They are assumed to not interfere with the 3D solution structure, fold, and/or biological activity of the target protein. Nevertheless, the excision of the carrier tag can be achieved but requires the careful selection of a protease.

Chemical vs. enzymatic cleavage of fusion proteins

The fusion partners can be separated from their counter parts either via chemical treatment or an enzymatic method. Even though very effective, chemical treatments are usually negatively associated with their complexity and their expensive procedures[113]. When selecting chemical reagents for cleavage, one will most likely choose from CNBr, formic acid, or hydroxylamine. In most scenarios, CNBr is ill advised as it recognizes methionine, which might be present the sequence of the target protein. Furthermore, all of the once mentioned above are attributed with harsh conditions, such as dramatic pH changes, which usually are unfavorable in biological systems as they denature proteins or induce modifications of the side chains[2].

In contrast, the enzymatic cleavage using proteases such as thrombin, factor Xa, or TEV can be accomplished under mild conditions and are economically introduced through DNA technology. Moreover, the use of recombinant fusion proteases brings additional advantages. For one, the expression and purification in lab can be more economical and the removal of the fusion protease can be handled along with the removal of the cleaved tag[114]. Lastly, recombinant fusion proteases are valuable, especially because the purity of commercially available enzymes, such as thrombin, can be problematic at times[115]. However, those endoproteases are also associated with

drawbacks such as the demand for high ratios of enzymes with regard to the fusion protein, which can be quite expensive in case of up scaling the protein production[2]. Secondary cleavage sites in which the protease is active at locations other than the intended position also represent a serious disadvantage when using for instance thrombin[116,115]. In most cases this is due to a prolonged incubation time, which can also be seen as uneconomical. In some examples an inefficient and incomplete digestion can also be related to steric hindrance, in which the cleavage site is too close to a folded structure of the protein of interest[25]. In order to circumvent this issue the introduction of additional residues, for example five glycine residues, might be able to enhance the cleavage efficacy[44]. Besides, the considerable time it might take to cleave the fusion protein, the protein of interest might not be folded correctly anymore, functionally inactive, or even instable after cleavage and precipitate[117]. Some enzymes also leave residues at the N-terminus of the protein of interest. This might be unfavorable for target proteins with therapeutic applications, in which case an additional cleavage step using exogenous proteases might be necessary[2].

Thrombin

Up until today, thrombin is isolated from bovine plasma as there has yet to be described a suitable method of expressing and purifying recombinant thrombin. Young et al mention that the purification process can be difficult, which leads to contaminated thrombin preparations in some cases[44,115]. Nonetheless, this trypsin-like serine protease is considered cost effective compared to factor Xa and the PreScission protease. Being a heterodimer, it is interconnected through disulfide bonds. These three intramolecular disulfide bonds ensure the stability of the protein[118]. Its ideal conditions are described to be a pH range of 5-10 with an optimum at pH 9.5 in the

absence of NaCl and 8.3 in the presence of 1M NaCl[116]. Additionally, thrombin is resistant to several detergents and shows optimal activity at a temperature of 45 °C. It can be disabled by PMSF (phenylmethysulfonyl fluoride) or AEBSF (4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride). The cleavage site for thrombin that is used in fusion proteins is LVPR | GS, which is related to the natural cleavage site of thrombin in human factor VIII (LVPR | GF). Like any trypsin-like serine protease it breaks the peptide bond on the carboxyl side of the basic residue arginine[32,115]. Surprisingly, the first one is cleaved with a better efficiency and was modified due to the need for a *BamHI* restriction enzyme recognition site[53]. However, even though the thrombin cleavage is considerably specific, it is not absolute. Multiple studies showed that it mistakenly hydrolyzed peptide bonds after the residue lysine[119]. Jenny et al compared several cleavage experiments and there seems to be a trend of secondary cleavage sites with prolonged incubation times[115]. In case of using GST as the carrier and in case there is any uncleaved fusion protein still present, both can be removed in a similar fashion by affinity chromatography to glutathione- agarose resin[53] which is the reason for its popularity. In addition, it has been reported that thrombin can be separated via benzamidine sepharose[44]. Enteropeptidases and viral proteases are lacking an affinity tag in order to be removed subsequent to the digestion[116]. In addition, due to the issue of having inefficient cleavage experiences with thrombin, some researches revised their clones by inserting a GlyGlyGlyGlyGly motif near the cleavage recognition site[120].

Factor Xa

Similar to thrombin, factor Xa is a blood-clotting enzyme, is considered a trypsin-like serine protease, and cleaves at the peptide bond at the carboxyl side of the basic amino acid

arginine but in the specific arrangement of I(D/E)GR | X. This linker sequence originated from the sequences in prothrombin, because factor Xa alters prothrombin to thrombin[115]. Two disulfide-linked subunits, 17 and 16 kDa, make up the active protein. Hence, reducing agents in the cleavage buffer will decrease the efficiency of this enzyme. It is an advantage this glycoprotein can be expressed recombinantly and secreted from mammalian cells[121] as well as isolated from blood plasma[122]. Even though Factor Xa has a higher specificity than for example thrombin does, its drawbacks are the high ratios of enzyme to fusion protein to ensure a successful, effective cleavage and the associated high costs. Nevertheless, there are no additional residues left at the N-terminus of the protein of interest when using Factor Xa, which is especially crucial for recombinant proteins or peptides that are intended to be used in clinical studies. Furthermore, this calcium binding protein is only affected by a few detergents but still not as tolerant towards them compared to thrombin[123].

PreScission Protease

The PreScission protease is a 46 kDa protein that was genetically engineered and is only available at GE Healthcare. It is derived from human rhinovirus (HRV 3C) that is responsible for diseases such as polio and hepatitis A. The protease specifically recognizes the amino acid sequence LFQ | GP, cleaving between glutamine and glycine. The optimal cleavage buffer is 50 mM Tris-HCl, containing 150 mM NaCl, 1 mM EDTA and 1 mM DTT at pH 7 according to the vendor. Under these conditions one unit of protease can separate 90% of 0.1 mg fusion protein at 5 °C within 16 hours. Moreover, Zn^{2+} can be used to inhibit the enzyme. PreScission protease's advantages are the low operating temperature (5 °C) and it is constructed as a GST- fused protein, which makes it possible to remove the protease and for instance the GST-tag at the same

time. According to Hunt et al, this protease seems to be strategically better as it has minimal non-specific cleavage sites[25].

TEV Protease

Recently, viral proteases have become increasingly more popular. It has been shown that they exhibit a more strict sequence specificity[116]. The tobacco etch virus (TEV) protease is possibly the best-illustrated enzyme of this type as William Dougherty et al initially described in 1989. Its optimum recognition site is a linear epitope comprised of seven residues (ENLYFQ | G/S) and the separation occurs between glutamine and glycine/serine. Many attempts in producing large quantities of recombinant TEV protease result in the need of solubility-enhancing fusion partner. His-tagged TEV protease is the most common clone with a yield of up to 400 mg/l, but it was also constructed with a GST-, MBP- or Streptag II[124,10,125]. The self-digestion of the catalytic domain near the C-terminus proofs to be problematic as it dramatically decreases the protease efficiency[124]. However, autolysis can be sidestepped by creating mutants with substituting residues close to the internal cleavage site[126]. It is 100-fold more unwilling to undergo self-cleavage and still shows moderate catalytic activity compared to the wildtype[44]. This S219V mutant is commercially available at Invitrogen, the so-called Ac-TEV. The TEV protease is active at a pH ranging from 6-9 and is reported to be most active in the absence of a monovalent salt. Although its optimal operating temperature is at 30-34 °C it still preserves considerable efficiency at 4 °C[127]. The TEV protease cannot be inactivated by PMSF or AEBSF, but is disabled by 0.01% SDS and temperatures above 37 °C[116]. Compared to thrombin and factor Xa, this protease has yet to report an instance in which it cleaves a fusion protein other than at its designed cleavage site.

Separation of cleaved fusion proteins

There are several ways to separate the protein of interest from the affinity tag. An easy and effective method is "on-column cleavage" in which the tag is cleaved off while the fusion protein is still bound to the resin. The advantages are that one does not have to introduce another chromatography step to remove the cleaved tag. In addition, one does not have to worry if the cleaved tag will completely bind to the resin as it is already interacting with the matrix. Therefore, the tag will be removed in the same step. Also any uncleaved fusion protein, which would be considered a contaminant would stay attached to the column[1]. However, the amount of protease necessary to completely separate the target protein from the tag is slightly higher than in off-column cleavages making the on-column approach less efficient.

In case of an off-column cleavage the isolation of the protein of interest from the affinity tag can be accomplished by re-chromatography to eliminate the cleaved tag and any un-cleaved fusion protein. Other ways of tag removal include gel filtration or other chromatography steps depending on the affinity tag used and the protein of interest, for example for highly charged target proteins ion exchange chromatography can also be an alternative[102]. In cases of the target molecule being a peptide, which has initiated high demands due to their vital roles in various biological signaling processes, high performance liquid chromatography (HPLC) has been the preferred method of purification post cleavage. Yet, the disadvantage of introducing another purification step for the tag removal is the possible loss of product with each additional chromatography technique. In a competitive market where production costs are high, the development of a more economical, sound purification process would add substantial benefits and may be a "trend to follow".

1.8. References

1. Smith, D. B., Purification of glutathione S-transferase fusion proteins. *Methods Mol. Cell. Biol.* **1993,** *4* (5), 220-9.

2. Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J., Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression Purif.* **2006,** *48* (1), 1-13.

3. Korf, U.; Kohl, T.; van der Zandt, H.; Zahn, R.; Schleeger, S.; Ueberle, B.; Wandschneider, S.; Bechtel, S.; Schnoelzer, M.; Ottleben, H.; Wiemann, S.; Poustka, A., Large-scale protein expression for proteome research. *Proteomics* **2005,** *5* (14), 3571-3580.

4. Rajan, S. S.; Lackland, H.; Stein, S.; Denhardt, D. T., Presence of an N-terminal polyhistidine tag facilitates stable expression of an otherwise unstable N-terminal domain of mouse tissue inhibitor of metalloproteinase-1 in Escherichia coli. *Protein Expression Purif.* **1998,** *13* (1), 67-72.

5. Sun, Q.-M.; Chen, L.-L.; Cao, L.; Fang, L.; Chen, C.; Hua, Z.-C., An Improved Strategy for High-Level Production of Human Vasostatin 120-180. *Biotechnol. Prog.* **2005,** *21* (4), 1048-1052.

6. Chen, H.; Xu, Z.; Xu, N.; Cen, P., Efficient production of a soluble fusion protein containing human beta-defensin-2 in E. coli cell-free system. *J. Biotechnol.* **2005,** *115* (3), 307-315.

7. Dyson, M. R.; Shadbolt, S. P.; Vincent, K. J.; Perera, R. L.; McCafferty, J., Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **2004,** *4*, No pp given.

8. Nallamsetty, S.; Waugh, D. S., Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expression Purif.* **2006,** *45* (1), 175-182.

9. Fox, J. D.; Kapust, R. B.; Waugh, D. S., Single amino acid substitutions on the surface of Escherichia coli maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* **2001,** *10* (3), 622-630.

10. Kapust, R. B.; Waugh, D. S., Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **1999,** *8* (8), 1668-1674.

11. Kou, G.; Shi, S.; Wang, H.; Tan, M.; Xue, J.; Zhang, D.; Hou, S.; Qian, W.; Wang, S.; Dai, J.; Li, B.; Guo, Y., Preparation and characterization of recombinant protein ScFv(CD11c)-TRP2 for tumor therapy from inclusion bodies in Escherichia coli. *Protein Expression Purif.* **2007,** *52* (1), 131-138.

12. Tang, W.; Sun, Z.-Y.; Pannell, R.; Gurewich, V.; Liu, J.-N., An efficient system for production of recombinant urokinase-type plasminogen activator. *Protein Expression Purif.* **1997,** *11* (3), 279-283.

13. Graeslund, S.; Hammarstroem, M., Affinity fusions for protein purification. *Downstream Ind. Biotechnol.* **2013**, 191-199.

14. Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **2003,** *60* (5), 523-533.

15. Smyth, D. R.; Mrozkiewicz, M. K.; McGrath, W. J.; Listwan, P.; Kobe, B., Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **2003,** *12* (7), 1313-1322.

16. Center, R. J.; Kobe, B.; Wilson, K. A.; Teh, T.; Howlett, G. J.; Kemp, B. E.; Poumbourios, P., Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein Sci.* **1998,** *7* (7), 1612-1619.

17. Fonda, I.; Kenig, M.; Gaberc-Porekar, V.; Pristovsek, P.; Menart, V., Attachment of histidine tags to recombinant tumor necrosis factor-alpha drastically changes its properties. *TheScientificWorld* **2002,** *2*, 1312-1325.

18. Chant, A.; Kraemer-Pecore, C. M.; Watkin, R.; Kneale, G. G., Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expression Purif.* **2005,** *39* (2), 152-159.

19. Goel, A.; Colcher, D.; Koo, J.-S.; Booth, B. J. M.; Pavlinkova, G.; Batra, S. K., Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochim. Biophys. Acta, Gen. Subj.* **2000,** *1523* (1), 13-20.

20. Bentley, W. E.; Mirjalili, N.; Andersen, D. C.; Davis, R. H.; Kompala, D. S., Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng.* **2009,** *102* (5), 1284-1297.

21. Harper, S.; Speicher, D. W., Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2011,** *681* (Protein Chromatography), 259-280.

22. Braun, P.; LaBaer, J., High throughput protein production for functional proteomics. *Trends Biotechnol.* **2003,** *21* (9), 383-388.

23. Yee, A.; Chang, X.; Pineda-Lucena, A.; Wu, B.; Semesi, A.; Le, B.; Ramelot, T.; Lee, G. M.; Bhattacharyya, S.; Gutierrez, P.; Denisov, A.; Lee, C.-H.; Cort, J. R.; Kozlov, G.; Liao, J.; Finak, G.; Chen, L.; Wishart, D.; Lee, W.; McIntosh, L. P.; Gehring, K.; Kennedy, M. A.; Edwards, A. M.; Arrowsmith, C. H., An NMR approach to structural proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (4), 1825-1830.

24. Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R., SUMO fusion technology for difficult-to-express proteins. *Protein Expression Purif.* **2005,** *43* (1), 1-9.

25. Hunt, I., From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression Purif.* **2005,** *40* (1), 1-22.

26. Holz, C.; Hesse, O.; Bolotina, N.; Stahl, U.; Lang, C., A micro-scale process for high-throughput expression of cDNAs in the yeast Saccharomyces cerevisiae. *Protein Expression Purif.* **2002,** *25* (3), 372-378.

27. Wurm, F.; Bernard, A., Large-scale transient expression in mammalian cells for recombinant protein production. *Curr. Opin. Biotechnol.* **1999,** *10* (2), 156-159.

28. Mackett, M.; Smith, G. L.; Moss, B., General method for production and selection of infectious vaccinia virus recombinants expressing foreign genes. *J. Virol.* **1984,** *49* (3), 857-64.

29. Janknecht, R.; De Martynoff, G.; Lou, J.; Hipskind, R. A.; Nordheim, A.; Stunnenberg, H. G., Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. U. S. A.* **1991,** *88* (20), 8972-6.

30. Derewenda, Z. S., The use of recombinant methods and molecular engineering in protein crystallization. *Methods (San Diego, CA, U. S.)* **2004,** *34* (3), 354-363.

31. Winzerling, J. J.; Berna, P.; Porath, J., How to use immobilized metal ion affinity chromatography. *Methods (San Diego)* **1992,** *4* (1), 4-13.

32. Zhao, X.; Li, G.; Liang, S., Several affinity tags commonly used in chromatographic purification. *J. Anal. Methods Chem.* **2013**, 581093/1-581093/9, 9 pp.

33. Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G., Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **1975,** *258* (5536), 598-9.

34. Sulkowski, E., Purification of proteins by IMAC. *Trends Biotechnol.* **1985,** *3* (1), 1-7; Chaga, G.; Bochkariov, D. E.; Jokhadze, G. G.; Hopp, J.; Nelson, P., Natural poly-histidine affinity tag for purification of recombinant proteins on cobalt(II)-carboxymethylaspartate crosslinked agarose. *J. Chromatogr. A* **1999,** *864* (2), 247-256.

35. Hochuli, E.; Doebeli, H.; Schacher, A., New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine residues. *J. Chromatogr.* **1987,** *411*, 177-84.

36. Hefti, M. H.; Van Vugt-Van der Toorn, C. J. G.; Dixon, R.; Vervoort, J., A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. *Anal. Biochem.* **2001,** *295* (2), 180-185.

37. Skerra, A., Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. *Gene* **1994,** *151* (1/2), 131-5.

38. Witte, C.-P.; Noel, L. D.; Gielbert, J.; Parker, J. E.; Romeis, T., Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Mol. Biol.* **2004,** *55* (1), 135-147.

39. Prinz, B.; Schultchen, J.; Rydzewski, R.; Holz, C.; Boettner, M.; Stahl, U.; Lang, C., Establishing a versatile fermentation and purification procedure for human proteins expressed in the yeasts Saccharomyces cerevisiae and Pichia pastoris for structural genomics. *J. Struct. Funct. Genomics* **2004,** *5* (1-2), 29-44.

40. Junttila, M. R.; Sarrinen, S.; Schmidt, T.; Kast, J.; Westermarck, J., Single-step Strep-tag purification for the isolation and identification of protein complexes from mammalian cells. *Proteomics* **2005,** *5* (5), 1199-1203.

41. Pahler, A.; Hendrickson, W. A.; Kolks, M. A. G.; Argarana, C. E.; Cantor, C. R., Characterization and crystallization of core streptavidin. *J. Biol. Chem.* **1987,** *262* (29), 13933-7.

42. Chaier, L.; Wolf, F. J., The properties of streptavidin, a biotin-binding protein produced by streptomycetes. *Arch. Biochem. Biophys.* **1964,** *106* (1), 1-5.

43. Barrette-Ng, I. H.; Wu, S.-C.; Tjia, W.-M.; Wong, S.-L.; Ng, K. K. S., The structure of the SBP-Tag-streptavidin complex reveals a novel helical scaffold bridging binding pockets on separate subunits. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2013,** *69* (5), 879-887.

44. Young, C. L.; Britton, Z. T.; Robinson, A. S., Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **2012,** *7* (5), 620-634.

45. Skerra, A.; Schmidt, T. G. M., Use of the Strep-tag and streptavidin for detection and purification of recombinant proteins. *Methods Enzymol.* **2000,** *326* (Applications of Chimeric Genes and Hybrid Proteins, Pt. A), 271-304.

46. Connelly, P. R.; Varadarajan, R.; Sturtevant, J. M.; Richards, F. M., Thermodynamics of protein-peptide interactions in the ribonuclease S system studied by titration calorimetry. *Biochemistry* **1990,** *29* (25), 6108-14.

47. Raines, R. T.; McCormick, M.; Van Oosbree, T. R.; Mierendorf, R. C., The S.Tag fusion system for protein purification. *Methods Enzymol* **2000,** *326*, 362-76.

48. Holmgren, A., Thioredoxin. *Annu Rev Biochem* **1985,** *54*, 237-71.

49. Thelander, L.; Reichard, P., Reduction of ribonucleotides. *Annu. Rev. Biochem.* **1979,** *48*, 133-58.

50. La Vallie, E. R.; Lu, Z.; Diblasio-Smith, E. A.; Collins-Racie, L. A.; McCoy, J. M., Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. *Methods Enzymol.* **2000,** *326* (Applications of Chimeric Genes and Hybrid Proteins, Pt. A), 322-340.

51. Katti, S. K.; LeMaster, D. M.; Eklund, H., Crystal structure of thioredoxin from Escherichia coli at 1.68 Å resolution. *J. Mol. Biol.* **1990,** *212* (1), 167-84.

52. LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; McCoy, J. M., A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Bio/Technology* **1993,** *11* (2), 187-93.

53. Smith, D. B.; Johnson, K. S., Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988,** *67* (1), 31-40.

54. Mitchell, D. A.; Marshall, T. K.; Deschenes, R. J., Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. *Yeast* **1993,** *9* (7), 715-22.

55. Medina, D.; Moskowitz, N.; Khan, S.; Christopher, S.; Germino, J., Rapid purification of protein complexes from mammalian cells. *Nucleic Acids Res.* **2000,** *28* (12), e61, ii-viii.

56. Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; Dirr, H., Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **1997,** *6* (2), 399-406.

57. Nemoto, T.; Ota, M.; Ohara-Nemoto, Y.; Kaneko, M., Identification of dimeric structure of proteins by use of the glutathione S-transferase-fusion expression system. *Anal. Biochem.* **1995,** *227* (2), 396-9.

58. Duplay, P.; Hofnung, M., Two regions of mature periplasmic maltose-binding protein of Escherichia coli involved in secretion. *J. Bacteriol.* **1988,** *170* (10), 4445-50.

59. Pryor, K. D.; Leiting, B., High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expression Purif.* **1997,** *10* (3), 309-319.

60. Gusarov, I.; Nudler, E., Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell (Cambridge, MA, U. S.)* **2001,** *107* (4), 437-449.

61. Davis, G. D.; Elisee, C.; Newham, D. M.; Harrison, R. G., New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnol. Bioeng.* **1999,** *65* (4), 382-388.

62. Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R., Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **2006,** *15* (1), 182-189.

63. Mah, T.-F.; Kuznedelov, K.; Mushegian, A.; Severinov, K.; Greenblatt, J., The α subunit of Escherichia coli RNA polymerase activates RNA binding by NusA. *Genes Dev.* **2000,** *14* (20), 2664-2675.

64. Rajan, S.; Plant, L. D.; Rabin, M. L.; Butler, M. H.; Goldstein, S. A. N., Sumoylation silences the plasma membrane leak K+ channel K2P1. *Cell (Cambridge, MA, U. S.)* **2005,** *121* (1), 37-47.

65. Martin, S.; Nishimune, A.; Mellor, J. R.; Henley, J. M., SUMOylation regulates kainate-receptor-mediated synaptic transmission. *Nature (London, U. K.)* **2007,** *447* (7142), 321-325.

66. Li, S.-J.; Hochstrasser, M., A new protease required for cell-cycle progression in yeast. *Nature (London)* **1999,** *398* (6724), 246-251.

67. Kishi, A.; Nakamura, T.; Nishio, Y.; Maegawa, H.; Kashiwagi, A., Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation. *Am. J. Physiol.* **2003,** *284* (4, Pt. 1), E830-E840.

68. Butt, T. R.; Jonnalagadda, S.; Monia, B. P.; Sternberg, E. J.; Marsh, J. A.; Stadel, J. M.; Ecker, D. J.; Crooke, S. T., Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (8), 2540-4.

69. Zuo, X.; Li, S.; Hall, J.; Mattern, M. R.; Tran, H.; Shoo, J.; Tan, R.; Weiss, S. R.; Butt, T. R., Enhanced Expression and Purification of Membrane Proteins by SUMO Fusion in Escherichia coli. *J. Struct. Funct. Genomics* **2005,** *6* (2-3), 103-111.

70. Khorasanizadeh, S.; Peters, I. D.; Roder, H., Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **1996,** *3* (2), 193-205.

71. Armstrong, R. N., Glutathione S-transferases: reaction mechanism, structure, and function. *Chem. Res. Toxicol.* **1991,** *4* (2), 131-40.

72. Korzekwa, K. R.; Jones, J. P., Predicting the cytochrome P450 mediated metabolism of xenobiotics. *Pharmacogenetics* **1993,** *3* (1), 1-18.

73. Hayes, J. D.; Pulford, D. J., The glutathione S-transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.* **1995,** *30* (6), 445-600.

74. Chasseaud, L. F., The role of glutathione and glutathione S-transferases in the metabolism of chemical carcinogens and other electrophilic agents. *Adv Cancer Res* **1979,** *29*, 175-274.

75. Hayes, J. D.; Flanagan, J. U.; Jowsey, I. R., Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **2005,** *45*, 51-88, 1 plate.

76. Dirr, H. W.; Reinemer, P., Equilibrium unfolding of class π glutathione S-transferase. *Biochem. Biophys. Res. Commun.* **1991,** *180* (1), 294-300.

77. Erhardt, J.; Dirr, H., Native dimer stabilizes the subunit tertiary structure of porcine class pi glutathione S-transferase. *Eur. J. Biochem.* **1995,** *230* (2), 614-20.

78. Eaton, D. L.; Bammler, T. K., Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicol. Sci.* **1999,** *49* (2), 156-164.

79. Pickett, C. B.; Telakowskihopkins, C. A.; Argenbright, L.; Lu, A. Y. H., Regulation of glutathione S-transferase mRNAs by phenobarbital and 3-methylcholanthrene: analysis using cDNA probes. *Biochem. Soc. Trans.* **1984,** *12* (1), 71-4.

80. Mannervik, B.; Danielson, U. H., Glutathione transferases - structure and catalytic activity. *CRC Crit. Rev. Biochem.* **1988,** *23* (3), 283-337.

81. Mannervik, B.; Alin, P.; Guthenberg, C.; Jensson, H.; Tahir, M. K.; Warholm, M.; Jornvall, H., Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. *Proc Natl Acad Sci U S A* **1985,** *82* (21), 7202-6.

82. Jakoby, W. B.; Ketterer, B.; Mannervik, B., Glutathione transferases: nomenclature. *Biochem. Pharmacol.* **1984,** *33* (16), 2539-40.

83. Allocati, N.; Federici, L.; Masulli, M.; Di Ilio, C., Glutathione transferases in bacteria. *Febs J.* **2009,** *276* (1), 58-75.

84. Warholm, M.; Guthenberg, C.; Mannervik, B., Molecular and catalytic properties of glutathione transferase μ from human liver: an enzyme efficiently conjugating epoxides. *Biochemistry* **1983,** *22* (15), 3610-17.

85. Board, P. G.; Webb, G. C., Isolation of a cDNA clone and localization of human glutathione S-transferase 2 genes to chromosome band 6p12. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84* (8), 2377-81.

86. Aalin, P.; Mannervik, B.; Joernvall, H., Structural evidence for three different types of glutathione transferase in human tissues. *FEBS Lett.* **1985,** *182* (2), 319-22.

87. Board, P. G.; Mannervik, B., The contribution of the C-terminal sequence to the catalytic activity of GST2, a human alpha-class glutathione transferase. *Biochem. J.* **1991,** *275* (1), 171-4.

88. Gardner, J. L.; Gallagher, E. P., Development of a Peptide Antibody Specific to Human Glutathione S-Transferase Alpha 4-4 (hGSTA4-4) Reveals Preferential Localization in Human Liver Mitochondria. *Arch. Biochem. Biophys.* **2001,** *390* (1), 19-27.

89. Armstrong, R. N., Glutathione S-transferases: Structure and mechanism of an archetypical detoxication enzyme. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1994,** *69*, 1-44.

90. Smith, D. B.; Davern, K. M.; Board, P. G.; Tiu, W. U.; Garcia, E. G.; Mitchell, G. F., Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83* (22), 8703-7.

91. Ji, X.; Zhang, P.; Armstrong, R. N.; Gilliland, G. L., The three-dimensional structure of a glutathione S-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-Å resolution. *Biochemistry* **1992,** *31* (42), 10169-84.

92. Guthenberg, C.; Mannervik, B., Glutathione S-transferase (transferase π) from human placenta is identical or closely related to glutathione S-transferase (transferase ρ) from erythrocytes. *Biochim. Biophys. Acta, Enzymol.* **1981,** *661* (2), 255-60.

93. Reinemer, P.; Dirr, H. W.; Ladenstein, R.; Schaeffer, J.; Gallay, O.; Huber, R., The three-dimensional structure of class π glutathione S-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *Embo J.* **1991,** *10* (8), 1997-2005.

94. Wang, T.; Arifoglu, P.; Ronai, Z. e.; Tew, K. D., Glutathione S-transferase P1-1 (GSTP1-1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. *J. Biol. Chem.* **2001,** *276* (24), 20999-21003.

95. Brophy, P. M.; Barrett, J., Glutathione transferase in helminths. *Parasitology* **1990,** *100* (2), 345-9.

96. Kloetzel, K.; Lewert, R. M., Pigment formation in Schistosoma mansoni infections in the white mouse. *Am J Trop Med Hyg* **1966,** *15* (1), 28-31.

97. Rufer, A. C.; Thiebach, L.; Baer, K.; Klein, H. W.; Hennig, M., X-ray structure of glutathione S-transferase from Schistosoma japonicum in a new crystal form reveals flexibility of the substrate-binding site. *Acta Crystallogr., Sect. F Struct. Biol. Cryst. Commun.* **2005,** *61* (3), 263-265.

98. Lim, K.; Ho, J. X.; Keeling, K.; Gilliland, G. L.; Ji, X.; Rueker, F.; Carter, D. C., Three-dimensional structure of Schistosoma japonicum glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV. *Protein Sci.* **1994,** *3* (12), 2233-44.

99. McTigue, M. A.; Williams, D. R.; Tainer, J. A., Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel. *J. Mol. Biol.* **1995,** *246* (1), 21-7.

100. Manoharan, T. H.; Gulick, A. M.; Reinemer, P.; Dirr, H. W.; Huber, R.; Fahl, W. E., Mutational substitution of residues implicated by crystal structure in binding the substrate glutathione to human glutathione S-transferase π. *J. Mol. Biol.* **1992,** *226* (2), 319-22.

101. Amann, E.; Brosius, J.; Ptashne, M., Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. *Gene* **1983,** *25* (2-3), 167-78.

102. Li, Y., Recombinant production of antimicrobial peptides in Escherichia coli: A review. *Protein Expression Purif.* **2011,** *80* (2), 260-267.

103. Frangioni, J. V.; Neel, B. G., Solubilization and purification of enzymically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **1993,** *210* (1), 179-87.

104. Maru, Y.; Afar, D. E.; Witte, O. N.; Shibuya, M., The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. *J. Biol. Chem.* **1996,** *271* (26), 15353-15357.

105. Marston, F. A. O., The purification of eukaryotic polypeptides synthesized in Escherichia coli. *Biochem. J.* **1986,** *240* (1), 1-12.

106. Beekman, J. M.; Austin; Cooney, J.; Elliston, J. F.; Tsai, S. Y.; Tsai, M.-J., A rapid one-step method to purify baculovirus-expressed human estrogen receptor to be used in the analysis of the oxytocin promoter. *Gene* **1994,** *146* (2), 285-9.

107. Rudert, F.; Visser, E.; Gradl, G.; Grandison, P.; Shemshedini, L.; Wang, Y.; Grierson, A.; Watson, J., pLEF, a novel vector for expression of glutathione S-transferase fusion proteins in mammalian cells. *Gene* **1996,** *169* (2), 281-2.

108. Zhan, Y.; Song, X.; Zhou, G. W., Structural analysis of regulatory protein domains using GST-fusion proteins. *Gene* **2001,** *281* (1-2), 1-9.

109. Lally, J. M.; Newman, R. H.; Knowles, P. P.; Islam, S.; Coffer, A. I.; Parker, M.; Freemont, P. S., Crystallization of an intact GST-estrogen receptor hormone binding domain fusion protein. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **1998,** *D54* (3), 423-426; Vikis Haris, G.; Guan, K.-L., Glutathione-S-transferase-fusion based assays for studying protein-protein interactions. *Methods Mol Biol* **2004,** *261*, 175-86.

110. Singh, C. R.; Asano, K., Localization and characterization of protein-protein interaction sites. *Methods Enzymol.* **2007,** *429* (Translation Initiation: Extract Systems and Molecular Genetics), 139-161.

111. Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M., Global analysis of protein activities using proteome chips. *Science (Washington, DC, U. S.)* **2001,** *293* (5537), 2101-2105.

112. Yip, Y. L.; Smith, G.; Ward, R. L., Comparison of phage pIII, pVIII and GST as carrier proteins for peptide immunization in Balb/c mice. *Immunol. Lett.* **2001,** *79* (3), 197-202.

113. Andersson, L.; Blomberg, L.; Flegel, M.; Lepsa, L.; Nilsson, B.; Verlander, M., Large-scale synthesis of peptides. *Biopolymers* **2000,** *55* (3), 227-250.

114. Leong, L. E. C., The use of recombinant fusion proteases in the affinity purification of recombinant proteins. *Mol. Biotechnol.* **1999,** *12* (3), 269-274.

115. Jenny, R. J.; Mann, K. G.; Lundblad, R. L., A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression Purif.* **2003,** *31* (1), 1-11.

116. Waugh, D. S., An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression Purif.* **2011,** *80* (2), 283-293.

117. Baneyx, F., Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **1999,** *10* (5), 411-421.

118. Bush-Pelc, L. A.; Marino, F.; Chen, Z.; Pineda, A. O.; Mathews, F. S.; Di Cera, E., Important Role of the Cys-191-Cys-220 Disulfide Bond in Thrombin Function and Allostery. *J. Biol. Chem.* **2007,** *282* (37), 27165-27170.

119. Gallwitz, M.; Enoksson, M.; Thorpe, M.; Hellman, L., The extended cleavage specificity of human thrombin. *PLoS One* **2012,** *7* (2), e31756.

120. Guan, K.; Dixon, J. E., Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **1991,** *192* (2), 262-7.

121. Heidtmann, H.-H.; Kontermann, R. E., Cloning and recombinant expression of mouse coagulation factor X. *Thromb. Res.* **1998,** *92* (1), 33-41.

122. Jackson, C. M.; Johnson, T. F.; Hanahan, D. J., Bovine Factor X. I. Large-scale purification of the bovine plasma protein possessing Factor X activity. *Biochemistry* **1968,** *7* (12), 4492-505.

123. Vergis, J. M.; Wiener, M. C., The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expression Purif.* **2011,** *78* (2), 139-142.

124. Blommel, P. G.; Fox, B. G., A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression Purif.* **2007,** *55* (1), 53-68.

125. Miladi, B.; Bouallagui, H.; Dridi, C.; El Marjou, A.; Boeuf, G.; Di Martino, P.; Dufour, F.; Elm'Selmi, A., A new tagged-TEV protease: Construction, optimisation of production, purification and test activity. *Protein Expression Purif.* **2011,** *75* (1), 75-82.

126. Kapust, R. B.; Toezser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S., Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **2001,** *14* (12), 993-1000.

127. Nallamsetty, S.; Kapust, R. B.; Toezser, J.; Cherry, S.; Tropea, J. E.; Copeland, T. D.; Waugh, D. S., Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expression Purif.* **2004,** *38* (1), 108-115.

2. Rapid and efficient purification of recombinant peptides and low molecular weight proteins

2.1. Abstract

The synthesis and purification of peptides of importance in the fields of research and medicine continue to be a challenging task. Chemical synthesis of oligopeptides, especially those greater than 25 amino acids, is cost prohibitive. On the other hand, several bottlenecks exist in the production of recombinant short peptides in heterologous expression hosts such as *Escherichia coli* (*E.coli*).

In this study, a rapid, cost-effective, and reliable method for the production and single-step-purification of peptides and small proteins was developed. Peptides/ proteins were overexpressed in *E.coli* as GST-fusion products in high yields. The recombinant peptides/ proteins were successfully purified after enzymatic cleavage followed with selective heat-induced precipitation of the GST-affinity tag. Qualitative and quantitative analysis using SDS-PAGE and mass spectrometric methods suggest that the recombinant peptides/ proteins were purified to >95% homogeneity. Results of biophysical experiments, including multi-dimensional NMR spectroscopy, show that the purified proteins/ peptides retain their native conformation. Isothermal titration studies indicate no significant change in the binding affinity of the heat treated purified product to their interacting partner(s) compared to the recombinant peptides purified by conventional chromatographic procedures without subjecting to heat treatment. In our opinion, the results reported are expected to render the purification of recombinant proteins/ peptides of biomedical relevance easy and reliable.

2.2. Introduction

Peptides and small proteins are known to play a key role in various biological processes. They can be hormones and neurotransmitters[1,2], or growth and differentiation factors[3,4], which trigger signaling cascades upon interaction with the cell surface receptors[1,5]. Other peptides and small proteins are also commonly used as inhibitors for targeting enzymes[6], biomarkers for the early prediction of several diseases[7,8,9] and also act as therapeutics and anti-microbial agents[10,11,12,13,14,15,16]. Current recombinant production procedures for the peptides and small proteins do not protect them completely from proteases present in the host expression platforms due to their small size or and the presence of highly charged residues[17,18,19]. Overexpression of recombinant peptides and small proteins with a larger affinity tag seem to give them greater stability and an increased proteolytic resistance[20]. These affinity tags also contribute to enhanced expression yields and accelerate the purification process[21,22,23].

Chemical synthesis of peptides was first introduced to the research community after du Vigneaud's synthesis of oxytocin in 1954[24,25]. Yet, factors such as coupling efficiency and steric hindrance of larger side chains or protective groups limit this process. Moreover, the peptides' tendencies to aggregate can often result in low yields[26,27]. For that reason, the recombinant protein production can be the preferred, more expandable, and viable method for target biomolecules, especially if larger than 25 amino acids[28,29]. In addition to experimental restrictions, the economic and environmental impacts of chemical peptide synthesis including disposal costs and complying with safety regulations should be considered[30]. Another shortcoming is the production of isotope labeled peptides. These "heavy" peptides can be valuable to acquire complete information on the protein structure and dynamics or for the elucidation of peptide-protein binding interactions. However, obtaining ^{15}N-labeled peptides

through chemical synthesis is expensive, less environment friendly, and more difficult with increasing length[30]. The use of recombinant proteins might be a better way for the production of ^{15}N-labeled peptides as their labeling process is very effective and genetically controlled with high fidelity[31].

Glutathione S-transferase (GST) is one of the most popular affinity tags used as fusion partner for expressing diverse proteins in both prokaryotic and eukaryotic expression systems[32,33,34]. It is well recognized to stabilize the fusion protein due to its high solubility in the *E.coli* cytosol[35,36] and can be purified using a single-step affinity chromatographic procedure[37]. In this research study, the overexpression and one-step purification of GST-tagged small proteins and peptides has been successfully demonstrated. Subsequently, the tagged protein/ peptide products were subjected to enzymatic cleavage and the cleavage products were purified to homogeneity by using a simple heat treatment. This purification procedure did not show any loss in biological activity of the target peptides/small proteins. Furthermore, there were no changes in the conformation of the biomolecule detected when compared to the versions purified by the conventional method and therefore can be used in a variety of physiological assays. Based on the diverse examples that were examined, we believe that this method can be generically used to purify peptides and proteins, whose T_m is greater than 65 °C.

2.3. Materials and Methods

LB Broth (Miller) and Amicon ultrafiltration centrifugal concentrators were purchased from EMD Millipore. Ampicillin, NaCl, KCl, Na_2HPO_4, Tris-HCl were purchased at J.T. Baker Chemicals, Isopropyl-1-thio-β -D-galactopyranoside (IPTG) at OMNI Chemicals. Reduced

glutathione and thrombin were obtained from Sigma Aldrich. The secondary anti-mouse IgG antibody conjugated with alkaline phosphatase is a product of Genescript Inc.

Expression and Purification of the GST-fused peptide/ proteins

LB broth containing ampicillin (100 μg/ml) was inoculated with 5% (v/v) of freshly grown bacterial culture under aseptic conditions and incubated at 37 °C and 250 rpm. Once the OD_{600} reached 0.6, the cells were induced with 1mM IPTG and further incubated for four hours. Cells were harvested at 6,000 rpm for 20 minutes at 4 °C and the pellets were washed with 1x PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2mM KH_2PO_4; pH 7.2) either for immediate use or for storage at -20 °C.

E.coli cells containing the recombinant GST-fusion protein was resuspended in 20 ml 1x PBS (pH 7.2) and subjected to cell lysis by ultrasonication (Mirsonic Inc). Insoluble cell debris were removed by centrifugation at 19,000 rpm for 30 minutes. The clear cell lysate was loaded onto a pre-equilibrated GSH-Sepharose column (GELifeSciences MA, USA) at a flow rate of 1 ml/min, followed by washing with 1x PBS buffer until a flat baseline was reached to eliminate all contaminating bacterial proteins. The GST-fusion protein was eluted with 10 mM reduced glutathione in 1x PBS buffer. For subsequent off-column thrombin cleavage, the eluted fraction was subjected to ultrafiltration using centrifugal spin concentrators (EMD Millipore MA, USA) with a molecular weight cut off of 10 kDa. Protein concentration was estimated by measuring the absorbance at 280 nm. Samples monitoring the purification were resolved on 15% SDS–PAGE under reduced conditions according to the method of Laemmli[38].

Enzymatic cleavage and purification of peptides/proteins by heat incubation

The GST-tag was cleaved by subjecting it to thrombin at the ratio of 1U of enzyme for every 0.25 mg of fusion protein. The pure small protein/ peptide was separated from the cleaved mixture by incubating the sample at 65 °C for 20 minutes, and then followed by two-time centrifugation at 13,000 rpm for 10 minutes.

In comparison, the conventional purification method involves off-column thrombin cleavage and reloading the sample onto a pre-equilibrated GSH-Sepharose column. While the GST-tag binds to the resin, the small recombinant protein or peptide of interest is found in the flow through fraction. Protein concentrations were estimated by measuring the absorbance at 280 nm, while peptide concentrations were assessed by using the Brij method[39].

Determination of the temperature of precipitation/ aggregation of GST

In order to establish the appropriate temperature for the heat treatment, cleavage products consisting of the GST-tag and the target recombinant protein were subjected to 20 minute incubations at temperatures ranging from 40-85 °C, followed by centrifugation to separate the aggregated protein in the pellet from the soluble component(s) in the supernatant fraction. Subsequently, to determine the fate of GST during the heat treatment process, the absorbance of pure GST-tag at 350 nm was monitored aggregation at temperatures ranging from 40-80 °C. Furthermore, it was verified that the selective removal of GST is feasible at various buffer conditions. The intrinsic fluorescence spectra of heat-treated GST were monitored at different salt concentrations (0 mM, 137 mM, 500 mM and 800 mM NaCl) as well as different pH conditions ranging from 6-8 to confirm that GST was still removed from the supernatant due to heat under those buffer conditions. Intrinsic tryptophan fluorescence spectra of the samples were

collected at 25 °C using a Hitachi F-2500 spectrofluorometer at 2.5 nm resolution, with an excitation wavelength at 280 nm.

Western Blot analysis using anti-GST antibodies to verify the purity of samples

In order to examine if the cleaved GST was completely removed during the heat treatment procedure, a Western Blot with monoclonal antibodies raised against the GST-tag was performed. Samples of purification were resolved on a 12% SDS-PAGE under reduced conditions and the protein bands were transferred onto a nitrocellulose membrane with 100 V and 75 mA for 90 minutes. Subsequently, the membrane was blocked in 5% skim milk (dissolved in 1x TBS-T: 10 mM Tris, 100 mM NaCl, 0.05% Tween-20; pH 7.4), washed, and then incubated overnight in 0.2% BSA in 1x TBS-T containing the primary antibody (titer 1:2500). After washing the membrane three times, the membrane was incubated for 2 hours with 0.2% BSA in 1x TBS-T including the secondary AP-conjugated antibody (titer 1:2500). After washing the blot, bands were visualized using NBT/BCIP (Thermo Fisher Scientific Inc., MA, USA) as a substrate for the alkaline phosphatase (AP).

Comparison of the Secondary Structure using Circular Dichroism (CD)

CD data were recorded as an average of 3 accumulations at 25 °C using a Jasco J-720 spectropolarimeter. Far UV CD spectra of CD2 and AlbM4 (100 μM) in 1x PBS pH 7.2 were recorded using a quartz cell of 0.1 mm path length in the standard sensitivity mode with a scan speed of 50 nm per minute. Appropriate blank corrections were made in the CD spectra. The CD data are expressed as molar ellipticity (deg x cm^2 x $dmol^{-1}$).

Binding Studies by Isothermal Titration Calorimetry

Isothermal titration calorimetric experiments were performed using the ITC200 (MicroCal Inc., Northampton, MA) at 25 °C to examine the functionality of the recombinant proteins/ peptides purified using the heat treatment method. Chloroplast signal recognition particles (cpSRP) function as a heterodimer, which consists of subunits cpSRP43 and cpSRP54. Particularly the CD2 domain of cpSRP43 binds to a 10-residue peptide fragment of cpSRP54. CD2 was dialyzed against 1× PBS pH 7.2. Samples were subjected to centrifugation to remove any aggregated or precipitated material and were degassed before the titration. Concentrations of synthetic 54-peptide (Peptides International, Louiseville, KY) to CD2 were maintained at a molar ratio of 10:1. The contents of the syringe (54-peptide) were added sequentially in 1.3 μL aliquots to the cell (CD2) with a 12 second interval between injections. Using Origin Version 7.0 software, heats of reaction per injection (μcalories/s) were determined by the integration of peak areas. Thermodynamic values were derived after fitting the data using a one-site of binding model available in Origin 7.0. The fit provides values of the heat of binding ($\Delta H°$), the stoichiometry of binding (n), and the dissociation constants (K_d) from plots of the heat evolved per mole of ligand injected *versus* 54-peptide/ CD2 ratio.

Comparison studies using Differential Scanning Calorimetry and thermal denaturation using intrinsic fluorescence

Heat capacities of the heat-treated CD2 and the fusion protein GST-CD2 were measured as a function of temperature at pH 7.2 using NANO DSCIII with a ramping temperature of 1 °C/min from 15-90 °C. Thermal denaturation scans were performed using a protein concentration of 1 mg/mL. The protein solution was degassed prior to acquisition of DSC data.

Both the heating and cooling cycles were recorded to examine the reversibility of the thermal unfolding process.

In case of the conventionally purified CD2, the intrinsic fluorescence of the protein was monitored at increasing temperatures ranging from 40-95 °C using a Hitachi F-2500 spectrofluorometer at 2.5 nm resolution, with an excitation wavelength of 280 nm. All fluorescence measurements were conducted at a protein concentration of 100 μg/ml in 1× PBS pH 7.2. Appropriate blank corrections were made to subtract for background noise.

Monitoring the backbone conformation of the target recombinant protein by ^{1}H-^{15}N HSQC

Multidimensional nuclear magnetic resonance spectroscopy (NMR) is a very powerful technique, which is regularly used for the characterization of 3D structure and backbone dynamics at atomic resolution. This technique allows the comparison of the 3D solution structure of heat-treated and conventionally purified CD2 through the acquisition of ^{1}H-^{15}N HSQC spectra of the heat-treated and conventionally purified CD2. ^{15}N enriched protein samples were prepared by growing the *E.coli* cells in M9 minimal medium supplemented with vitamin solutions. ^{1}H-^{15}N HSQC spectra of protein samples in 1x PBS were acquired at room temperature and at a concentration of 0.5 mM.^{1}H-^{15}N cross-peaks were assigned in the spectra of CD2 and a ^{1}H-^{15}N chemical shift perturbation plot was generated. NMR experiments were carried out on Bruker 700 MHz and 500 MHz spectrometers, which are equipped with cryo-probes.

MALDI-MS Analysis of WAP and HB peptide

MALDI-MS was performed to analyze the purity and size of the recombinant peptides. Prior to MALDI-TOF (Bruker Daltonics) analysis, the recombinant WAP and HB-peptide (~50-

100µg) samples were desalted by passing through "ZIPTM" tips (C-18 matrix). The theoretical molecular weight of WAP and the HB peptide were calculated using the ProtParam tool from Expasy[40] and was found to be 7315Da and 3767Da, respectively, and was compared with the experimental value.

Binding studies of WAP and HB-peptide by ITC

ITC experiments monitoring WAP and the HB-peptide's ability to bind heparin were performed as stated earlier. WAP and HB-peptide were dialyzed against 1x PBS pH 7.2 or 10 mM Phosphate Buffer containing 100 mM NaCl respectively. Concentrations of heparin to WAP or heparin to HB-peptide were maintained at 10:1 and 20:1 respectively.

2.4. Results

Purification of the GST-fusion protein products and cleavage using thrombin

Glutathione-based affinity chromatography of GST-tagged fusion proteins is one of the most popular purification techniques and can undoubtedly be scaled up to generate milligram or gram quantities of recombinant proteins[32]. CD2 (6 kDa Chromo-domain 2 of chloroplast signal recognition particle 43) was purified to homogeneity (> 95%) using this well-established affinity chromatography method yielding 40 mg per 1 liter culture (Figure-1b lane 5). Furthermore, complete cleavage of the fusion proteins with thrombin was successfully achieved, as can be observed in the SDS-PAGE gel stained with Coomassie blue to monitor the purification of CD2 (Figure-1b, lane 7).

Purification of the cleaved fusion product using heat procedure

The cleaved fusion protein mixture was subjected to 65 °C for 20 minutes, precipitating the GST-tag. Subsequently, the affinity tag was efficiently separated from the supernatant by high-speed centrifugation leaving recombinant CD2 in solution. Coomassie stained SDS-PAGE gels clearly show the purity of fusion protein GST-CD2 (Figure-1b, lane 5) and the heat-treated recombinant target protein CD2 (Figure-1b, lane 9) as well as the complete cleavage of the fusion protein using thrombin. Results of the more sensitive Western Blot show that the contaminating GST was completely removed as GST was not detected by the monoclonal antibodies raised against the affinity tag (Figure-1c, lane 3). As panel A of Figure 1 shows, the alternative of the heat treatment method requires the introduction of another chromatography in order to isolate the protein of interest. While the affinity tag binds to the GSH-Sepharose column, the protein of interest elutes in the flow through. Due to the large volume of this fraction, it probably needs to be concentrated in order to use it for subsequent experiment. When comparing the yields of CD2 from these two purification methods, our method provides a more time-efficient and economic. Approximately 97% of the pure target molecule CD2 was recovered using the heat treatment compared to about 88% when reloading the cleavage mixture back onto the GSH- Sepharose column.

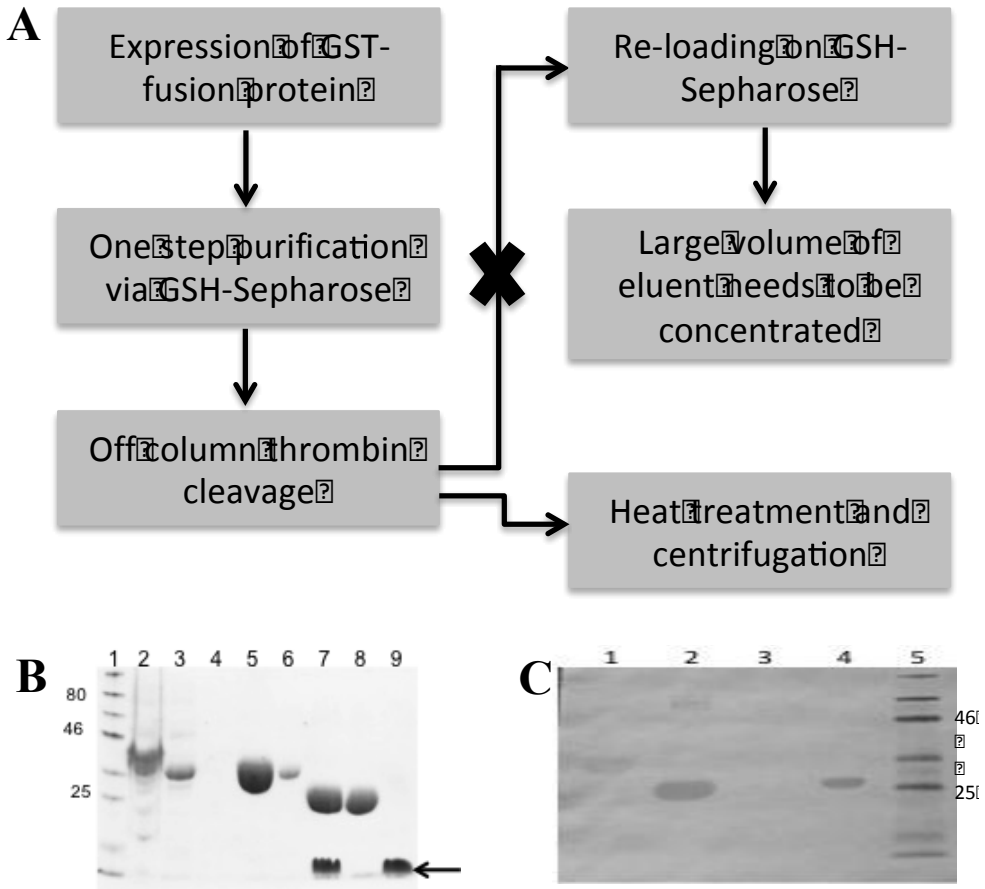


Figure 6: *a) Flow chart comparing conventional purification method and heat treatment method. b) SDS-PAGE of Purification of CD2 (6kDa) using heat treatment method: lane-1 pre-stained protein marker, lane-2 pellet post cell lysis, lane-3 supernatant post lysis, lane-4 flow through, lane-5 eluted GST-CD2, lane-6 8M urea, lane-7 cleaved GST-CD2 using thrombin, lane-8 pellet after heat treatment, lane-9 supernatant after heat treatment. c) Western Blot of heat treatment method: lane-1 GST-CD2, lane-2 cleaved GST-CD2, lane-3 supernatant post heat treatment, lane-4 pellet post heat treatment, lane-5 pre-stained protein marker.*

GST aggregates under the influence of heat

As one can observe in the heat treatment experiments, GST partly precipitates at a temperature of 50 and 55 °C. This is consistant with Kaplan et al's observations of GST's loss of enzyme activity and melting temperature, which was found to be at 51 °C[41]. At 65 °C, the GST-tag denatured and was found in the pellet (Figure-2, lane 11) while CD2 was still detected in the supernatant (Figure-2, lane 12). Consequently, we are able to conclude that this temperature is

required to completely remove the GST-tag from the supernatant. Moreover, by monitoring the absorbance of GST at 350 nm at increasing temperatures, a decrease in absorbance at 280 nm but an increase in turbidity was observed. Hence, we can conclude that GST aggregates during to the heat treatment process. These observations are independent of salt concentrations or pH of the buffer.

Figure 7 *SDS-PAGE of Heat treatment: Lane-1 pre-stained protein marker, lane-2 GST-CD2, lane-3 cleaved GST CD2, lane-4 supernatant 45 °C, lane-5 pellet 50 °C, lane-6 supernatant 50 °C, lane-7 pellet 55 °C, lane-8 supernatant 55 °C,* lane-*9 pellet 60 °C, lane-10 supernatant 60 °C, lane-11 pellet 65 °C, lane-12 supernatant 65 °C, lane-13 pellet 70 °C, lane-14 supernatant 70 °C, lane-15 pellet 75 °C, lane-16 supernatant 75 °C, lane-17 pellet 80 °C, lane-18 supernatant 80 °C, lane-19 pellet 8 5°C, lane-20 supernatant 85 °C. Circle represents temperature at which for the first time GST was completely removed from the supernatant.*

Comparison studies of the heat treated and conventionally purified small protein or peptide

The proposed heat treatment technique has shown to yield pure protein of interest. However, this method is only valuable if the recombinant target protein is alike to its conventionally purified counterpart in terms of structure, stability, and biological functionality.

Spectroscopic characterization of the secondary and tertiary structure of recombinant CD2

Far UV CD spectra measurements between 190 nm – 250 nm were used to observe changes in the secondary structure of the proteins or peptides. When overlaying the far UV CD

spectra of the heat treated as well as the conventionally purified CD2 shown in Figure 3a, it can be concluded that heat treatment did not disrupt the secondary structure of the recombinant CD2, as they are the same for both CD2 samples. Their CD profiles show similar secondary strcutural conformations of predominantly α-helical structures with the minima centered at 208 nm and 222 nm. Intrinsic steady-state tryptophan fluorescence gives insight on the tertiary structure of the protein of interest. An overlay of the emission spectra of both heat-treated and conventionally purified CD2 (Figure-3B) indicates that the tryptophans are located in a partially solvent exposed environment as indicated by the emission maximum at 341 nm. In order to elucidate that the heat treatment process did not disorganize the 3D solution structure, the ^{1}H-^{15}N HSQC spectra of heat-treated and conventionally purified CD2 were acquired. Superimposition of both ^{1}H-^{15}N HSQC spectra and the insignificant ^{1}H-^{15}N chemical shifts show that the heat treatment did not cause any changes in the solution structure of CD2 (Figure 3C).

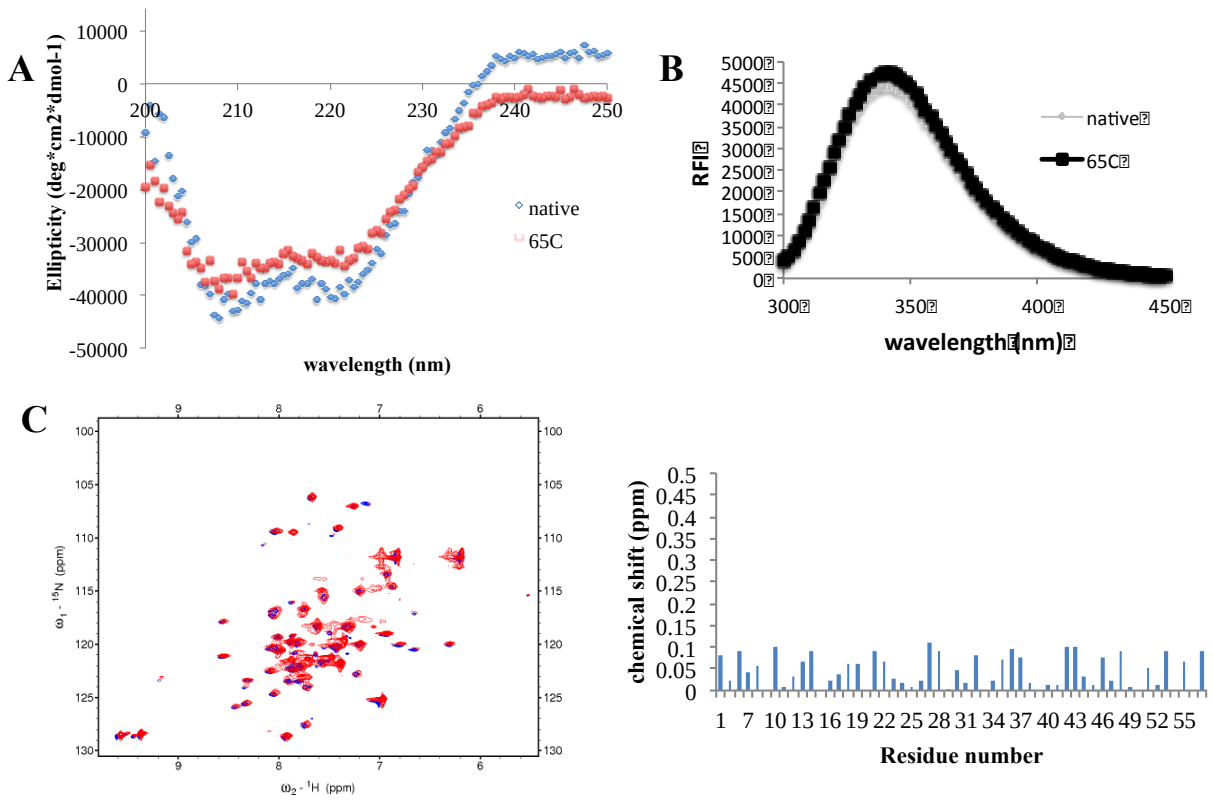


Figure 8 *A) Far UV Circular Dichroism spectra. B) intrinsic fluorescence emissions spectra and C) overlay of 2D $^{1}H^{15}N$-HSQC spectra and chemical shift perturbation plot of heat treated and conventionally purified CD2.*

Studies on the stability of the purified proteins

DSC and thermal denaturation experiments, based on intrinsic fluorescence, are able to directly measure and compare the thermal stability of heat-treated and conventionally purified CD2. The DSC profiles and thermal denaturation plot in Figure 4 show that the melting temperatures (T_m, the temperature at which 50% of the protein population exists in its folded conformation while the rest is in the unfolded conformation) of the recombinant CD2 purified by heat treatment are very similar to the protein purified by conventional GSH-Sepharose chromatography (83 °C). Therefore, it can be verified that the heat treatment method does not

significantly change the thermodynamic stability of the protein of interest. Interestingly, DSC experiments of the fusion protein give a T_m of 56 °C. This indicates that in case of an incomplete thrombin cleavage of the fusion protein, the contaminating fusion protein would also precipitate during the heat treatment procedure, leaving only the protein of interest in solution.

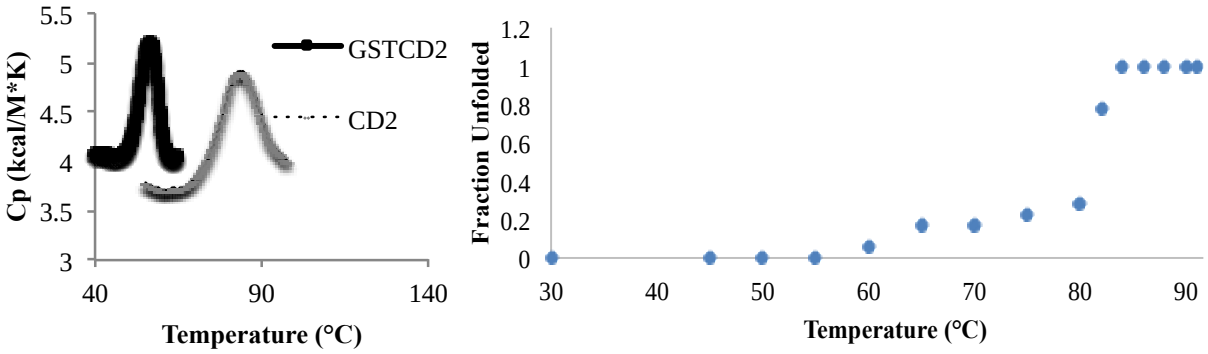


Figure 9 *Differential Scanning thermogram of heat treated CD2 and GST-CD2 and thermal denaturation of conventionally purified CD2.*

Comparison of the Functionality of the purified recombinant CD2

ITC experiments are a resourceful tool, which can directly measure the binding affinity, stoichiometry, and thermodynamics of an interaction. Chloroplast signal recognition particles (cpSRP) function as a heterodimer, which consists of subunits cpSRP43 and cpSRP54. Particularly the CD2 domain of cpSRP43 binds to a 10-residue peptide fragment of cpSRP54. The ITC profiles of the interaction between the CD2-domain of cpSRP43 with the 54-peptide motif are shown in Figure-5. Both heat-treated and conventionally purified CD2 display the characteristic one-site binding model with similar binding affinities (1.27 μM for heat-treated CD2 *vs*. 54-peptide in Figure-5A and 1.42 μM for conventionally purified CD2 *vs*. 54-peptide in Figure-5B). This correlates with results that have already been reported in previous studies by

the Kumar group[42]. This indicates that the heat treatment did not affect the biological interaction of recombinant CD2.

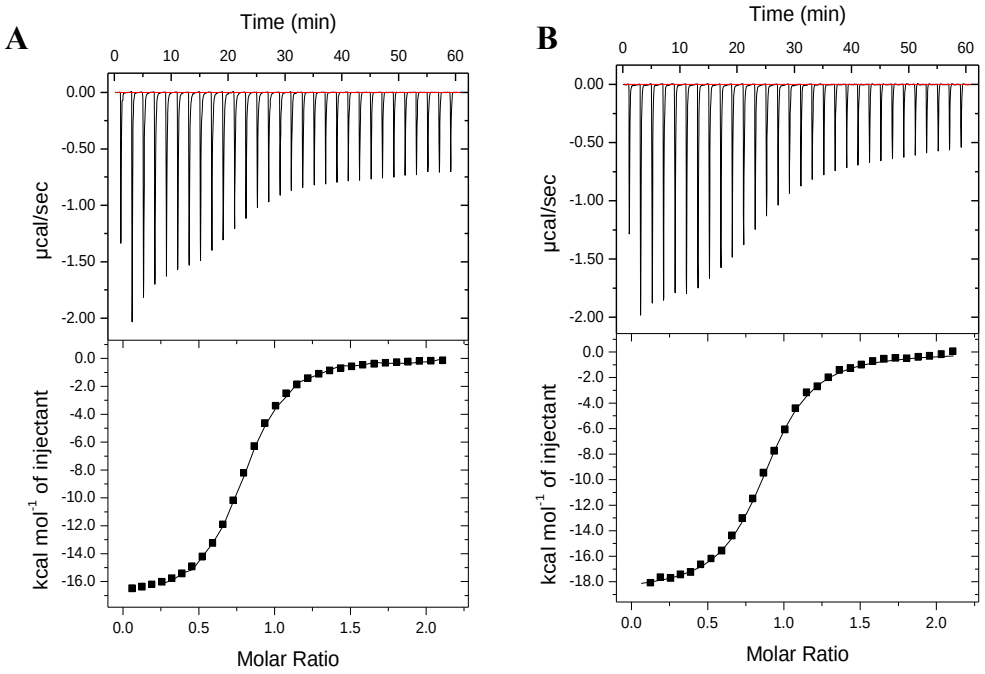


Figure 10*: Isothermal Titration Calorimetry of A) heat treated CD2 vs. 54 peptide (K_d= 1.27uM) and B) conventionally purified CD2 vs. 54 peptide (K_d= 1.42uM).*

Other examples of protein and peptide purified using the heat treatment method

In an attempt to expand this method, several diverse recombinant proteins and peptides were purified by this new method under the same conditions that were applied to CD2. The results of this study show that the heat treatment method is especially convenient when purifying recombinant peptides. AlbM4 is a 10-residue peptide motif of the protein cAlb, which is a ligand of the chloroplast signal recognition particle 43. This 1.5 kDa peptide is prone to aggregation during expression due to its highly positively charged character. By fusing the peptide to GST and applying the heat treatment method, the AlbM4 peptide was successfully isolated to homogeneity (Figure 6A). Moreover, an additional chromatography step was bypassed with the

heat treatment method, which prevented challenges such as reduced recovery of the target peptide due to the introduction of a secondary chromatography step. The far UV CD spectrum of AlbM4 (Figure-6B) displays a similar profile to the synthetic AlbM4 that was used in previous studies from the Kumar group. The CD profile (Figure-6B) shows a minimum at 205 nm, which is characteristic for a random coil structure and distinctive for most peptides.

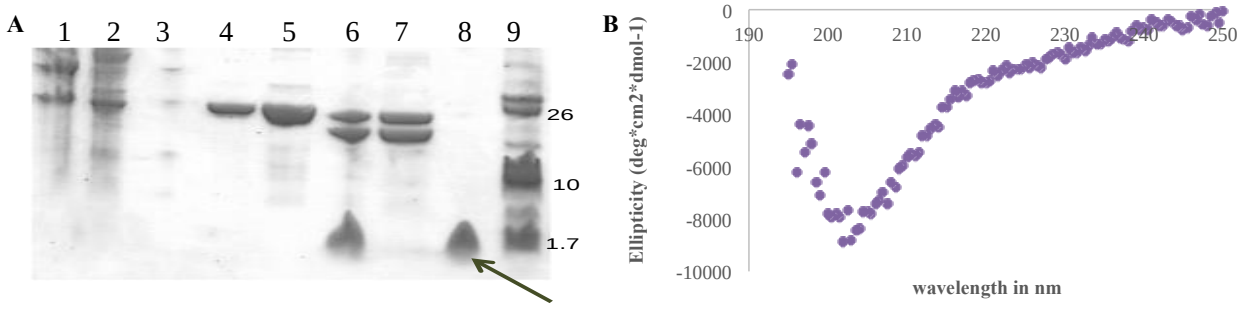


Figure 11 A) *Purification of AlbM4 peptide (1.5kDa). Lane-1 pellet after lysis, lane-2 supernatant after lysis, lane-3 flow through, lane-4&5 GST-AlbM4, lane-6 cleaved, lane-7 pellet after heat treatment, lane-8 supernatant after heat treatment, lane-9 pre-stained ultra low protein marker. B) Far UV CD spectrum confirms characteristic random coil secondary structure.*

The WAP-domain (7 kDa) of Anosmin-1 and the constructed heparin-binding (HB) peptide (3.7 kDa), which both have been shown to interact with heparin, represent more examples for the usefulness of the heat treatment method. MALDI-MS analysis of the small protein and the peptide confirm the size of product gained(Figure-7B and Figure-8B). The abubndace of the impurities present as additional peaks in the Mass Spectrum need to be quantified, for example by HPLC. Nevertheless, after applying just one chromatography step and the heat treatment method one can obtain a highly homogenous peptide or small protein sample as the ITC experiments confirmed similar binding capabilities. Both WAP and the HB-peptide retained their ability to interact with heparin. WAP and its ligand heparin display a characteristic

one-site binding model with a moderate binding affinity (K_d) of 590 μM and a binding stoichiometry of 1:1 (Figure 7C). This data confirms previously published work suggesting that WAP is a heparin-binding domain of Anosmin-1[43]. Furthermore, the affinity of the Hb-peptide to heparin was displayed, giving a K_d of 170 nM (Figure-8C).

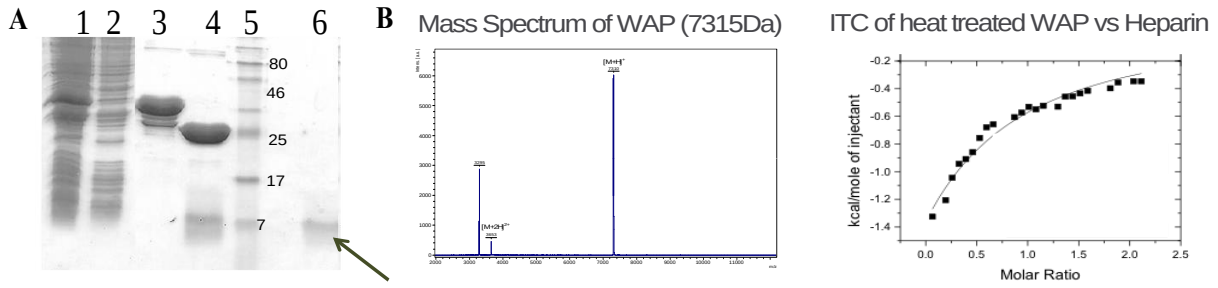


Figure 12 *A) Purification of WAP: Lane-1 pellet after lysis, lane-2 supernatant after lysis, lane-3 GST-WAP, lane-4 cleaved, lane-5 pre-stained protein marker, lane-6 WAP. B) Mass Spectrum confirms size of WAP. C) ITC of WAP vs. heparin.*

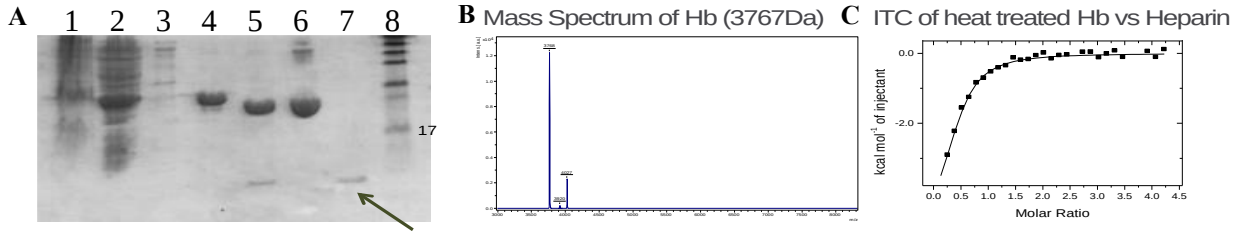


Figure 13 *A) Purification of the HB-peptide: Lane-1 pellet after lysis, lane-2 supernatant after lysis, lane-3 flow through, lane-4 GST-HB, lane-5 cleaved, lane-6 pellet after heat treatment, lane-7 HB-peptide, lane-8 pre-stained protein marker. B) Mass Spectrum confirms size of HB-peptide. C) ITC of HB-peptide vs. heparin.*

Likewise, the feasibility of this method to purify larger proteins was examined by applying the heat treatment procedure to the calcium-binding protein S100A13 (11.5 kDa) and the copper-binding domain C2B (18 kDa), both of which are important for the secretion of the fibroblast growth factor 1 (FGF1). It was discovered that while the 11.5kDa S100A13 is still

present in the supernatant (Figure-9A, lane 4), the 18kDa C2B precipitates along with GST upon being heated and were detected in the pellet (Figure-9B, lane 8). This might indicate the possible limitation as a function of protein molecular weight of this novel technique.

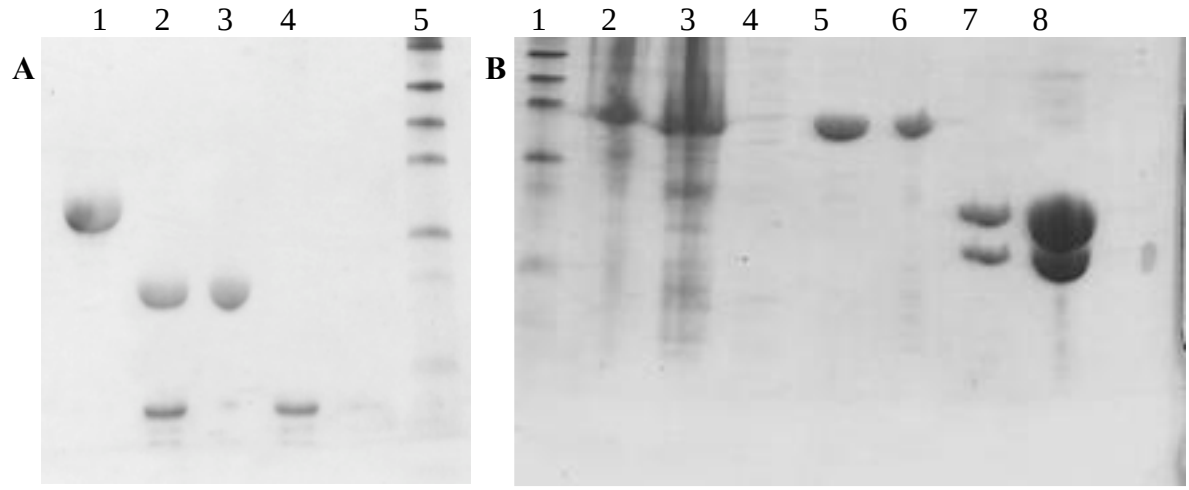


Figure 14 *A) SDS-PAGE depicting purification of S100A13. Lane-1 GST-S100A13, Lane-2 cleaved, Lane-3 pellet after heat treatment, Lane-4 supernatant after heat treatment, Lane-5 pre-stained protein marker. B) SDS-PAGE of the purification of C2B. Lane-1 pre-stained protein marker, Lane-2 pellet after lysis, Lane-3 supernatant after lysis, Lane-4 flow through, Lane-5 eluted GST-C2B, Lane-6 8M urea, Lane-7 cleaved, Lane-8 pellet after heat treatment.*

2.5. Discussion

The current study is mainly focused on the overexpression and purification of recombinant peptides and small proteins using the GST-affinity tag as a fusion partner. GST is known to dictate and improve the solubility of the fusion partner. In addition, because of its large size, the fused peptide is less susceptible to proteolytic degradation. Therefore, this is a commonly used affinity tag in the fields of molecular biology and is identified to express in very large quantities resulting in high yields of the fusion protein in diverse expression platforms. A rapid and efficient purification of various small proteins and peptides, which were expressed with the GST-tag and cleaved with thrombin, were successfully demonstrated. While chemical

treatments, such as CNBr, formic acid or hydroxylamine are very effective, they are also related to fairly harsh cleavage conditions, i.e. dramatic pH changes, which most likely are not useful due to their ability to denature proteins or induce modifications of the side chains[30,35]. Furthermore, in most cases CNBr is not preferred because most proteins contain methionine in their amino acid sequence. On the other hand, an enzymatic cleavage, like a thrombin cleavage, can be performed under mild, physiological conditions[44]. This is especially desired for the production of biologically active proteins and clinically important peptides. The cleavage recognition site can be readily engineered during the cloning process or is already located on the cloning vector of choice.

The heat treatment procedure successfully separated the digested fusion protein by exclusively eliminating the tag. Our studies confirm that the GST-tag completely precipitates when heat (65 °C) is applied. In 1997, Kaplan et al showed that Sj26GST can undergo thermal inactivation with a melting temperature at 52 °C[41]. In contrast, the small protein CD2 is resistant to this temperature. Our results show that this method does not significantly affect the 3D solution structure, stability, or biological activity of CD2. Moreover, the higher yield of recovery (97%) of the small molecule after thrombin cleavage that was obtained from the heat treatment method was demonstrated. Therefore, this new method is a valuable alternative purification approach for recombinant peptides and small proteins. Other methods described above require an additional time-consuming chromatography step. In addition, it was concluded that the heat treatment can be extended to the gain large quantity of various other proteins and peptides of clinical interest as well as other proteins that are known to be thermally stable at 65 °C. Another significant benefit of this new method is the practical and widely accessible production of isotope labeled peptides and small proteins because of using recombinant protein expression.

2.6. References

1. Kortenoeven, M. L. A.; Pedersen, N. B.; Rosenbaek, L. L.; Fenton, R. A., Vasopressin regulation of sodium transport in the distal nephron and collecting duct. *Am. J. Physiol.* **2015,** *309* (2, Pt. 2), F280-F299.

2. Ondrejcakova, M.; Ravingerova, T.; Bakos, J.; Pancza, D.; Jezova, D., Oxytocin exerts protective effects on in vitro myocardial injury induced by ischemia and reperfusion. *Can. J. Physiol. Pharmacol.* **2009,** *87* (2), 137-142.

3. Boonstra, J.; Rijken, P.; Humbel, B.; Cremers, F.; Verkleij, A.; van Bergen en Henegouwen, P., The epidermal growth factor. *Cell Biol. Int.* **1995,** *19* (5), 413-30.

4. Nevalainen, M. T.; Valve, E. M.; Ingleton, P. M.; Nurmi, M.; Martikainen, P. M.; Harkonen, P. L., Prolactin and prolactin receptors are expressed and functioning in human prostate. *J. Clin. Invest.* **1997,** *99* (4), 618-627.

5. Robinson, S. D.; Safavi-Hemami, H.; McIntosh, L. D.; Purcell, A. W.; Norton, R. S.; Papenfuss, A. T., Diversity of conotoxin gene superfamilies in the venomous snail, Conus victoriae. *PLoS One* **2014,** *9* (2), e87648/1-e87648/13, 13 pp.

6. Eldar-Finkelman, H.; Eisenstein, M., Peptide inhibitors targeting protein kinases. *Curr. Pharm. Des.* **2009,** *15* (21), 2463-2470.

7. Du, Q.-S.; Xie, N.-Z.; Huang, R.-B., Recent Development of Peptide Drugs and Advance on Theory and Methodology of Peptide Inhibitor Design. *Med. Chem. (Sharjah, United Arab Emirates)* **2015,** *11* (3), 235-247.

8. Bischoff, R.; Luider, T. M., Methodological advances in the discovery of protein and peptide disease markers. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2004,** *803* (1), 27-40.

9. Doust, J. A.; Glasziou, P. P.; Pietrzak, E.; Dobson, A. J., A systematic review of the diagnostic accuracy of natriuretic peptides for heart failure. *Arch. Intern. Med.* **2004,** *164* (18), 1978-1984.

10. Banting, F. G.; Campbell, W. R.; Fletcher, A. A., Further clinical experience with insulin (pancreatic extracts) in the treatment of diabetes mellitus. *Br. Med. J.* **1923,** (I), 8-12.

11. del C. Dominguez, M.; Lorenzo, N.; Barbera, A.; Padron, G.; Torres, A. M.; Hernandez, M. V.; Hernandez, I.; Gil, R.; Sanchez, A.; Besada, V.; Gonzalez, L. J.; Garay, H.; Reyes, O.; Perez, E.; Lopez, M.; Mazola, Y.; Cosme, K.; Ancizar, J., Therapeutic effect of two altered peptide ligands derived from the human heat shock protein 60 in experimental models of rheumatoid arthritis. *Biotecnol. Apl.* **2013,** *30* (2), 153-156.

12. Mansour, S. C.; Pena, O. M.; Hancock, R. E. W., Host defense peptides: front-line immunomodulators. *Trends Immunol.* **2014,** *35* (9), 443-450.

13. Padhi, A.; Sengupta, M.; Sengupta, S.; Roehm, K. H.; Sonawane, A., Antimicrobial peptides and proteins in mycobacterial therapy: Current status and future prospects. *Tuberculosis (Oxford, U. K.)* **2014,** *94* (4), 363-373.

14. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D., The future of peptide-based drugs. *Chem. Biol. Drug Des.* **2013,** *81* (1), 136-147.

15. Kaspar, A. A.; Reichert, J. M., Future directions for peptide therapeutics development. *Drug Discovery Today* **2013,** *18* (17-18), 807-817.

16. Fosgerau, K.; Hoffmann, T., Peptide therapeutics: current status and future directions. *Drug Discovery Today* **2015,** *20* (1), 122-128.

17. Itakura, K.; Hirose, T.; Crea, R.; Riggs, A. D.; Heyneker, H. L.; Bolivar, F.; Boyer, H. W., Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science* **1977,** *198* (4321), 1056-63.

18. Li, Y., Recombinant production of antimicrobial peptides in Escherichia coli: A review. *Protein Expression Purif.* **2011,** *80* (2), 260-267.

19. Kim, H.; Jang, J. H.; Kim, S. C.; Cho, J. H., De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *J. Antimicrob. Chemother.* **2014,** *69* (1), 121-132.

20. Tang, W.; Sun, Z.-Y.; Pannell, R.; Gurewich, V.; Liu, J.-N., An efficient system for production of recombinant urokinase-type plasminogen activator. *Protein Expression Purif.* **1997,** *11* (3), 279-283.

21. Sun, Q.-M.; Chen, L.-L.; Cao, L.; Fang, L.; Chen, C.; Hua, Z.-C., An Improved Strategy for High-Level Production of Human Vasostatin 120-180. *Biotechnol. Prog.* **2005,** *21* (4), 1048-1052.

22. Chen, H.; Xu, Z.; Xu, N.; Cen, P., Efficient production of a soluble fusion protein containing human beta-defensin-2 in E. coli cell-free system. *J. Biotechnol.* **2005,** *115* (3), 307-315.

23. Nallamsetty, S.; Waugh, D. S., Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expression Purif.* **2006,** *45* (1), 175-182.

24. du Vigneaud, V.; Ressler, C.; Swan, J. M.; Roberts, C. W.; Katsoyannis, P. G., The synthesis of oxytocin. *J. Am. Chem. Soc.* **1954,** *76*, 3115-21.

25. Merrifield, R. B., Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963,** *85* (14), 2149-54.

26. Kent, S. B. H., Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **1988,** *57*, 957-90.

27. Pennington, M. W.; Byrnes, M. E., Procedures to improve difficult couplings. *Methods Mol. Biol. (Totowa, N. J.)* **1994,** *35* (PEPTIDE SYNTHESIS PROTOCOLS), 1-16.

28. Rodriguez, V.; Asenjo, J. A.; Andrews, B. A., Design and implementation of a high yield production system for recombinant expression of peptides. *Microb. Cell Fact.* **2014,** *13*, 65/1-65/10, 10 pp.

29. Yin, L. M.; Edwards, M. A.; Li, J.; Yip, C. M.; Deber, C. M., Roles of Hydrophobicity and Charge Distribution of Cationic Antimicrobial Peptides in Peptide-Membrane Interactions. *J. Biol. Chem.* **2012,** *287* (10), 7738-7745.

30. Andersson, L.; Blomberg, L.; Flegel, M.; Lepsa, L.; Nilsson, B.; Verlander, M., Large-scale synthesis of peptides. *Biopolymers* **2000,** *55* (3), 227-250.

31. Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R., An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. *J. Biomol. NMR* **1998,** *11* (1), 97-102.

32. Harper, S.; Speicher, D. W., Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2011,** *681* (Protein Chromatography), 259-280.

33. Mitchell, D. A.; Marshall, T. K.; Deschenes, R. J., Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. *Yeast* **1993,** *9* (7), 715-22.

34. Medina, D.; Moskowitz, N.; Khan, S.; Christopher, S.; Germino, J., Rapid purification of protein complexes from mammalian cells. *Nucleic Acids Res.* **2000,** *28* (12), e61, ii-viii.

35. Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J., Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression Purif.* **2006,** *48* (1), 1-13.

36. Dyson, M. R.; Shadbolt, S. P.; Vincent, K. J.; Perera, R. L.; McCafferty, J., Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **2004,** *4*, No pp given.

37. Smith, D. B.; Johnson, K. S., Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988,** *67* (1), 31-40.

38. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London, U. K.)* **1970,** *227* (5259), 680-685.

39. Scopes, R. K., Measurement of protein by spectrophotometry at 205 nm. *Anal Biochem* **1974,** *59* (1), 277-82.

40. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein identification and analysis tools on the ExPASy server. *Proteomics Protoc. Handb.* **2005**, 571-607.

41. Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; Dirr, H., Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **1997,** *6* (2), 399-406.

42. Kathir, K. M.; Rajalingam, D.; Sivaraja, V.; Kight, A.; Goforth, R. L.; Yu, C.; Henry, R.; Kumar, T. K. S., Assembly of Chloroplast Signal Recognition Particle Involves Structural Rearrangement in cpSRP43. *J. Mol. Biol.* **2008,** *381* (1), 49-60.

43. Jayanthi, S.; Kachel, B.; Morris, J.; Kumar, T., Molecular Cloning and Overexpression of WAP domain of Anosmin-1 in *Escherichia coli*. *Molecular Cloning – Selected Applications in Medicine and Biology* **2011,** *4*, 1-15.

44. Waugh, D. S., An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression Purif.* **2011,** *80* (2), 283-293.

3. Application(s) of GST-affinity tag in NMR

3.1. Abstract

With the advances in biological sciences, there is a consistent demand for structural information of biologically active polypeptides. High throughput screenings are necessary in fields such as proteomics, genomics, and bioinformatics as they provide valuable clues on proteins, which can have therapeutic, diagnostic, and industrial applications. In this context, recombinant protein expression is a good approach to obtain large amounts of the target protein and thus has become a commonly used way of production. A very commonly used affinity tag is Glutathione S-Transferase (GST, 26kDa), which is known to increase expression yields by enhancing the solubility of the protein of interest and therefore making it a valuable tool for the purification of recombinant proteins. Multidimensional NMR spectroscopy is a popular technique to elucidate the 3D structure of proteins in solution. However, obtaining the structural information of peptides and small proteins can be challenging. In this study, we show that multidimensional NMR data can be successfully acquired on recombinant proteins even without removing the GST-affinity tag. Our results show that the GST-affinity tag does not appear to have an effect on the quality of NMR data of its associated recombinant target protein. It is well known that GST isozymes exist as dimers, but there has been little research on the oligomeric state of GST-fused proteins. Our results also suggest that the GST-tag and the fused partner can be observed as two separate entities in multidimensional NMR spectra. Furthermore, small angle X-ray scattering (SAXS) is employed to study the low-resolution structure and flexibility of GST- fusion proteins. The results of the SAXS experiment support that GST-fused proteins predominantly exist as dimers in solution. We predict the loss of signals in the ^{1}H-^{15}N HSQC

spectrum corresponding to the GST-tag is primarily due to the decrease in the T2 relaxation rate upon the symmetric dimerization of GST. Additionally, the six residues located between the GST-tag and the target protein, which represent the recognition site for the enzymatic thrombin cleavage, act as a flexible linker and may play an important role in this observation. Furthermore, we were able to optimize the effects of the thrombin linker by introducing 12 glycine residues, which increased the flexibility between the GST-affinity tag and the protein of interest. As a result we were able to obtain better quality NMR data and are optimistic that these findings can be used to extend this application to larger proteins.

3.2. Introduction

Glutathione S-transferase represents a critical component in the phase II detoxification of xenobiotic agents, including environmental toxins and therapeutic drugs. According to Armstrong et al, GST acts as one of the most essential enzymes in the removal of electrophilic toxins in animals, plant, and many microorganisms[1]. The family of GST includes isozymes in prokaryotes and eukaryotes. However, each species' GSTs are categorized separately, for example in case of human cytosolic GSTs, Greek letters are used to classify them. Crystal structures are accessible for each group of human cytosolic GST and show that the 3D conformation is homologous among these different classes of human GST[2]. In the late eighties, GST derived from the helminthic parasite *Schistosoma japonicum* (Sj26GST), which is categorized as a mammalian class mu GST, has been found to be an efficient affinity tag for the expression and purification of recombinant proteins[3,4,5]. One of its main advantages is the high solubility of GST, which in-turn is expected to be extend to the fused target protein[6]. Furthermore, GST is not toxic for the host and can be isolated using an easy one-step purification

procedure, thus this protein is one of the most commonly used affinity tags[7,8]. In comparison to the Maltose-binding-protein-tag and the polyhistidine-tag, Dyson et al showed that among 32 different target proteins with molecular weights ranging from 17-110 kDa, GST-fusion proteins generated the largest amount of soluble protein[9].

Another important characteristic of Sj26GST, along with soluble GSTs from other classes, is the formation of dimers. The dimerization has been shown to be important for GST's enzyme activity[5,10]. McTigue et al and Lim et al solved the crystal structures for Sj26GST in the absence and in the presence of its substrate glutathione, respectively[11,12]. The interface of the two GST monomers has been described to be comprised of a GST specific "lock-and-key" type interactions in addition to hydrophobic contacts that are stabilized by several salt bridges and electrostatic interactions[11,12,13,14]. Mutational studies performed by Sayed et al and Hornby et al suggest though that the phenylalanine that has been indicated to be crucial for the "lock-and-key" type interaction is more critical for the tertiary structure than the dimerization process[15]. Moreover, Abdalla et al's results demonstrate that rather 10 site-specific mutations are necessary for preventing the dimer formation of pi class GSTP1-1[16]. In addition, Dirr and Reinemer discovered that the dimerization of class pi GST is advantageous because of the increased thermostability of the enzyme. Their findings also demonstrated that the separation of the GST dimer and the unfolding of the protein are intricately connected[17]. Erhardt and Dirr's results also suggest the absence of a folded monomer intermediate and therefore the researchers advocate the direct transition from a folded dimer to unfolded monomers[18]. On the other hand, the debate of the conformation state in the transition of dimers to monomers is continued with experimental data presented by Aceto et al. In stead of a direct transition they propose a multi-step process based on their studies on pi-class GSTs[19]. Aceto et al revealed that at low concentrations of

detergent the GSTP1-1 dimer separates into enzymatically inactive monomers. Fabrini et al also support the presence of a folded intermediate[2].

Studying GST-fusion proteins, Lally et al found that GST also dimerizes when it is fused to a peptide[20]. By using electron microscopy it was shown that the attached peptide, which was subject of their analysis, is extended away from the GST dimer. Furthermore, when Lim et al acquired crystal structures for a GST-fusion peptide, they experienced high temperature factors for the residues representing the thrombin cleavage recognition side that was placed between the affinity tag and the peptide of interest. The increase of this parameter indicates higher mobility of the six-residue peptide while attached to the C-terminus of GST, leading other researchers to remove the GST affinity tag when growing crystals in order to avoid these inter-domain movements[21]. Nevertheless, in GST-pull down assays, Vikis Harris and Guan portrayed how essential this flexibility between the affinity tag and the protein of interest is. It ensures that the GST-tag does not interfere with the fusion partner's ability to interact with its substrate[7]. Both research groups attribute this capability to the extendable linker region at the C-terminus of GST.

In our studies, we confirm the observation of the loss of cross-peaks corresponding to the structure of GST in the fusion protein that was reported by Liew and colleagues[22]. We further explored this phenomenon by employing multidimensional NMR spectroscopy, size exclusion chromatography, and small angle X-ray scattering. Experiments performed by Liew et al already indicated the loss of NMR signals corresponding to GST due to the dimerization of GST and the resulting increase in T2 relaxation times for GST[22]. In addition, we hypothesize that the 6 amino acids, representing the thrombin cleavage recognition site and located between the affinity tag and the protein of interest, is imperative for the phenomenon as they act as a flexible linker. This allows the fusion protein to be seen as two separate entities, therefore not affecting each other's

3D solution structure. In the following, GST-CD2 (CD2 6kDa), GST-CD2CD3 (CD2CD3 11 kDa), and the control GST were investigated to elucidate the rationale of the circumstance of disappearing GST peaks in the ^{1}H-^{15}N HSQC spectrum of the fusion protein.

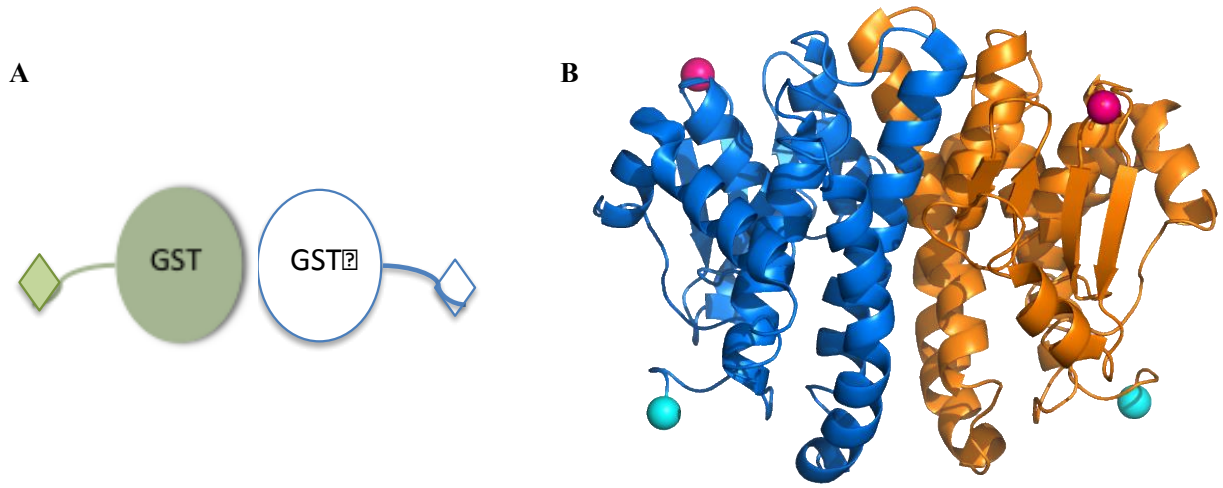


Figure 1: A) Schematic illustration explaining the rationale for the non-appearance of the ^{1}H-^{15}N cross-peaks representing the amide protons of the GST-affinity tag. B) Pymol illustration depicting GST-dimer (PDB: 1Y6E).

As described by Liew et al, a limit for the feasibility of this phenomenon can be expected in terms of molecular weight of the fusion partner. Larger proteins are more likely to interact with the residues in the linker region or with GST, which can result in a different 3D solution structure or the disappearance of such residues along with GST's cross-peaks[22]. In order to extend the applicability of this method to larger proteins, we introduced 12 additional glycine residues after the thrombin linker region. We expect that the now more extended thrombin linker increases the flexibility of the fusion protein and allows us to obtain structural information on a wider range of proteins without having to remove the affinity tag.

3.3. Materials and Methods

Expression and Purification of GST-CD2 and GST-CD2CD3

LB broth (Miller; EMD Millipore, MA) containing ampicillin (J.T. Baker Chemicals, PA, 100 μg/ml) was inoculated with 5% (v/v) overnight culture under sterile conditions and incubated at 37 °C and 250 rpm. When the OD_{600} of 0.6 was reached, the cells were induced with 1mM isopropyl-1-thio-β-D-galactopyrannoside (IPTG, OMNI Chemicals, IN), and further incubated for four hours. Later, the cells were harvested at 6,000 rpm for 20 minutes at 4 °C using a Beckman JA-10 rotor. The collected pellets were washed using 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4; pH 7.2; J.T. Baker Chemicals) and were used immediately or stored at -20 °C.

E.coli BL21 (DE3) cells containing the expressed recombinant fusion protein(s) were resuspended in 25 ml 1xPBS pH 7.2 and lysed by sonication. Cell debris was removed by centrifugation at 19,000 rpm for 30 minutes. The supernatant containing the soluble GST-CD2 or GST-CD2CD3 was loaded onto a pre-equilibrated GSH-Sepharose column at a flow rate of 1 ml/min. Subsequently, unbound *E.coli* protein contaminants were removed by washing the column with 1x PBS until a flat baseline was reached. The fusion protein was eluted with 10 mM reduced glutathione (Sigma Aldrich, MO) dissolved in 1x PBS. Glutathione had to be removed for the following structure elucidation studies by concentrating and buffer exchanging using an Amicon concentrator (EMD Millipore, MA). Protein concentration of the pure fusion protein was evaluated by measuring the absorbance at 280 nm and its specific molar extinction coefficient. Samples obtained while monitoring the purification were resolved on 15% SDS–PAGEs under reduced conditions according to the method of Laemmli[23].

Cleavage and separation

The pure fusion partner CD2 or CD2CD3 was obtained by off column thrombin cleavage. Complete cleavage was standardized in previous studies to 1U for every 0.25 mg of fusion protein. The GST-tag was removed from the cleavage product mixture by reloading on a pre-equilibrated GSH-Sepharose column. While the GST-tag binds to the resin, the small proteins are found in the flow through fraction, which subsequently was concentrated using an Amicon concentrator. Protein purity was monitored by SDS-PAGE and the target proteins concentrations were calculated using their absorbance at 280 nm and specific molar extinction coefficients.

Acquisition of ^{1}H-^{15}N HSQC spectra

Multidimensional nuclear magnetic resonance spectroscopy (NMR) is a technique capable of elucidating the 3D structure and backbone dynamics of a protein at atomic resolution. This technique allows comparing the 3D solution structure of the fusion proteins and just the target molecules. ^{1}H-^{15}N HSQC spectra were acquired using ^{15}N isotope enriched proteins. The concentration of the protein(s) was in the range of 0.4-0.5 mM, in 1x PBS buffer (90% H_2O + 10% D_2O, pH 7.2). All experiments were conducted at 298 K using the Bruker Avance 700 MHz or Bruker 500 MHz NMR spectrometer. The chemical shifts of the assigned peaks of CD2 were visualized by using the ^{1}H-^{15}N chemical shift perturbation plot.

Size Exclusion Chromatography of GST-CD2

Gel filtration experiments were performed using an AKTA FPLC on a Superdex 75 column (GE Healthcare, Pittsburgh, PA) in a running buffer of 1x PBS, pH 7.2, and at a flow of 1 ml/min and 22 °C. The protein peak of GST-CD2 was detected by its 280 nm absorbance.

Under the experimental conditions used, no shrinkage of the resin was observed. A standard plot of the logarithm of the molecular mass versus the elution time was constructed using standard proteins in the molecular mass range of 17−66 kDa. Experimental conditions used for the elution of standard proteins were the same as those for GST-CD2.

Analysis of Small angle X-ray scattering (SAXS) data acquired for GST, GST-CD2, and GST-CD2CD3

A dilution series of GST, GST-CD2, and GST-CD2CD3 was created in the range of 10 mg/ml to 1 mg/ml. The Cornell High Energy Synchrotron Source (CHESS) was used as the source of high-energy X-rays and hutch G1 was set up with a robotic sampling unit[24]. For each sample, 10 spectra were acquired by exposing for 2 seconds and scattering was collected while oscillating samples to reduce X-ray damage. The spectra were averaged and blank samples of 1x PBS buffer were subtracted from each sample. The concentration of each sample was measured before beam exposure and using a Guinier fit [I (q) *versus* q^2], it was possible to determine aggregation and estimate the radius of gyration for each sample. The distance distribution was calculated using the Primus program from the ATSAS package (Europäisches Labor für Molekularbiologie, Hamburg)[25]. In the Primus shape wizard, the DAMMIF algorithm was employed to predict 10 envelope structures for GST[26]. The computed envelopes were then aligned to overlap using the program set DAMAVER to compare and test the similarity of the structures[27]. Of the 10 envelopes predicted for GST, 9 were predicted to be probable. In the following, DAMMIN refined the shape of the model through simulated annealing using a single-phase dummy atom model[28]. In case of the fusion proteins GST-CD2 and GST-CD2CD3 an all atom program, developed by a member of the Kumar lab, was used due to their flexible

character. 10,000 random conformers of the protein of interest were built and sub-ensembles of isomers co-occurring in solution were chosen based on their fit to the experimental SAXS data. The optimized ensembles were then compared to the pool of 10,000 random structures in a size-distribution plot. Furthermore, the molecular weight was determined using the software RAW[24] based on the scattering intensity and the measured concentrations of the analyzed samples, and the standard lysozyme of known concentration (c= 4.12 mg/mL).

SAXS data can provide several indicators for the presence of flexibility within a protein. Customarily, the Kratky plot gives a qualitative assessment of disordered states within a protein and is able to distinguish them from globular, compact proteins[29]. The Kratky illustration is a transformation of the scattering profile (q^2 x I (q) as a function of q) that allows an easier visualization of the degree of flexibility within a protein. Another tool to detect flexibility within biopolymers and macromolecules is the Porod-Debye Law[30]. Here, the scattering data is transformed as q^4 x I (q) *vs*. q or q^4 x I (q) *vs*. q^4, which should display a curve asymptotically approaching a constant value as q approaches infinity for globular, compact proteins. Moreover, the scattering data was transformed using an indirect Fourier transformation in PRIMUS to obtain the distance distribution function[31]. It is defined to be a positive curve that ends at the maximum linear dimension in the scattering particle (D_{max}) and therefore to equal 0 at p(0) and p(r>D_{max}).

PCR-based cloning of GST-G_{12}-CD2CD3

The pGEX-KG vector (GE Healthcare) was used for the expression of GST-G_{12}-CD2CD3. The fusion protein can be cleaved with thrombin (cleavage sequence Leu-Val-Pro-Arg | Gly-Ser). The gene of interest G_{12}-CD2CD3 was amplified by using gene specific primers (FP: 5'ATGCACGGATCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTCAAG

TGTTCGA–3', BP: 5'ATGCACCTCGAGTCGACCCGGccCTATTCATTCATTGGTTGTTGT TGTTGGTAGA-3') and the Taq polymerase (NEB). The conditions for the PCR were as described in the vendor's protocol at an annealing temperature of 69 °C. The PCR product was monitored by agarose gel electrophoresis and purified from unused nucleotides, primers, and polymerase using the QIAGEN PCR purification Kit (QIAGEN, CA) according to the manufacturer's instructions. The pGEX-KG vector and PCR product were subjected to double digestion using the restriction enzymes *BamHI* and *XhoI* (NEB). The double digested products were purified using the QIAGEN PCR purification Kit, ligated at a molar ratio of 8:1 (insert: vector) and transformed into DH5α chemical competent cells (NEB). Recombinant plasmids were isolated from bacterial colonies using the QIAGEN Miniprep Kit and subjected to both DNA sequencing and restriction analysis to confirm its identity.

Expression, purification, and analysis of GST-G_{12}-CD2CD3

Recombinant plasmids containing the gene of interest were transformed into *E.coli* BL21 (DE3) cells. The expression, purification, and NMR analysis was performed in a similar fashion as described earlier. The SAXS data was acquired at different protein concentrations (1.25 mg/ml -10 mg/ml) at the Sybils Beamline, Lawrence Berkeley National Lab, CA. This work was conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project MINOS (R01GM105404).

3.4. Results

Purification of GST-CD2 and CD2

All fusion proteins (GST-CD2, GST-CD2CD3, GST-G_{12}-CD2CD3) as well as CD2 and CD2CD3 were purified to homogeneity using Glutathione- Sepharose affinity chromatography (Figure 2). GST-CD2, which was recovered from the affinity column (Figure 2, lane 5), was subjected to thrombin cleavage. It was observed in earlier studies that 1U of thrombin is sufficient to effectively cleave 0.25mg of the fusion protein. The purified fusion protein (Figure 2, lane 5 and target protein (Figure 2, lane 9) migrated as a single band according to their expected molecular masses on a SDS-PAGE under reducing conditions.

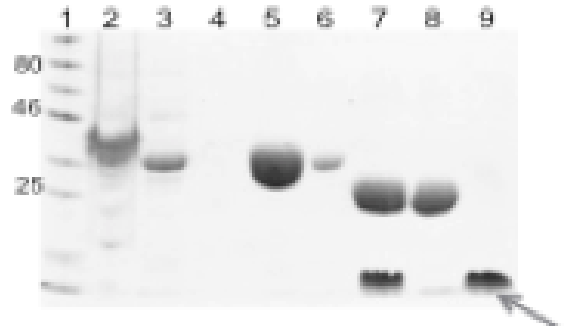


Figure 2: SDS-PAGE depicting the purification of GST-CD2 and CD2: Lane-1 pre-stained protein marker, Lane-2 pellet after lysis, Lane-3 supernatant after lysis, Lane-4 flow-through, Lane-5 GST-CD2, Lane-6 8M urea, Lane-7 cleaved GST and CD2, Lane-8 GST, Lane-9 CD2.

^{1}H-^{15}N HSQC spectra of GST-CD2 and CD2

Multidimensional NMR experiments were performed to elucidate the 3D solution structure of GST-CD2 and the target protein of interest CD2. Two-dimensional ^{1}H-^{15}N HSQC spectra yield a fingerprint of the backbone conformation of proteins. Each cross- peak in a ^{1}H-^{15}N HSQC spectrum represents an amino acid in a particular backbone conformation of the protein. The ^{1}H-^{15}N HSQC spectra of GST-CD2 and CD2 are interesting for two reasons. First,

the peaks found in the ^{1}H-^{15}N HSQC of GST-CD2 overlay well on the peaks of CD2 (Figure 2A). Careful inspection of the spectra revealed that the ^{1}H-^{15}N chemical shift perturbations are insignificant (Figure 2B). Therefore, the 3D solution structure of CD2 is the same whether it is acquired individually or as fusion protein. Second, having the peaks of the GST-CD2 spectrum identified as residues corresponding to CD2 means in turn that the cross-peaks corresponding to GST have disappeared.

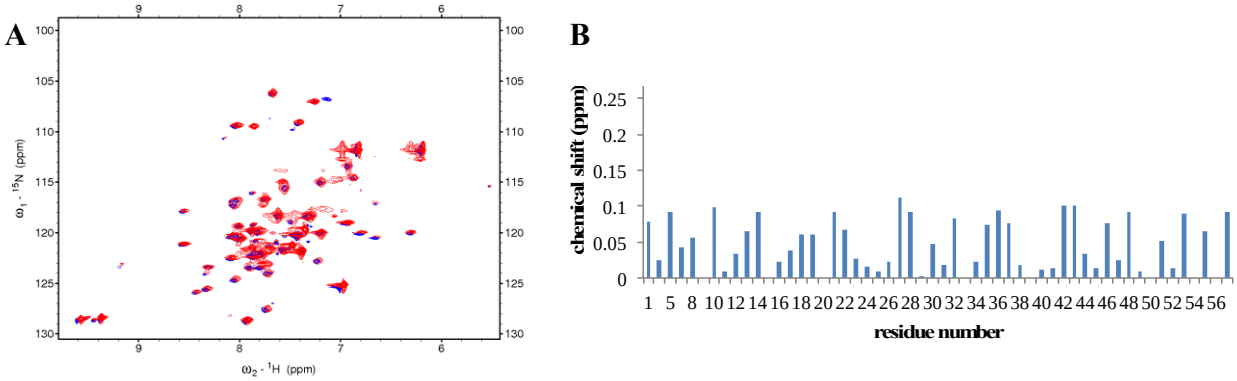


Figure 3: Panel A: Overlay of ^{1}H-^{15}N HSQC spectra of GST-CD2 (red) and recombinant CD2 (blue); Panel B: ^{1}H-^{15}N chemical shift perturbation plot of CD2.

GST-fusion proteins form dimers

Both gel filtration chromatography and SAXS analysis were employed to confirm the multimeric state of GST and its fusion proteins in solution. When comparing the elution time of GST-CD2 (Figure 4) to the molecular weight standard proteins during the size exclusion chromatography it can be inferred that GST-CD2 forms a dimer in solution. In consummation with these findings, the SAXS results in Table 1 are in agreement with the results of the gel filtration experiments. Both these data suggest that the proteins have a molar mass that correlates with the size of a dimer (theoretical molecular mass of the monomers of GST, GST-CD2 and

GST-CD2CD3 are 26 kDa, 32 kDa, and 37 kDa respectively). It needs to be mentioned that the accuracy of determination of molecular weights using SAXS data lies within a systematic deviation error of 10%[32].

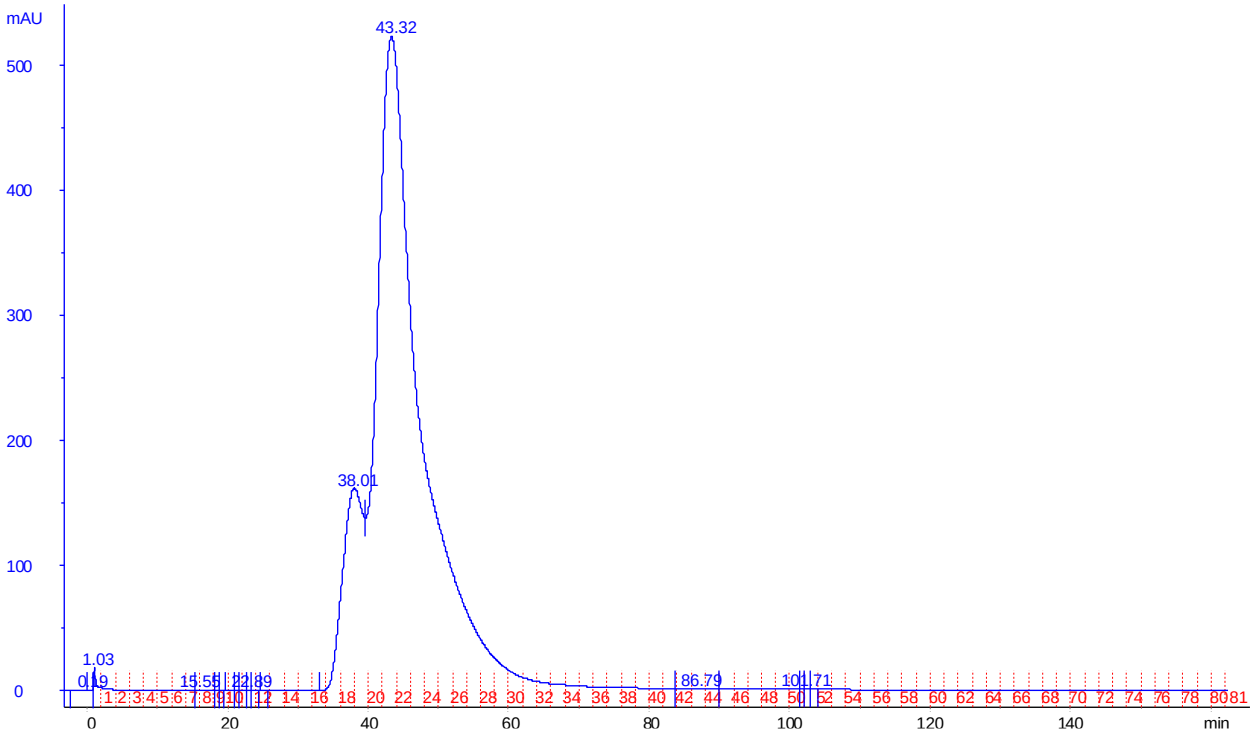


Figure 4: Size-Exclusion Chromatogram of GST-CD2.

Table 1 Estimated molecular mass obtained from SAXS data

Sample	**Theoretical molecular weight**	**Experimental molecular weight**
GST	26 kDa	57 kDa
GST-CD2	32 kDa	70 kDa
GST-CD2CD3	37 kDa	76 kDa

GST-fusion proteins are flexible due to thrombin linker

The normalized Kratky plots of GST, GST-CD2, and GST-CD2CD3 are depicted (Figure 5). The curve representing GST displays a typical bell-shape characteristic for this globular protein. For GST-CD2 and GST-CD2CD3, the peak amplitude does not decrease and the bell-shape is mostly intact, indicating that there are folded portions of the protein present. The folded portions are attributed to GST and CD2/CD2CD3 based on the acquired ^{1}H-^{15}N HSQC spectra. Only when GST is properly folded, it dimerizes resulting in the loss of cross-peaks in the ^{1}H-^{15}N HSQC spectrums. In a similar fashion CD2 shows its native conformation state in GST-CD2 as the cross peaks of the ^{1}H-^{15}N HSQC spectrum of GST-CD2 superimpose well on the cross-peaks of the ^{1}H-^{15}N HSQC of CD2 (Figure 3A). Furthermore, when comparing the normalized Kratky plot of GST and the GST-fusion proteins, the fusion proteins are identified to have more flexibility than the individual GST-protein. The Kratky plot of the GST-fusion proteins exhibits a broadened bell-shape curve as well as a plateau at $q \times Rg > 5$ instead of reaching to the x-axis of the plot. We believe the flexibility stems from the thrombin recognition site that is positioned between GST and the target protein. These results are in agreement with observations made based on the GST-pull down assays of other fusion proteins[7]. CD2 is fused to the C-terminus of GST with the thrombin recognition site (linker) introducing enough flexibility to view the two partners as separate physical entities. In addition to the thrombin linker, the fusion partner is attached to a flexible loop region that is found at the C-terminus of GST. This loop can also contribute to the inter-domain flexibility as it can be extended[21]. However, it was shown that without a fusion partner, the loop is folded more compactly[11].

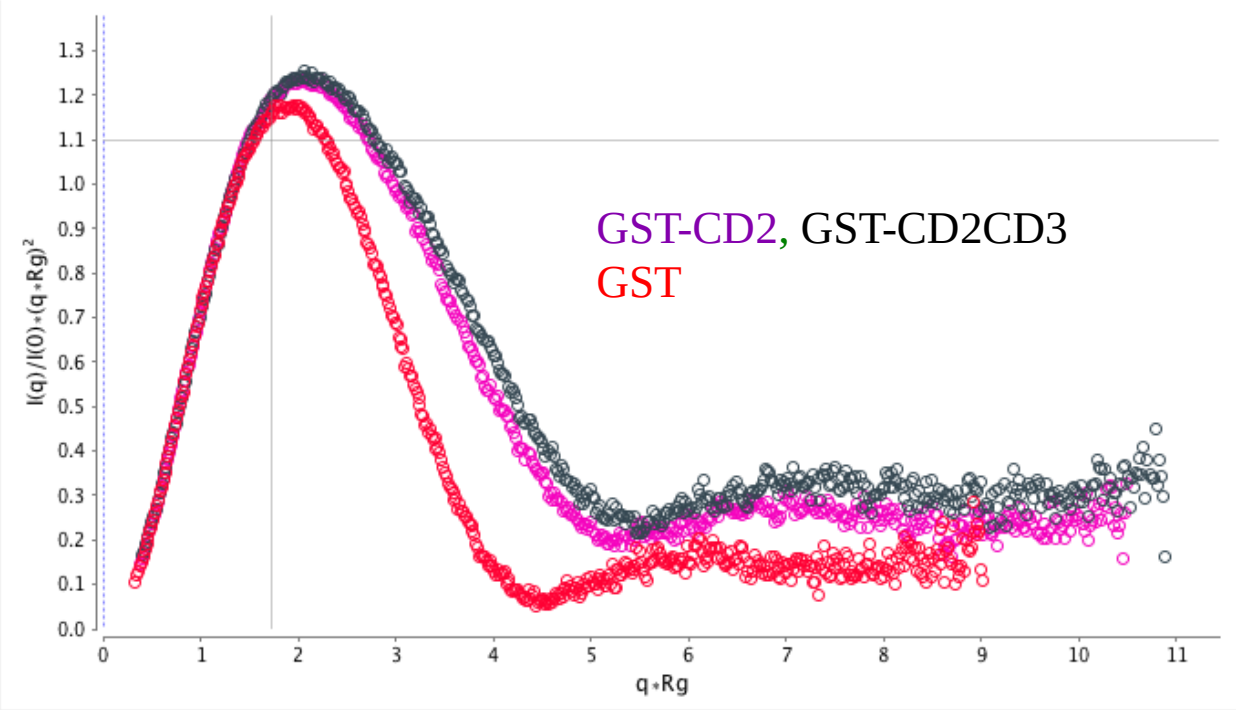


Figure 5: Dimensionless Kratky- Plot of GST, GST-CD2, and GST-CD2CD3.

In comparison, the Porod-Debye law is another useful tool for revealing flexibility within macromolecules from SAXS data. Rambo and Tainer claim it to be more powerful and conclusive than the Kratky analysis, especially when comparing protein flexibilities and needing to look for more confined flexibility[30]. A globular protein demonstrates a plateau when transforming the scattering profile to q^4 x I(q) *vs*. q^4, while fully flexible particles will show a characteristic plateau when the SAXS data is converted to q^2 x I(q) *vs*. q^2. The Porod exponents for GST, GST-CD2, and GST-CD2CD3 are calculated and are listed in Table 2. GST forms a compact, globular dimer, giving it a characteristic Porod exponent of almost 4. Both fusion proteins, however, have a lower Porod value but not as low as 2, which would be indicative for an intrinsically disordered protein. Therefore, it can be concluded that the decreased Porod exponent of GST-CD2 and GST-CD2CD3 reveals local flexibility within the fusion protein,

which can be attributed to the short linker sequence between the affinity tag and the protein of interest.

Table 2 Porod- Exponents for GST, GST-CD2, and GST-CD2CD3

Sample	**Porod- Exponent**
GST	3.7
GST-CD2	3.1
GST-CD2CD3	3.2

In a further attempt to analyze the SAXS data, a low-resolution model of GST-CD2 was generated based on the scattering profile and using an all-atom algorithm. Out of a pool of 10,000 possible structures, 4 were highlighted as very likely based on angular and distance constraints and their fit to the experimental SAXS scattering. Both the pair distance distribution plot (Figure 6) and the Pymol illustration of the most probable *ab initio* models (Figure 7A, C) are in agreement that CD2 is predominantly extended away from the GST dimer. The shape of the distance distribution plot can also be used as an indicator for the structural properties of the sample[31]. In particular, globular compact particles exhibit a symmetrical bell shaped curve, whereas unfolded particles have a stretched tail. In case of GST, the pair distance distribution function is consistent for a globular protein, while GST-CD2 still shows features of the compact folded GST and CD2 but also displays a tail that could be accredited to the extended conformation and the flexibility because of the thrombin linker (Figure 6).

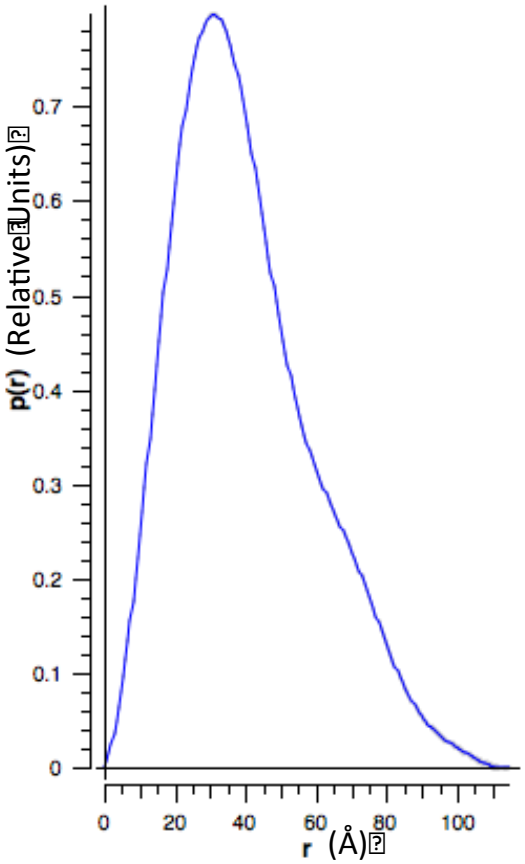
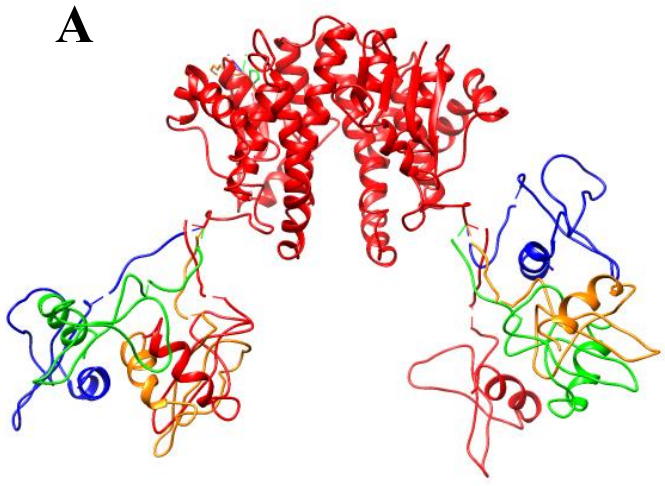


Figure 6: Pair distance distribution plot of GST-CD2.



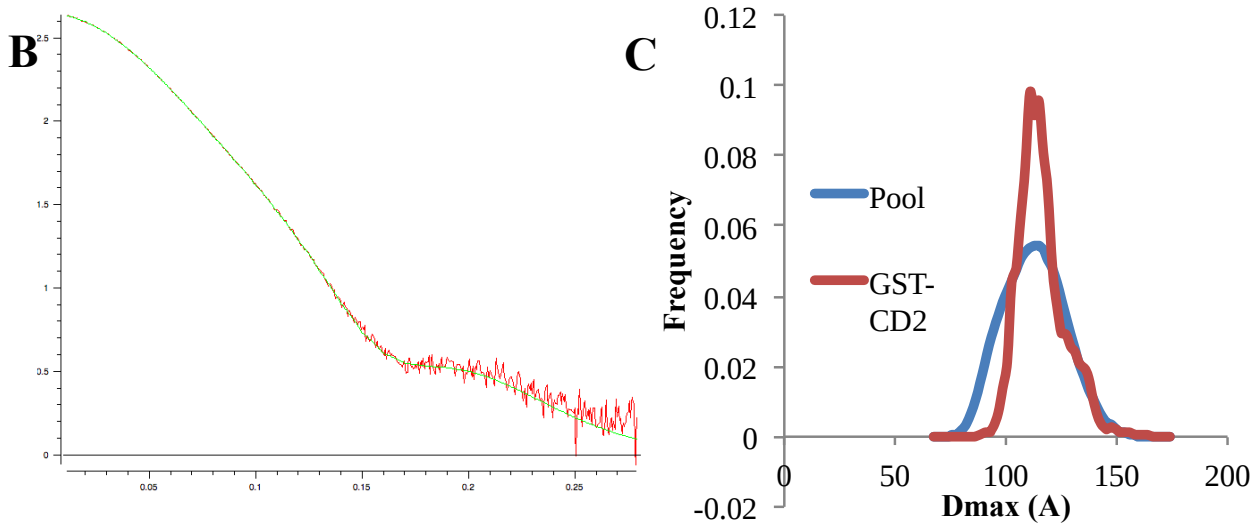


Figure 7: Panel A: Ensemble model of GST-CD2 obtained from SAXS data. Panel B: Calculated SAXS scattering profile (green) of GST-CD2 model fits the experimental SAXS data (red). Panel C: Size-distance distribution of comparing Dmax of random pool and ensemble models for GST-CD2.

Engineering of a more flexible linker

When comparing the ^{1}H-^{15}N HSQC spectra of GST-CD2CD3 and CD2CD3 (Figure 8), one would expect the same overlay as it was observed for GST-CD2 and CD2. However, the fingerprint of the backbone confirmation is different for the fusion protein GST-CD2CD3 and the individual CD2CD3. It is possible that the target protein interacts with the linker or the GST dimer with the result of shifted peaks or loss thereof in the ^{1}H-^{15}N HSQC of GST-CD2CD3. Nevertheless, this result is surprising when viewing the Kratky plot and Porod exponent of GST-CD2CD3 (Figure 5a, Table 2). The results of the SAXS analysis of GST-CD2CD3 and GST-CD2 are quite similar. Therefore, one would expect that the peaks corresponding to GST in the ^{1}H-^{15}N HSQC spectrum of GST-CD2CD3 should disappear and the peaks resulting from CD2CD3 should overlay with the backbone confirmation of the acquisition of the ^{1}H-^{15}N HSQC of CD2CD3. When taking the ensemble models of GST-CD2CD3 and their size distribution plot,

which compares the maximal diameter found in the ensemble models and the pool of 10000 random structures, into consideration, both indicate two states. Especially in case of the size distribution plot, the two maxima indicate an enclosed and a more extended conformation (Figure 9C). The *ab initio* model (Figure 9A) suggests that the fused partner CD2CD3 is often located in close proximity to the GST dimer, which could lead to interactions between CD2CD3 and GST and therefore to the loss of signal for those CD2CD3 residues in the ^{1}H-^{15}N HSQC or the association with the linker region which could explain the shift of peaks for some residues. In an attempt to increase the inter-domain flexibility of the fusion partner, 12 glycine residues were introduced after the recognition site for the enzymatic thrombin cleavage. It is also assumed that this extension of the linker will lead to CD2CD3 predominantly protruding away from GST. Therefore, there should be lesser interactions of the protein of interest with the affinity tag or the linker allowing the acquisition of the ^{1}H-^{15}N HSQC spectrum of the native conformation of the target protein while still fused to GST. Through PCR-based sub-cloning, 12 additional glycine residues were incorporated following the thrombin cleavage recognition site and upstream of the protein of interest. The yield and purification method is similar to the other GST-fusion proteins as described earlier. Interestingly, the ^{1}H-^{15}N HSQC of GST-G_{12}-CD2CD3 overlays better with CD2CD3 than GST-CD2CD3 did in earlier studies (Figure 10A). With the new fusion protein clone, there are no peaks missing when comparing the ^{1}H-^{15}N HSQC spectrum of GST-G_{12}-CD2CD3 to the one of CD2CD3. The amount of shifted peaks that was observed in the spectrum of GST-CD2CD3 is also less when acquiring the HSQC spectrum of GST-G12-CD2CD3. The few additional cross-peaks of GST-G12-CD2CD3 are located in the center and might resemble residues from the extended linker. Moreover, an increased flexibility for the engineered GST-

G_{12}-CD2CD3 clone was observed in the Kratky plot (Figure 10B) and its Porod exponent was calculated to be 3.1 both supporting the hypothesis.

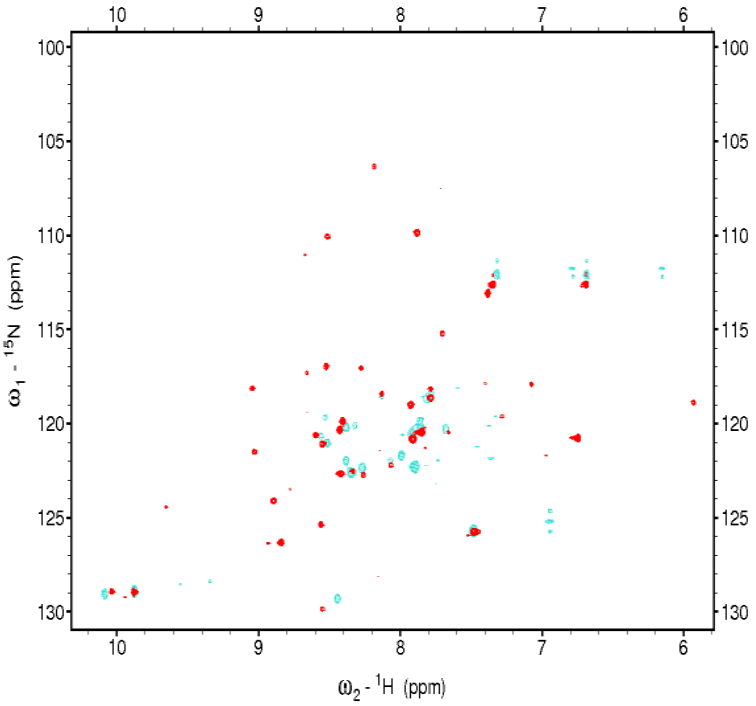
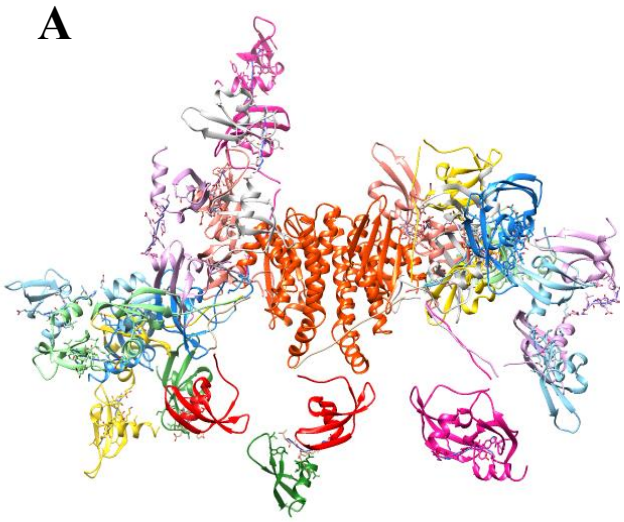


Figure 8: Overlay of ^{1}H-^{15}N HSQC spectra of GST-CD2CD3 (red) and recombinant CD2CD3 (turquois).



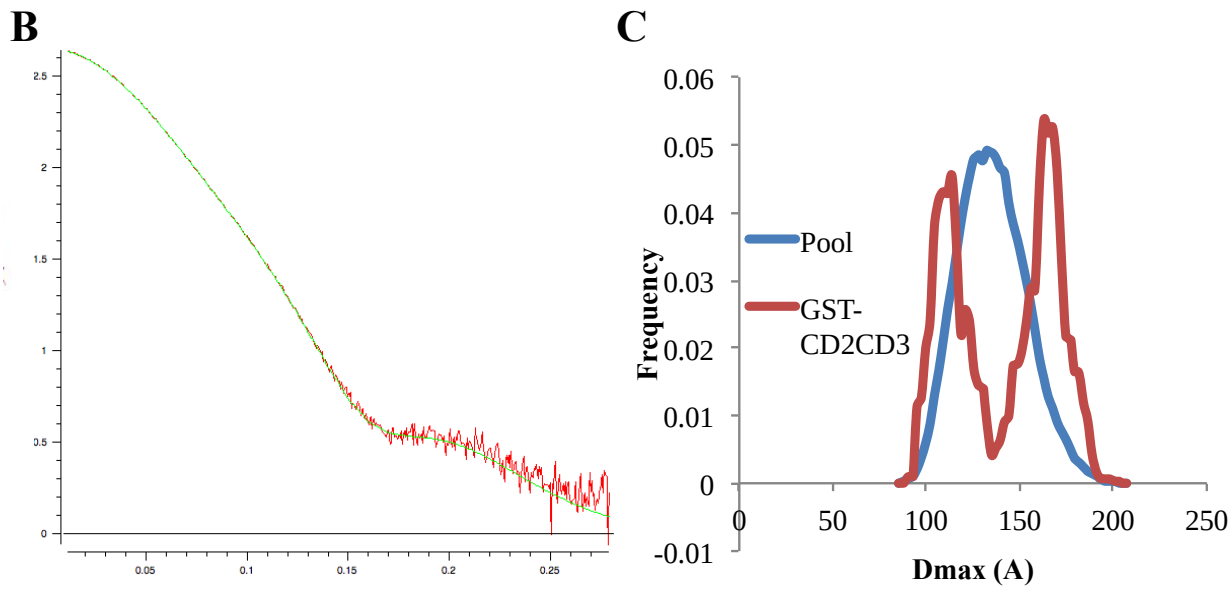
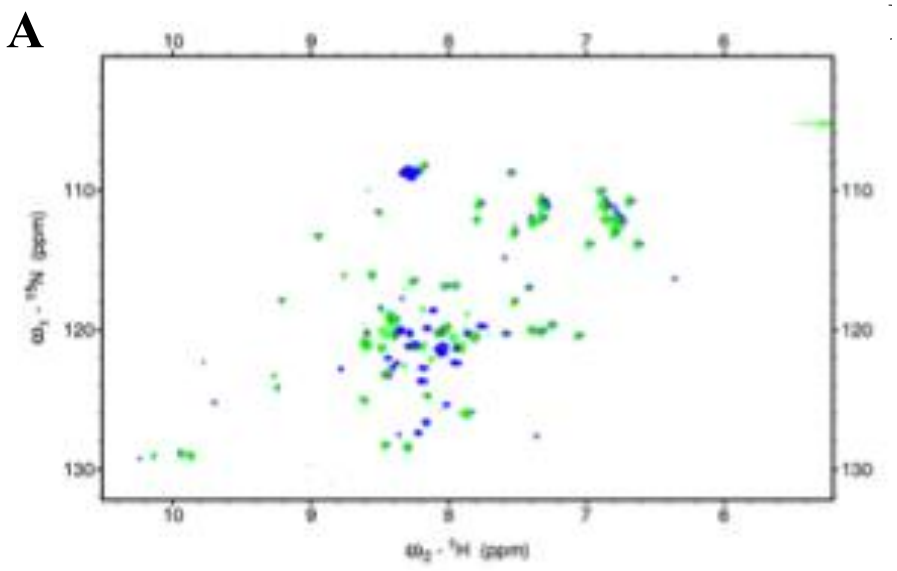


Figure 9: Panel A: Model of GST-CD2CD3 obtained from SAXS data. Panel B: Calculated SAXS scattering profile of GST-CD2CD3 model fits the experimental SAXS data. Panel C: Size-distance distribution of random pool and ensemble models for GST-CD2CD3.



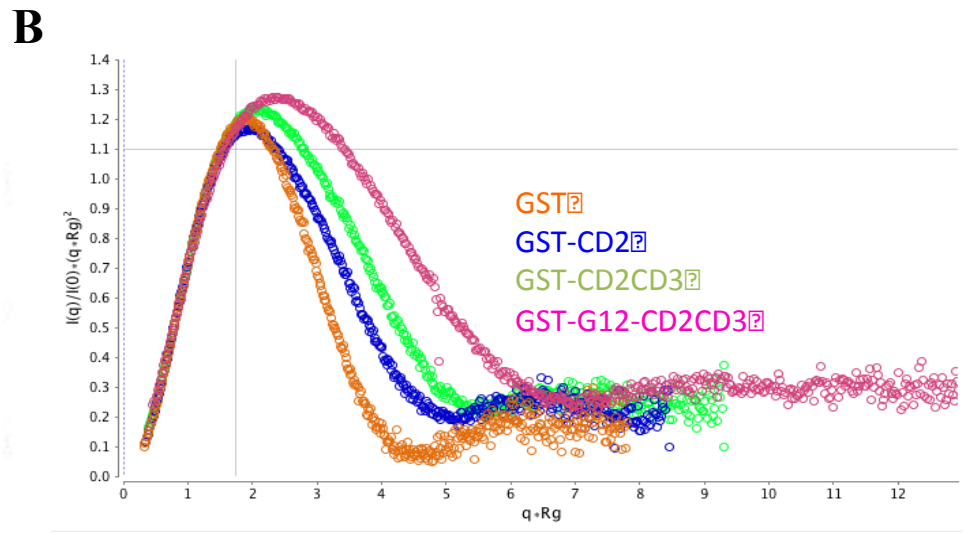


Figure 10: Panel A: Overlay of ^{1}H-^{15}N HSQC spectra of GST-G12-CD2CD3 (blue) and recombinant CD2CD3 (green). Panel B: Dimensionless Kratky- Plot of GST, GST-CD2, and GST-CD2CD3.

3.5. Discussion

The GST-affinity tag is a very versatile molecule and has found use in various biological applications[5,7,33,34,35]. All soluble GSTs that have been studied so far are found to form dimers implying the importance of this quaternary structure for the enzyme's function and activity[36]. This is also verified as engineered monomers lose their ability to bind to GSH and cannot be purified *via* GSH-Sepharose column but require the introduction of another affinity tag or using different chromatography methods[16]. Moreover, due to GST's high solubility and easy purification it is a popular affinity tag for the expression and production of recombinant proteins in high yield[5]. GST-fusion proteins are also useful in studies on protein-protein interaction using enzymatic assays or immunoassays. The GST pull-down assay is probably one of the most prominent techniques for the detection of proteins of interest and their interaction partners[37]. Besides, the GST-affinity tag has aided in the determination of structures of proteins, which were

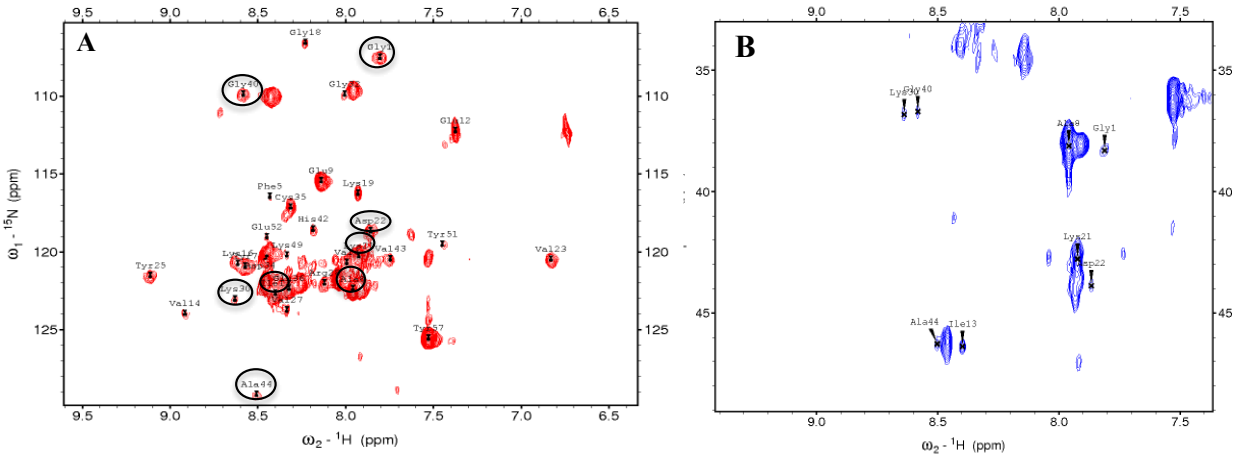
difficult to crystallize, with the help of GST in a protein-driven crystallization and subsequently the structure of the target protein could be determined by molecular replacement methods[12,21]. The conditions for the crystallization of GST-fusion proteins have proven to be similar, while one would have to spend more time and effort to optimize the crystallization settings for the individual target proteins[38]. Nevertheless, crystallization can only give a static picture, which is why NMR is often employed to elucidate the 3D solution structure of a protein of interest as well as to give insights on target protein's backbone dynamics. This study demonstrates that a GST-fusion protein can be a tool of elucidating the structural details of the target protein without having to remove the affinity tag. It is believed that the explanation for this observation of non-appearance of the cross-peaks corresponding to GST is two-fold. First, it has already been shown and this study confirms that just like GST individually, the fusion protein forms a dimer in solution. Consequently, the increased size due to dimerization leads to a decreased tumbling relaxation rates of GST portion compared to the target protein on a NMR timescale. Therefore, the cross-peaks corresponding to the GST residues broaden and disappear. Moreover, the results in this study indicate that internal flexibility and extended position, which have already been reflected by high temperature factors and crystal structures, are equally important as it allows viewing the two fusion partners as separate entities. Therefore, GST does not have an effect on the solution structure of CD2. Zhan et al and Vikis et al have also described the impact of such inter-domain flexibility. GST-fusion proteins with a thrombin cleavage site were crystallized and showed that the fusion partner protruded from the GST domain[21]. Moreover, GST-fusion proteins are used in GST pull-down assays because the fusion partner and affinity tag are "separated" by a linker that ensures full accessibility for both protein partners[7].

SAXS data gave better insight on the oligomeric state of GST-fusion proteins and the inter-domain flexibility of the two fusion partners. The determination of molecular weight based on the SAXS data confirmed results of the size exclusion chromatography, in which it was found that GST-CD2 forms a dimer in solution. On the other hand, the Kratky representation only showed a shift of the curve rather than a transformation in the profile. This suggests that once there is a fusion partner attached to GST, the thrombin linker helps the target protein to protrude from the structure of GST as well as also allows the target protein a certain degree of flexibility. The all atom-model generated for GST-CD2 visualized the multiple possible conformations, in which CD2 can occupy more than one specific location. In addition, the pair distance distribution function is in agreement with the extended state of the fusion protein GST-CD2. For a more discrete differentiation, the Porod-Debye Law was employed. It also confirms an increase in flexibility of the target protein within the fusion protein. The application of GST-fusion proteins for the structure determination of the target protein seems to work well for smaller proteins and peptides. Therefore, this approach could be applicable for the study of peptide-protein interactions and help in the elucidation of the binding interface of both interaction partners. However, when trying to extend it to larger proteins, the examples used in this study, CD2CD3 (11kDa), showed inconsistencies with previous observations. The overlay revealed that the backbone conformation of CD2CD3 when fused to GST was not the same when acquired individually. Even though SAXS indicated that this fusion protein was still flexible as the Porod exponent and Kratky plot indicated, the smaller GST-CD2 is well extended. Further, GST-CD2CD3's size distribution function suggests that the fusion protein can exist in both an extended or enclosed conformation. The computed models indicate that CD2CD3 can also fold back towards the GST dimer. It is therefore possible that the target protein interacts with GST

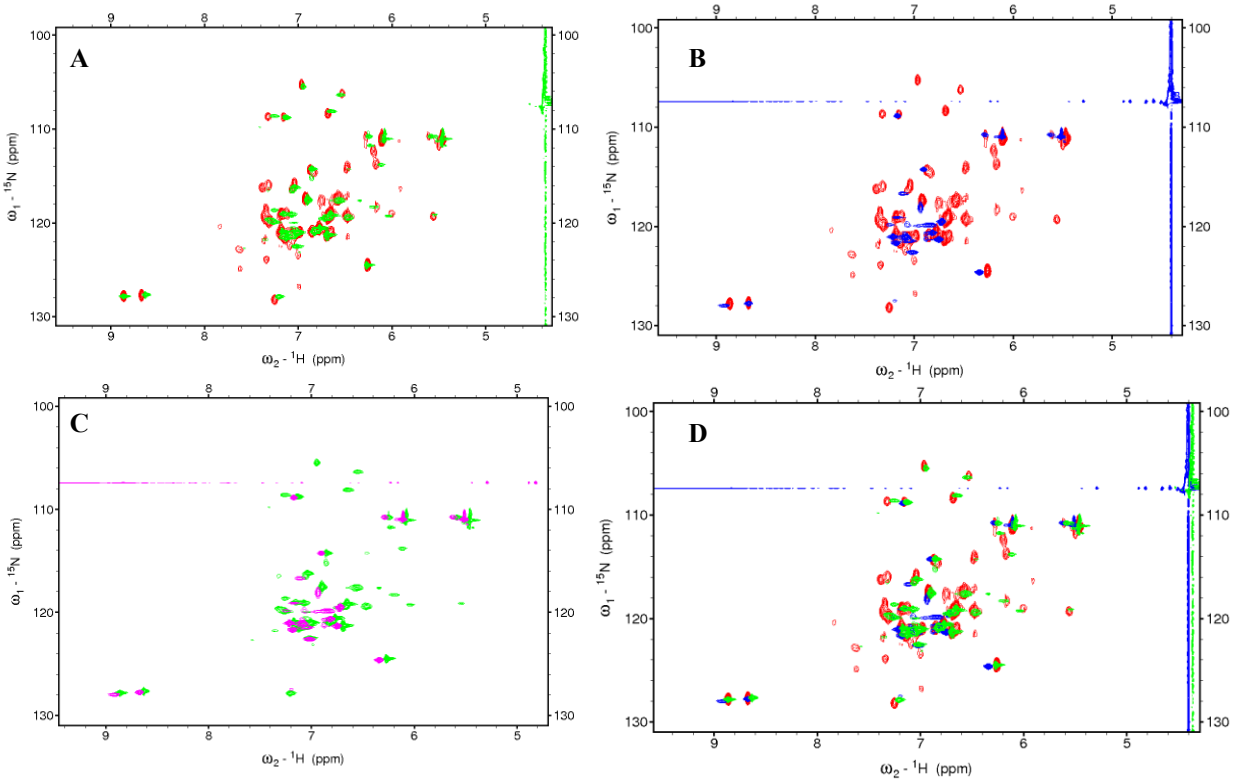
allowing for those residues to disappear along with GST or bind to the linker causing a different backbone conformation of CD2CD3. We successfully improved the applicability of this phenomenon for larger proteins by introducing 12 glycine residues into the linker that led to greater flexibility within GST-CD2CD3. The extension of the linker also appeared to prevent any associations with the GST dimer or linker as now the ^{1}H-^{15}N HSQC spectrum overlaid better with the spectrum for CD2CD3.

In summary, we believe that the use of GST-fusion proteins is a valuable alternative high throughput method in the elucidation of the 3D solution structure of the target protein without having to remove the affinity tag and can also be employed in the study of protein-peptide interactions.

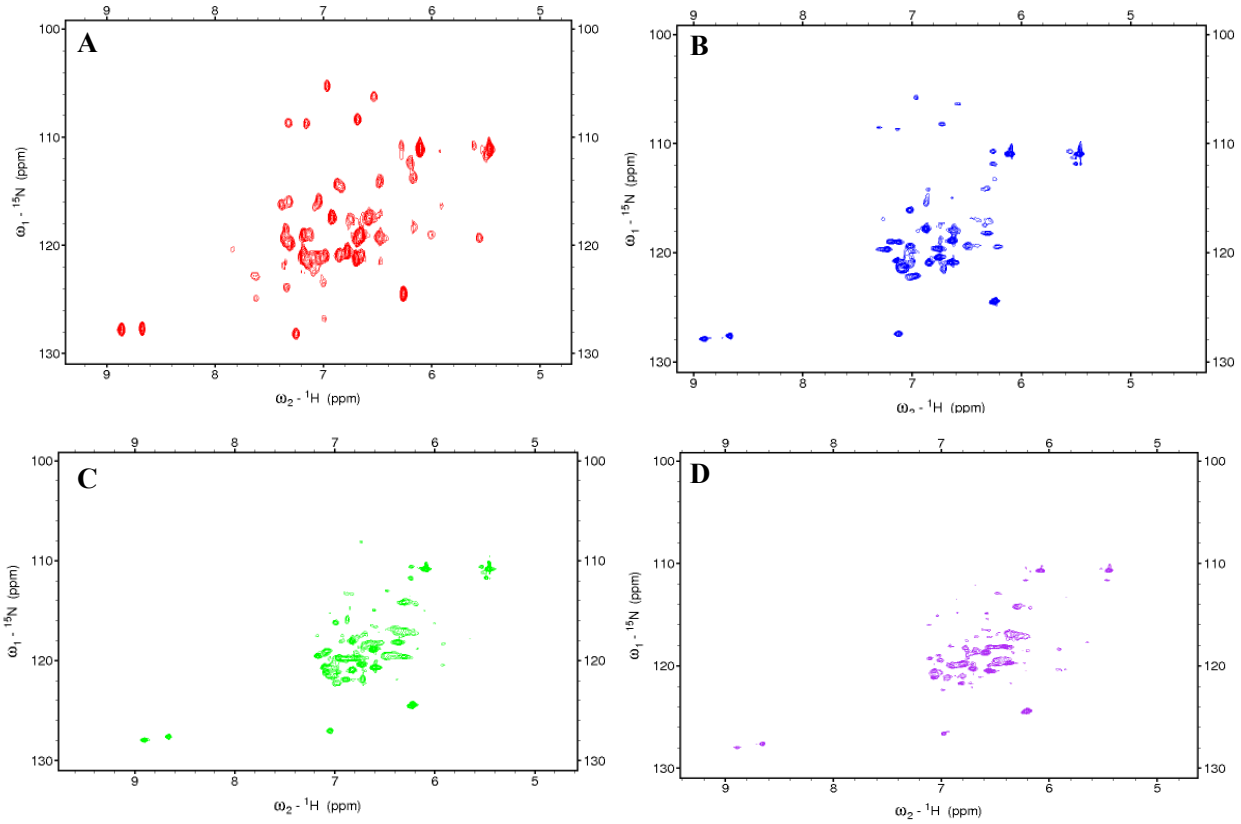
3.6. Supplement



Supplemental Figure 1: Panel A: ^{1}H-^{15}N HSQC spectra of GST-CD2. Panel B: 3D HNCA of GST-CD2.



Supplemental Figure 2: Overlay of ^{1}H-^{15}N HSQC spectra of the urea denaturation of GST-CD2. Panel A: 0 & 1 M urea, Panel B: 0 & 2.5 M urea, Panel C: 1 & 2.5 M urea, Panel D: 0 & 1 & 2.5 M urea)



Supplemental Figure 3: ^{1}H-^{15}N HSQC of temperature-induced denaturation of GST-CD2 at 295 K (Panel A), 305 K (Panel B), 310 K (Panel C), and 315 K (Panel D).

3.7. References

1. Armstrong, R. N., Glutathione S-transferases: reaction mechanism, structure, and function. *Chem. Res. Toxicol.* **1991,** *4* (2), 131-40.

2. Fabrini, R.; De Luca, A.; Stella, L.; Mei, G.; Orioni, B.; Ciccone, S.; Federici, G.; Lo Bello, M.; Ricci, G., Monomer-Dimer Equilibrium in Glutathione Transferases: A Critical Re-Examination. *Biochemistry* **2009,** *48* (43), 10473-10482.

3. Smith, D. B.; Davern, K. M.; Board, P. G.; Tiu, W. U.; Garcia, E. G.; Mitchell, G. F., Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83* (22), 8703-7.

4. Smith, D. B.; Johnson, K. S., Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988,** *67* (1), 31-40.

5. Smith, D. B., Purification of glutathione S-transferase fusion proteins. *Methods Mol. Cell. Biol.* **1993,** *4* (5), 220-9.

6. Harper, S.; Speicher, D. W., Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2011,** *681* (Protein Chromatography), 259-280.

7. Vikis Haris, G.; Guan, K.-L., Glutathione-S-transferase-fusion based assays for studying protein-protein interactions. *Methods Mol Biol* **2004,** *261*, 175-86.

8. Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; Dirr, H., Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **1997,** *6* (2), 399-406.

9. Dyson, M. R.; Shadbolt, S. P.; Vincent, K. J.; Perera, R. L.; McCafferty, J., Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **2004,** *4*, No pp given.

10. Hornby, J. A. T.; Codreanu, S. G.; Armstrong, R. N.; Dirr, H. W., Molecular Recognition at the Dimer Interface of a Class Mu Glutathione Transferase: Role of a Hydrophobic Interaction Motif in Dimer Stability and Protein Function. *Biochemistry* **2002,** *41* (48), 14238-14247.

11. McTigue, M. A.; Williams, D. R.; Tainer, J. A., Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel. *J. Mol. Biol.* **1995,** *246* (1), 21-7.

12. Lim, K.; Ho, J. X.; Keeling, K.; Gilliland, G. L.; Ji, X.; Rueker, F.; Carter, D. C., Three-dimensional structure of Schistosoma japonicum glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV. *Protein Sci.* **1994,** *3* (12), 2233-44.

13. Rufer, A. C.; Thiebach, L.; Baer, K.; Klein, H. W.; Hennig, M., X-ray structure of glutathione S-transferase from Schistosoma japonicum in a new crystal form reveals flexibility of the substrate-binding site. *Acta Crystallogr., Sect. F Struct. Biol. Cryst. Commun.* **2005,** *61* (3), 263-265.

14. Reinemer, P.; Dirr, H. W.; Ladenstein, R.; Schaeffer, J.; Gallay, O.; Huber, R., The three-dimensional structure of class π glutathione S-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *Embo J.* **1991,** *10* (8), 1997-2005.

15. Sayed, Y.; Wallace, L. A.; Dirr, H. W., The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: Implications for catalysis, ligandin function and stability. *Chem.-Biol. Interact.* **2001,** *133* (1-3), 60-62.

16. Abdalla, A.-M.; Bruns, C. M.; Tainer, J. A.; Mannervik, B.; Stenberg, G., Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein. *Protein Eng.* **2002,** *15* (10), 827-834.

17. Dirr, H. W.; Reinemer, P., Equilibrium unfolding of class π glutathione S-transferase. *Biochem. Biophys. Res. Commun.* **1991,** *180* (1), 294-300.

18. Erhardt, J.; Dirr, H., Native dimer stabilizes the subunit tertiary structure of porcine class pi glutathione S-transferase. *Eur. J. Biochem.* **1995,** *230* (2), 614-20.

19. Aceto, A.; Caccuri, A. M.; Sacchetta, P.; Bucciarelli, T.; Dragani, B.; Rosato, N.; Federici, G.; Di Ilio, C., Dissociation and unfolding of Pi-class glutathione transferase. Evidence for a monomeric inactive intermediate. *Biochem. J.* **1992,** *285* (1), 241-5.

20. Lally, J. M.; Newman, R. H.; Knowles, P. P.; Islam, S.; Coffer, A. I.; Parker, M.; Freemont, P. S., Crystallization of an intact GST-estrogen receptor hormone binding domain fusion protein. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **1998,** *D54* (3), 423-426.

21. Zhan, Y.; Song, X.; Zhou, G. W., Structural analysis of regulatory protein domains using GST-fusion proteins. *Gene* **2001,** *281* (1-2), 1-9.

22. Liew, C. K.; Gamsjaeger, R.; Mansfield, R. E.; Mackay, J. P., NMR spectroscopy as a tool for the rapid assessment of the conformation of GST-fusion proteins. *Protein Sci.* **2008,** *17* (9), 1630-1635.

23. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London, U. K.)* **1970,** *227* (5259), 680-685.

24. Nielsen, S. S.; Moller, M.; Gillilan, R. E., High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. *J. Appl. Crystallogr.* **2012,** *45* (2), 213-223.

25. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I., PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2003,** *36* (5), 1277-1282.

26. Franke, D.; Svergun, D. I., DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009,** *42* (2), 342-346.

27. Volkov, V. V.; Svergun, D. I., Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2003,** *36* (3, Pt. 1), 860-864.

28. Svergun, D. I., Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999,** *76* (6), 2879-2886.

29. Glatter, O.; Kratky, O.; Editors, *Small Angle X-ray Scattering*. 1982; p 515 pp.

30. Rambo, R. P.; Tainer, J. A., Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **2011,** *95* (8), 559-571.

31. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **2007,** *40* (3), 191-285.

32. Mylonas, E.; Svergun, D. I., Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **2007,** *40* (S1), s245-s249.

33. Yip, Y. L.; Smith, G.; Ward, R. L., Comparison of phage pIII, pVIII and GST as carrier proteins for peptide immunization in Balb/c mice. *Immunol. Lett.* **2001,** *79* (3), 197-202.

34. Nemoto, T.; Ota, M.; Ohara-Nemoto, Y.; Kaneko, M., Identification of dimeric structure of proteins by use of the glutathione S-transferase-fusion expression system. *Anal. Biochem.* **1995,** *227* (2), 396-9.

35. Derewenda, Z. S., The use of recombinant methods and molecular engineering in protein crystallization. *Methods (San Diego, CA, U. S.)* **2004,** *34* (3), 354-363.

36. Dirr, H., Folding and assembly of glutathione transferases. *Chem.-Biol. Interact.* **2001,** *133* (1-3), 19-23.

37. Singh, C. R.; Asano, K., Localization and characterization of protein-protein interaction sites. *Methods Enzymol.* **2007,** *429* (Translation Initiation: Extract Systems and Molecular Genetics), 139-161.

38. Carter, D. C.; Ruker, F.; Ho, J. X.; Lim, K.; Keeling, K.; Gilliland, G.; Ji, X., Fusion proteins as alternate crystallization paths to difficult structure problems. *Protein Pept. Lett.* **1994,** *1* (3), 175-8.

4. Conclusion

GST is an effective affinity tag to purify recombinant proteins and to investigate the 3D solution structure as well as protein-protein interactions of the fusion partner. Literature on the GST-tag suggests that this affinity tag can be utilized in both eukaryotic and prokaryotic expression systems and can easily be detected by means of antibodies. Moreover, this affinity tag yields large quantities of homogenous pure fusion protein in a timely manner.

In summary, we were able to develop a cost-effective alternative for the production of small proteins and peptides. We exploited the already established purification procedure for the fusion protein of this valuable solubility tag. Our studies showed that GST aggregates and can therefore specifically are removed by subjecting it to heat. Subsequently, we demonstrated the quick separation of the tag from the target peptide or small protein *via* ultracentrifugation. Experiments comparing the stability, secondary and tertiary structure, and biological activity suggested that our novel heat treatment method does not impair the small protein or peptide of interest. We are confident that this procedure saves time and resources and makes isotope enriched labeling of peptides now more realistic. A future direction of this method is the scalability of the process. Larger expression volumes lead to larger amounts of fusion protein that needs to be cleaved by thrombin, which in turn could be seen as cost-ineffective. An improvement of the existing procedure could be the use of chemical cleavage instead of employing enzymes in order to avoid high costs due to thrombin. However, due to the unfavorable cleavage conditions (i.e. dramatic pH changes), this adjustment might only be useful for the purification of peptides that are unstructured. Protein structure, stability, or biological activity could be disrupted under these harsh conditions associated with chemical cleavage. Also, in order to avoid cleavage within GST, one would need to mutate putative residues that would

represent a recognition site of the chemical reagent used and ensure that those alterations in the GST would not affect the affinity tags expression or purification.

In addition to developing a new purification method, we also optimized a technique for the acquisition of multidimensional NMR data for peptides and proteins by using the GST-fusion protein. This could be useful for the mapping of protein-peptide interaction interface without having to remove the affinity tag. Due to the flexibility that is introduced by the linker that is placed between the tag and the protein of interest, and the symmetrical dimerization of GST, we found that GST does not significantly affect the 3D solution structure of the fusion partner and the peaks corresponding to GST disappear. Extending the linker, by introducing 12 glycine residues, the flexibility was increased and better quality NMR spectra were obtained. In order to be able to use this method on larger proteins, the linker sequence would need to be optimized further. Not only does the linker provide flexibility that is important for the acquisition of NMR data, but also represents a spacer to ensure that GST does not interfere with the interaction of the target protein/peptide with its reaction partner.

5. Appendix – Towards the characterization of Anosmin-1

5.1. Abstract

Anosmin-1 is a secreted extracellular matrix associated glycoprotein that is encoded by the KAL1 gene[1]. It assists in the growth of olfactory and Gonadotropin- releasing hormone (GnRH) secreting neurons, which originate in the nasal compartment. Absence or damage of the protein has been shown to cause Kallmann Syndrome (KS), manifesting itself in the loss of olfactory bulbs and GnRH secretion most likely resulting from the unsuccessful embryonic migration of GnRH- synthesizing neuron[2]. Consequently, it leads to anosmia and hypogonadotropic hypogonadism respectively.

To date, the structure of the full-length Anosmin-1 has not been fully determined. However, it has been demonstrated that Anosmin-1 interacts with the fibroblast growth factor receptor 1 (FGFR1). It was concluded that it might be able to interfere with the fibroblast growth factor (FGF) signaling, which is known to play a vital role in cell growth and cell proliferation. The protein consists of a N-terminal cysteine-rich domain, a whey acidic like- protein domain (WAP) forming a so called four-disulfide core, followed by four fibronectin type III (FnIII) domains and a histidine-rich domain located at the C-terminus. Interestingly, both Anosmin-1 and FGFR1 interact with heparan sulfate to maximize their activity; this could be the linking feature to explain their relationship to each other.

In this context, the main goal in this study is to characterize and elucidate the 3D solution structure of Anosmin-1 in order to understand its role in the regulation of the activity of the FGF receptor.

5.2. Introduction

Anosmin-1 – its role in the development of sex organs

Anosmin-1 is an extracellular matrix associated glycoprotein that is encoded by the KAL1 gene[1]. It assists in the growth of olfactory and Gonadotropin- releasing hormone (GnRH) secreting neurons, which originate in the nasal compartment. Anosmia and hypogonadotropic hypogonadism are the defining symptoms of Kallmann Syndrome (KS), a rare disease acting rather on males than on females. Approximately 1 in 8,000 males and 1 in 40,000 females are affected [3,4,5]. Source of the disorder is very likely the unsuccessful embryonic migration of GnRH- synthesizing neurons in the olfactory bulb. Usually, these cells "migrate from the olfactory epithelium to the forebrain along the olfactory nerve pathway" as early as in the 6th embryonic week[6,7,8]. However, upon the deletion of KAL1 symptoms such as anosmia (lack of smell) due to deficiency of the olfactory bulb and dislocated GnRH neurons were observed[7]. It has been discussed that the deficit in GnRH is due to the failure of embryonic migration as described by Dode and Hardelin. In consequence, hormones important for gonadal maturation and function will not be released, providing the most common phenotype with delayed puberty[7,8]. Hence, it can be concluded that Anosmin-1 is involved in neurite outgrowth and axon branching and acts as an extracellular adhesion molecule[9].

Structural domains in Anosmin-1

KAL1 encodes for 680 amino acids, which have a theoretical molecular weight of 74 kDa; posttranslational modifications include N-glycosylation resulting in an observed molecular weight of 85-100 kDa[3,10]. Nevertheless, it is still unknown which of the possible glycosylation sites are actually implicated. Interestingly, Anosmin-1 does not contain a transmembrane domain or a glycosyl phosphotidyl inositol anchorage domain[3,11]. Structurally, the protein of interest

comprises of a N-terminal signal peptide, a cysteine-rich domain (CR), a whey acidic protein-like domain (WAP), four continuous fibronectin-like type III domains (FnIII), and a histidine-rich domain (HR) at the C-terminal region.

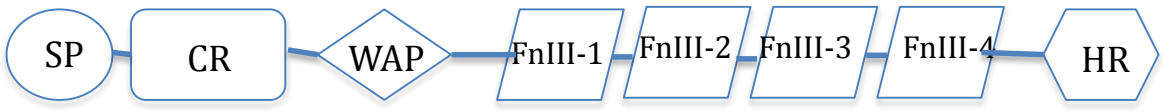


Figure 15 Structure of Anosmin-1 depicting the different structural domains

The WAP domain, evolutionary conserved, shows four disulfide core motives, a characteristic that is found in proteins associated with the primary immune defense, cell proliferation, and wound healing by inhibition of protease activity[3,9]. The exact disulfide bond pattern has not been described yet. Nevertheless, C151-C163 and C157-C172 have been identified to be essential for the stability of the protein[10].

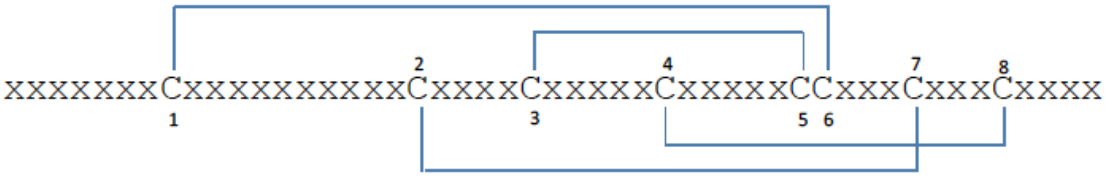


Figure 16 Possible motif of conserved disulfide bonds in the WAP domain of Anosmin-1 (Jayanthi et al 2011)

Similar to the WAP domain, the FnIII.1 domain is conserved, which implies its vital role in the function of the protein. The FnIII domains display significant similarity to the cell adhesion molecule (CAM) family which are usually related not only to cell-cell interactions but

also to migration during neural development[3]. Interaction studies have shown that of all the Anosmin-1 domains, the FnIII.1 domain's affinity to bind heparan sulfate is maximal, most likely due to its relatively large positively charged surface[10]. The WAP domain has also been demonstrated to have the ability to bind to heparan sulfate[10,12]. Nevertheless, it was described that the combination of WAP, FnIII.1 and CR domains has the highest affinity to interact with heparan sulfate[1].

According to Choy and Kim, mutational studies made it possible to recognize the most common modification causing loss-of-function of Anosmin-1[3]. Most of them are missense mutations, for example as can be found upon frame shifts and deletions. The WAP domain as well as the FnIII.1 domain seem to play a major role as 10 missense mutations were localized in these regions[3]. Moreover, most of these mutations were identified of being near proposed heparan sulfate binding sites. Examples are N267K[13,14] in FnIII.1 and E514K[14] in FnIII.3, which have been identified for X-linked KS patients. The N267K mutation leads to the loss of binding capability to FGFR, possibly due to changes in the protein conformation[12,15]. It is argued that the later alteration (E514K) increases the interactions with heparan sulfate due to the fact that the usual neutral or negative residues are now positively charged. Consequently, the electrostatic binding of the negatively charged heparan sulfate is promoted. Yet, as a result of this favored binding, Anosmin-1 seems to be more rigid and therefore moves less in the ECM than the wildtype protein[3]. Other mutations, like C172R and C163Y in the WAP domain have also been shown to hinder its activity. It is believed that the disruption of the disulfide core motif likely causes the destabilization of the protein[10]. These mutations have also been found in KS patients. Mutational studies have identified some common sites that affect the activity of Anosmin-1.

Still, little is known about the effects on the structure and function of the protein of these mutations[10].

FGF signaling

The Fibroblast Growth Factor (FGF) signaling plays an essential role in a wide range of cellular responses, such as cell proliferation, migration and differentiation, but is also important for tissue repair and tumor genesis[16,17]. There are 22 human FGF analogues, ranging in size from 17 – 34 kDa[18]. FGF has two types of molecules to interact with: on the one hand FGF binds to heparan sulfate. On the other hand it interacts with the five types of FGF Receptors (FGFR), an integral membrane protein consisting of a cytoplasmatic tyrosine kinase domain, a single transmembrane helix, and three extracellular immunoglobulin-like domains[17,19]. The binding of heparan sulfate of both FGF as well as FGFR is viewed as a necessary step for the modulation of cell activity. It has been reported that FGF interactions with the extracellular domain of its receptor induce the receptor tyrosine kinase and the intracellular signaling cascades that regulate various cellular processes[8]. This extracellular domain includes 3 immunoglobulin-like domains called D1-3[4]. The areas of D2 and D3 are said to be the FGF ligand-binding site, while D2 can interact with both FGF and heparin[4]. Once bound to the receptor, receptor dimerization and autophosphorylation of certain tyrosine residues are triggered. Moreover, more signaling processes, such as the phosphoinosidtide 3-kinase, phospholipase Cy and the classic mitogen-activated protein kinase are induced[12]. Important in this case is that FGFR activation can be controlled by the assembly of the extracellular signaling complex or by the intracellular signaling arrangement[12]. It has been shown that this is achieved by extracellular receptor modulators, for instance neuronal cell adhesion molecules, which are necessary for axon growth and neuronal migration[12].

FGFR- Anosmin-1 connection – Regulation of FGFR

Recent studies have indicated that in cases of a disrupted FGF signaling symptoms as in KS were observed. Anosmin-1, the first molecule to be implicated in the development of X-linked KS, is believed to interact with heparin and regulate the activation of the FGFR[6,9]. Extensive work has already been performed on the interaction of FGFR1 with heparin[17]. As mentioned earlier, the D2 domain of FGFR has the ability to bind heparin. Thus, it has been anticipated that it also can offer an interaction site for Anosmin-1. Hu et al suggested that certain domains in Anosmin-1, including the CR, the WAP, and the FnIII.1 domain, are able to directly interact with the FRFR and therefore might quite possibly play an active role in FGF signaling[18]. GST-pull down assays showed that individually WAP or FnIII.1 are reported to not bind to FGFR1[1]. Nevertheless, in combination they interact with the receptor[1]. Choy and Kim demonstrated that Anosmin-1 binds directly to FGFR1 through the FnIII.1 domain[20]. In addition, both domains, WAP and FnIII.1, have been shown to interact with heparin[6,9]. Moreover, it was observed that when Anosmin-1 modulates the initiation of FGFR1 signaling, neurite outgrowth and cytoskeletal rearrangement in human embryonic GnRH olfactory neuroblasts was stimulated[9,12]. However, it is still questionable, whether Anosmin-1 acts as the antagonist to FGF or supports its function. It is also unclear how the complexes are formed. It has been shown that Anosmin-1, when interacting with heparin, assists FGFR in the binding of its ligands FGF and thus helps in the dimerization of the receptor via heparan sulfate, while when bound to FGFR Anosmin-1 prevents the complex formation of FGFR with its substrates[12,18].

Rubredoxin from Pyrococcus furiosus

Rubredoxin is a small iron-sulfur protein and is considered to be a hyperthermostable protein[21,22,23]. Its molecular weight is 7.2 kDa and auto-oxidizes in the presence of air[23]. Understanding its ability to keep its proper fold at temperatures when other proteins already are denatured has been a challenging task. Hydrophobic packing interactions, van der Waals interactions, hydrogen bonding, and/or salt bridges have all been mentioned to be features that contribute to its thermostability[22].

Purification and Refolding of Proteins forming Inclusion Bodies

The formation of inclusion bodies during expression is more likely with increasing molecular weight which in turn means higher complexity of its fold[24] and has been considered unwanted in the effort of producing soluble recombinant proteins[25,26]. Inclusion bodies are attributed to impurities such as membranes, cell debris, membrane bound proteins and even still viable cells that can be found after cell lysis[27]. These accumulations of insoluble proteins are not suitable for applications such as characterization and any attempts to refold the proteins present in the inclusion bodies are quite challenging. Still, several isolation strategies of those clusters of insoluble proteins have been published as inclusion bodies show certain advantages. It is said that there is less proteolytic degradation in the inclusion bodies so that the expression yields can be up to 30% higher than that for cellular proteins[28].

In the past, the traditional method of purifying protein trapped in inclusion bodies consisted of several steps. First, one separates the inclusion bodies from the cellular debris, which is mostly done through centrifugation after cell lysis. The next step is to solubilize the cleaned protein aggregates using chaotropic reagents such as urea, guanidine hydrochloride, or detergents

(i.e. SDS). In the following, the solubilized proteins are refolded by removing said reagents[27]. Both of those steps are very critical for a satisfactory recovery of the protein of interest. Finally, the refolded protein of interests needs to be purified from any other contaminants.

More recently though new strategies have been proposed that focus on dilution, dialysis, or solid-phase separation as aggregation has become quite problematic[27,29]. The main idea is to physically separate partially folded protein in order to reduce intermolecular interactions and therefore decreasing aggregation. Dilution is especially used for the refolding of small-scale recombinant proteins. Even though it is problematic when trying to scale it up, it is still one of the most commonly used methods for refolding.

5.3. Materials and Methods

5.3.1 Cloning, overexpression and purification of Anosmin-1 and its subdomains into different expression systems

Cloning of His-Anosmin-1 and its subdomains in Pichia pastoris

The purchased pPICZ(α)B-Anosmin-1 clone was single and double digested with *EcoR1* and *Xba1* (NEB, New England Biolabs, MA) and the digestion products were separated on a 0.8% Agarose gel (Sigma-Aldrich, MO). The internal *EcoR1* site was mutated via site-directed mutagenesis (Agilent Quik Change II XL SDM Kit, Agilent Technologies, CA) to avoid digestion of *EcoR1* within Anosmin-1 during the cloning of His-Anosmin. Then the N-terminal His-tag was introduced through PCR amplification using the Taq-Polymerase (NEB). The 5' primer contains the nucleotide sequence for the His-tag. Both the PCR amplified insert and the vector were double digested with *EcoR1* and *Xba1* and ligated at a ratio of 1:8 (vector: insert). The clone was verified by colony PCR, single and double digestion with *EcoR1* and *Xba1*, and gene sequencing. In the following, multiple constructs of shortened versions of the full-length Anosmin-1 were created by introducing Stop-codons *via* site-directed mutagenesis (Agilent Quik Change II XL SDM Kit) creating the clones CR (HC), CR-WAP (HCW), CR-WAP-FnIII.1 (HCWF1), CR-WAP-FnIII.3 (HCWF3).

Growth curve of HSA and Overexpression of His- Anosmin-1

Human Serum Albumin (HSA) was supplied as a control protein for the overexpression in *Pichia pastoris* (EasySelect Pichia Expression Kit, Invitrogen, life technologies, CA). It was expressed according to the provided protocol in BMGY/BMMY medium (Buffered media with

glycerol as carbon source during cell growth and methanol as carbon source during induction; EasySelect Pichia Expression Kit,).

In order to transform Anosmin-1 or the other constructs in to the Pichia host strains (GS115, KM71H) the plasmid containing the DNA of interest was linearized *via* a single digestion with *Pme1* (NEB). Then, it was transformed either *via* electroporation or *via* Kit "EasyComp" (Invitrogen, CA) and plated on YPDS-plates containing 100 μg/ml Zeocin (Invitrogen, CA). A single colony was expressed according to the provided protocol (EasySelect Pichia Expression Kit) in BMGY/BMMY.

Subcloning and Overexpression of aB-CWF1

The gene coding for CR, WAP, and FnIII.1 domain (CWF1) was amplified *via* PCR then both the amplified insert and the vector were double digested followed by ligation at a ratio of 8:1 (insert: vector). As described before, the new clone was confirmed by gene sequencing and double digestion to release the insert, linearized with *Pme1* and transformed into GS115 using electroporation. The transformants were grown selectively on YPDS plates containing 100 μg/ml Zeocin. Once colonies were formed, they were streaked out again on YPDS plates containing 100 μg/ml Zeocin to confirm their antibiotic resistance due to incorporation of our gene of interest.

A single colony was then picked and inoculated in BMGY medium. After 2 days, the cell density was high enough, so that the cells were removed from the BMGY medium by centrifugation. The cells were resuspended in the induction medium BMMY containing 0.5% methanol upon the first induction, followed by 1% and 3% methanol for the subsequent inductions every 24 hours to maintain a constant expression pressure on the cells. Samples of broth, as well as pellet and supernatant after cell lysis were taken every 24 hours and run on a

Novex 4-20% Tris-Glycine mini gels (ThermoFisher Scientific, CA), followed by Western Blot. Furthermore, due to the large amount of samples, dot blots were performed in order to quickly spot check the samples and narrow down the samples used for Western Blot.

Subcloning, Expression and Purification of Anosmin-1's structural domains in E.coli

WAP

The vector pGEX-KG (GE Healthcare, MA) was used for the expression of WAP as a fusion protein with a N-terminal sequence coding for Glutathione S-transferase (GST). The affinity tag can be removed *via* thrombin cleavage (cleavage sequence Leu-Val-Pro-Arg-|| Gly-Ser) to obtain the recombinant protein of interest. The gene coding for WAP was PCR amplified from the *E.coli* codon optimized human full-length Anosmin-1 (GeneArt, Life Technologies, CA). Both pGEX-KG and the PCR product were double digested with *Nco1* and *Xho1* and ligated using Ligase (NEB). The plasmid holding the recombinant protein was transformed into BL21 (DE3) competent cells. A single colony was picked and grown in 10 ml LB medium (100 µg/ml Ampicillin) at 37 °C, 250 rpm, overnight. LB medium, containing ampicillin at the same concentration, was inoculated with 5% (v/v) overnight culture and induced with 1 mM IPTG for four hours once the OD_{600} had reached 0.6-0.8. The cells were harvested by centrifugation (20 minutes, 4 °C, 6,000 rpm) and washed with 1x PBS (pH 7.2).

A cell pellet from a 1 liter culture was resuspended in 20 ml 1x PBS (pH 7.2) and subjected to the French press for three passes at 1,000 psi pressure. After centrifugation of the lysate (20 minutes, 4 °C, 20,000 rpm) the supernatant was loaded onto the pre-equilibrated GSH-Sepharose column. Subsequently, the column was washed with 1x PBS (pH 7.2) until the baseline was reached and the protein was cleaved on-column with thrombin. For complete

cleavage 1U of thrombin for every 0.2 mg fusion protein was used. The reaction mixture was incubated for 24 hours on the rocker at room temperature. After cleavage, the protein of interest was eluted in 1x PBS (pH 7.2) and concentrated using Millipore concentrators. The molecular weight of the protein is 7.4 kDa and its yield is approximately 4 mg per 1 liter culture.

Rd-FnIII.1

Both pGEX-KG-FnIII.1 and pET22b-Rd-D2 were double digested with *BamH1* and *Xho1* (NEB) to release FnIII.1 and D2 respectively. Antarctic Phosphatase removed the 5' - phosphate group of the gel extracted vector pET22b-Rd, and the FnIII.1 insert and the pET-22b-Rd vector were ligated at a ratio of 3:1. To confirm the colonies as positive clones, the plasmid was double digested to release the newly inserted FnIII.1, checked using colony PCR, and the gene was sequenced. The plasmid holding the recombinant Rd-FnIII.1 was transformed into Rosetta competent cells. A single colony was picked and grown in 10 ml Terrific Broth medium (100 μg/ml Ampicillin and 100 μg/ml Chloramphenicol) overnight at 37° C, 250 rpm. TB medium, containing both antibiotics at the same concentration, was inoculated with 5% (v/v) overnight culture and grown at 32 °C, 250 rpm until the OD_{600} reached 0.5. Subsequently, the cultures were induced with 0.5 mM IPTG and incubated for 14 hours at 16 °C, 250 rpm. The cells were harvested by centrifugation (20 minutes, 4°C, 6,000 rpm) and washed with 1x PBS (pH 7.2).

A cell pellet of a 1 liter culture was resuspended in 20 ml 1x PBS (pH 7.2) and subjected to the French press for four passes at 1,000 psi pressure, followed by 10 cycles of sonication. After centrifugation of the lysate (30 minutes, 4 °C, 18,000 rpm) the supernatant was loaded onto a pre-equilibrated Ni-NTA column. Then, the column was washed with 1x PBS (pH 7.2) until

the baseline was reached. An imidazole gradient, consisting of 20 mM, 50 mM, 100 mM, 250 mM, and 500 mM was used to elute the protein of interest to purity. The molecular weight of Rd-FnIII.1 is approximately 20 kDa. The purification protocol was monitored by SDS-PAGE as well as Western Blot, which utilized antibodies raised against the His-tag.

Isolation and refolding of His-FnIII.1 from inclusion bodies

A cell pellet of 2 liter culture was resuspended by vortexing in 15ml 10 mM Phosphate buffer (PB buffer) and 100 mM NaCl (pH 7.2), sonicated, and centrifuged for 10 minutes at 13,000 rpm. After discarding the supernatant, the previous steps were repeated with 5 ml 10 mM PB buffer (10 mM EDTA, 0.5% TritonX-100, pH 7.2), 5ml 10 mM PB buffer (1M NaCl, pH 7.2), 5 ml 10 mM PB buffer (2 M urea), and finally 5 ml 10mM PB buffer (1% sodium lauroyl sarcosinate). Subsequently, the pellet was resuspended in 3 ml solution buffer (50 mM Tris buffer, 25% sucrose, 1 mM EDTA, 0.1% sodium azide, 10 mM DTT), sonicated, and lysozyme (0.4 mg per ml solution buffer) and magnesium chloride (final concentration 2 mM) were added. In the following, 3 ml of lysis buffer (50 mM Tris buffer, 1% TritonX-100, 1% deoxycholate, 100 mM NaCl, 0.1% sodium azide, 10 mM DTT) were added, and incubated for 30 minutes at room temperature. EDTA (final concentration 7 mM) was pipetted to the mix, flash frozen in liquid nitrogen, and thawed for 30 minutes at 37 °C. More magnesium chloride (final concentration 1 mM) and EDTA (final concentration 7 mM) were added to the mixture, incubated for 30 minutes at room temperature, and then centrifuged for 15 minutes at 13,000 rpm. Finally, the newly formed pellet was washed twice with wash buffer (50 mM Tris buffer, 100 mM NaCl, 1 mM EDTA, 0.1% sodium azide, 1 mM DTT). For the first round of washing, 0.5% of TritonX-100 was also added to the buffer.

5.3.2 Identification and Characterization of the Heparin-Binding regions in Anosmin-1

Binding affinity of the WAP domain to Heparin

The binding affinity of the WAP domain to heparin was observed and measured by Isothermal Titration Calorimetry using the ITC-200 (Microcal Inc, MA). The experiments were conducted at room temperature, at a protein concentration of 100 μM *vs*. 2 mM heparin in the presence and absence of 500mM NaCl. WAP was dialyzed against 1× PBS pH 7.2. Samples were centrifuged to remove any aggregated or precipitated protein and were degassed before the titration. Heparin was added sequentially in 1.3μL aliquots to WAP with a 12-s interval between injections. The heats of reaction per injection (μcalories/s) were determined by the integration of peak areas by the Origin Version 7.0 software. The dissociation constant K_d was derived after fitting the data using a one-site of binding model.

Studies on the Structure of the WAP domain of Anosmin-1

Secondary Structure of the WAP domain

Left and right polarized light are differentially absorbed due to optically active chiral molecules giving insight in the secondary structurural changes of the WAP domain upon binding to heparin. 145 μM WAP in 1x PBS (pH 7.2) were analyzed in the presence and absence of 5 mM DTT or 725 μM heparin in a far-UV CD spectrum under standard sensitivity mode. 10 accumulations of each sample were acquired at room temperature, a path length of 0.2 mm and a scan speed of 50 nm/min and averaged. The spectra were corrected by subtracted buffer blanks and plotted as molar ellipticity.

HSQC of the WAP domain

In order to elucidate the 3D solution structure of the WAP domain at atomic resolution NMR experiments were performed on the Bruker 500MHz spectrometer, which is equipped with a cryo-probe. GST-WAP was expressed in ^{15}N enriched M9 minimal media and the protein of interest was purified to homogeneity. The ^{1}H-^{15}N HSQC was acquired of a 0.5 mM WAP sample and gives the fingerprint of the backbone conformation of the protein of interest.

Studies on the Stability of the WAP domain of Anosmin-1

Thermodynamic Stability of the WAP domain

Heat capacities of the WAP domain were measured during the thermal denaturation using the NANO DSCIII in order to understand the stabilization of WAP in the presence and absence of heparin. WAP was dialyzed against 1x PBS (pH 7.2), centrifuged to remove any aggregated or precipitated protein, and degassed before the obtaining the DSC data. The scans were performed at a ramping temperature of 1 C/min from 15-80 °C. The concentration of WAP was 1 mg/ml.

Limited Trypsin Digestion of the WAP domain in presence and absence of heparin

In this experiment changes in the stability of WAP upon binding to heparin were explored. It was conducted at a protein concentration of 15 μM in the presence or absence of 150 μM heparin. Moreover, 0.01mg/ml of trypsin were used and its digestion over 20 minutes was monitored *via* SDS-PAGE.

5.4. Results and Discussion

5.4.1 Cloning, overexpression and purification of Anosmin-1 and its subdomains into different expression systems

Cloning of His-Anosmin and combinations of its subdomains in Pichia pastoris

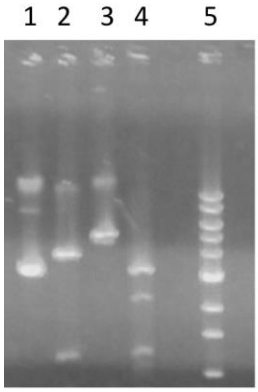


Figure 17 Single and Double Digestion of Anosmin. Lane 1 undigested, Lane 2 digested with *EcoR1*, Lane 3 digested with *Xba*, Lane 4 double digested

In figure 3, the purchased pPICZ(α)B-Anosmin-1 clone was single and double digested with *EcoR1* and *Xba1* in order to confirm its authenticity. Lane 1 shows the undigested sample. The characteristic bands for the supercoiled plasmid can be seen. In Lane 2 Anosmin was incubated with *EcoR1*. We only expected the linearized band. However, this lane shows that the internal *EcoR1* site was not mutated, giving us an 800 bp band and a 4.5 kb band. Lane 3 depicts the almost completely linearized sample with *Xba1*. Double Digestion with *EcoR1* and *Xba1* was performed for the sample in lane 4. Again, we can see the vector band at 3.5 kb, and the released insert, which is digested due to its internal *EcoR1* site giving us the 1200 bp and 800 bp bands. Other ways to establish the authenticity of the clones can be done by transformation and expression. For one, colony PCR should show the 2 kb band of the insert and Western Blot can specifically display the protein bands due to their His-tag.

In Figure 4, the SDS-PAGE and Western Blot of the expression of Anosmin-1 is depicted. Even though this is the secretory clone, the protein of interest seems to be detected in

the pellet. Moreover, it does not show the expected molecular weight. The reason for this observation could be that the protein is already degraded and only the part with the His-tag is detected. Therefore, expression conditions need to be optimized.

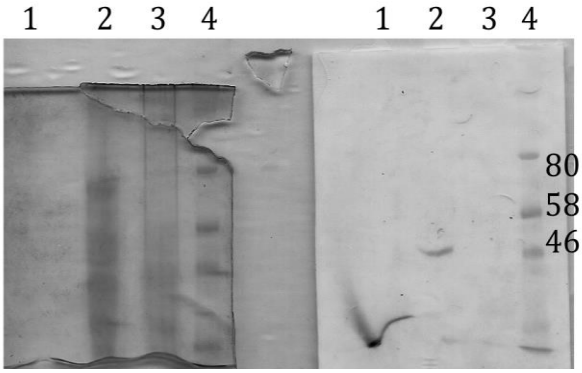


Figure 18 SDS-PAGE (left) and Western Blot (right) of pPICZ(a)B-Anosmin in KM71H. Lane 1 positive control, Lane 2 pellet after cell lysis, Lane 3 supernatant after lysis, Lane 4 prestained protein marker

For the sub-cloning, a N-terminal His-tag was introduced for purification purposes. Authenticity of this clone was verified by colony PCR. The amplified product with its expected size of 2 kb can be observed in lane 4 and 6 of Figure 5.

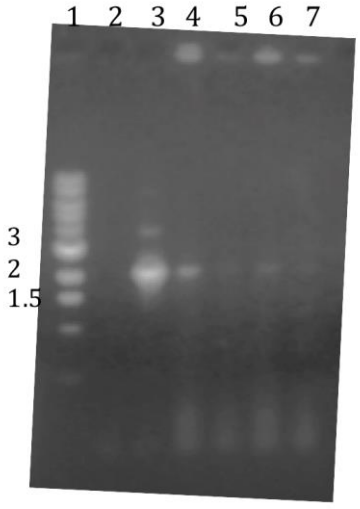


Figure 19 Colony PCR of His-Anosmin; Lane 1 1kb ladder, Lane 2 neg. control, Lane 3 pos. control, Lane 4-7 colony 1-4

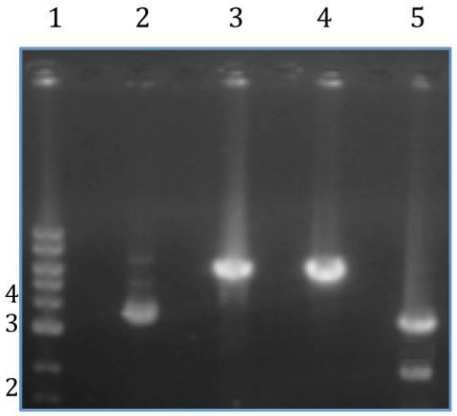


Figure 20 Single and Double Digestion of pPICZaB-His-Anosmin. Lane 1 1kb Ladder, Lane 2 undigested, Lane 3 and 4 Single digestion with *EcoR1* or *Xba* respectively, Lane 5 double digestion with *EcoR1* and *Xba*

In addition, the new clone was subjected to single and double digestion (Figure 6). Lane 2 depicts the undigested plasmid containing His-Anosmin-1. The characteristic bands for the supercoiled plasmid can be seen. In lanes 3 and 4 the plasmid was single digested with *EcoR1* and *Xba* respectively. The bands in lane 3 and 4 migrate at their expected size of 5.6 kb (vector: 3.6 kb, inserted gene: 2 kb). Lane 5 shows the bands of the double digestion with *EcoR1* and *Xba*, with a vector band detected at 3.6kb and the released insert at 2kb.

In the following, site directed mutagenesis was performed to introduce a stop codon, shortening the full-length protein from the back. The following clones were created in pPICZaB:

- His-CR (HC)
- His-CR-WAP (HCW)
- His-CR-WAP-FnIII.1 (HCWF1)
- His-CR-WAP-FnIII.3 (HCWF3)

Again, the new clones were confirmed by sequencing. All constructs and the full-length protein were transformed into the yeast strain GS115, which requires linearization of the plasmid by *PmeI* (Figure 8). After transformation into the yeast host, their integration into the chromosomal DNA was verified by colony PCR. For this purpose, the genomic DNA of the yeast colonies was isolated using the ZymoResearch YeaStar Genomic DNA Kit. All constructs and the full-length protein were confirmed. The expected size of the amplified product for each clone, using the provided AOX primers, is 2 kb (Figure 7). In case of the positive control, the genomic DNA of the control protein HSA was isolated and subjected to PCR, giving an intense band at the expected size of 2.1kb.

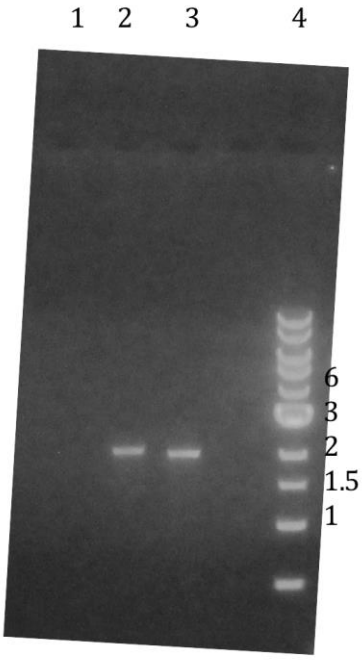


Figure 21 Colony PCR of His-Anosmin in GS115; Lane 1 neg. control, Lane 2 pos. control HSA, Lane 3 His-Anosmin GS115, Lane 4 1kb ladder

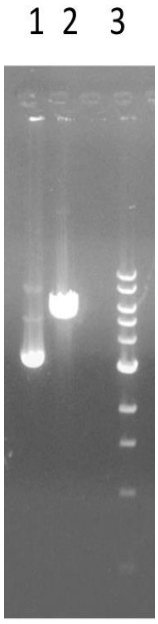


Figure 8 Single Digestion with *Pme*. Lane 1 undigested, Lane 2 single digestion with Pme, Lane 3 1kb ladder

Growth curve of HSA and Overexpression of His- Anosmin-1

Based on this result, the media was changed to the induction medium after 16-17 hours of incubation during its exponential growth phase (Figure 9). In Figure 10 is the expression of the control strain Human Serum Albumin (HSA) pictured. HSA has a molecular weight of 66 kDa. Being a secreted protein it was, as expected, detected with good yield in the broth (Figure 10, lane 2). Important for this expression is aeration, which can be achieved with baffled flasks. During induction, 0.5% Methanol was added every 24 hours, creating a stress on the cells. As a consequence, the promoter of AOX (alcohol oxidase) was induced and the gene coding for the protein of interest, which was integrated after the AOX promoter, is translated, and due to its signal sequence was secreted into the broth.

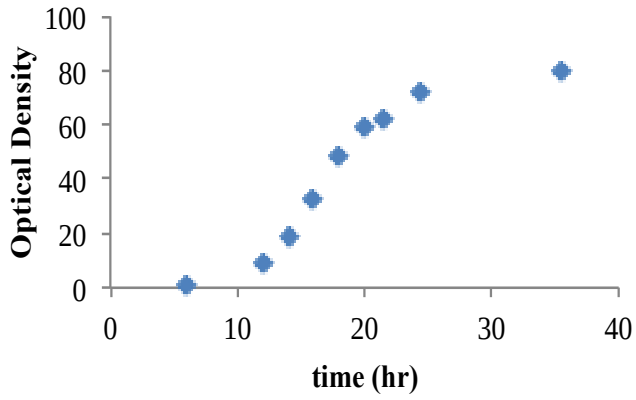


Figure 9 Growth Curve of HCWF1 in GS115

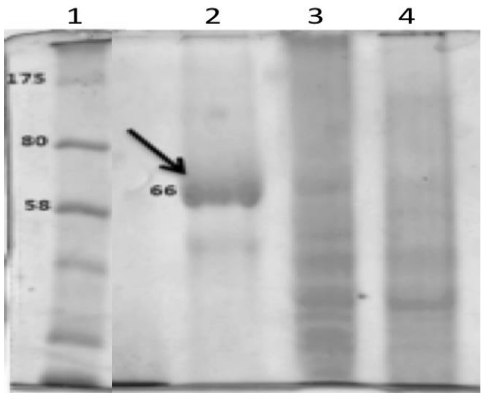


Figure 10 10% SDS-PAGE of expression of HSA. Lane 1 prestained protein marker, Lane 2 broth, Lane 3 or 4 pellet or supernatant after cell lysis respectively

The secreted (Figure 11, lane 4-6) and the intracellular (Figure 11, lane 2/3) clones of Anosmin-1 were expressed. Both hosts, GS115 and KM71H were examined. The advantage with the later cell line is that in this strain the gene, coding for AOX, was genetically disrupted. Both AOX, a homooctomeric protein with 8 80 kDa subunits, and Anosmin-1 (~74 kDa) have about the same molecular weight. This posed to be difficult to see Anosmin-1 in GS115.

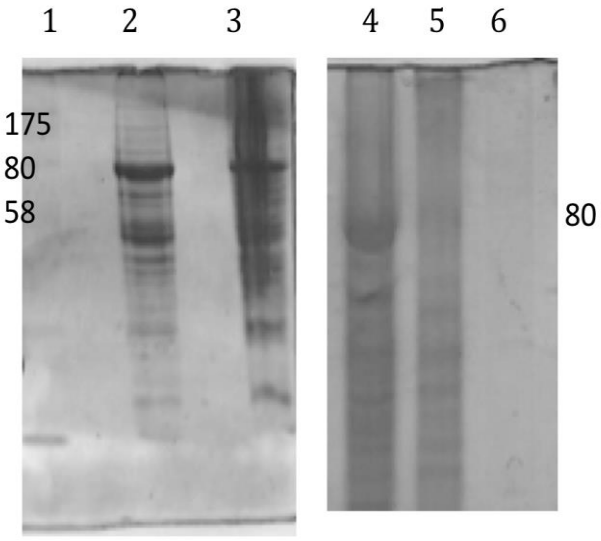


Figure 11 SDS-PAGEs showing the overexpression of Anosmin-1. Lane1 protein marker, Lane2 pellet after lysis, Lane3 supernatant after lysis, Lane4 supernatant after lysis, Lane5 pellet after lysis, Lane6 broth

Subcloning and overexpression of aB-CWF1

Due to the problems with the full length Anosmin-1 expression, we chose to focus on the combination of CR-WAP-FnIII.1. It has been sown that these domains are crucial for Anosmin-1's biological activity[14,20]. However, the N-terminal His-tag might be problematic during transformation or secretion. Therefore, CR-WAP-FnII.1 was subcloned without said N-terminal tag that was introduced at the beginning (Figure 12 and 13). Moreover, the empty vector pPICZaB and the control protein prolactin, which was already been shown to express in *Pichia pastoris,* were also transformed and expressed.

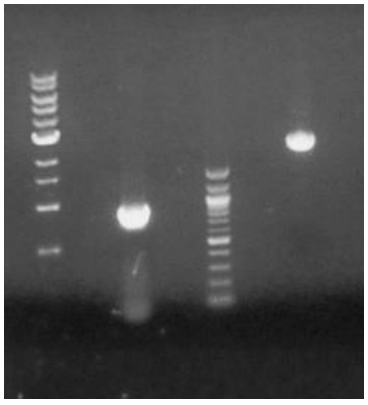


Figure 12 PCR amplification of aB-CWF1. From left to right: 1kb DNA ladder, PCR product, 100bp ladder, plasmid pPICZaB

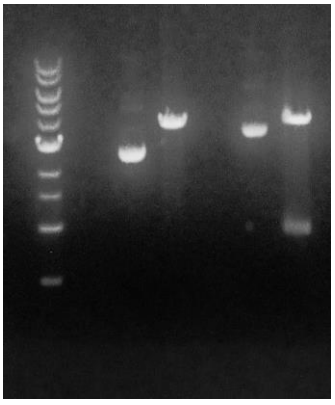


Figure 13 Double Digestion of transformants to check to positive clones. From left to right: 1kb ladder, undigested pPICZaB, undigested transformant, double digested transformant

Even though the control protein prolactin was expressed after 24 and 48 hours (Figure 14, Lane 2 and 3) and could be confirmed by Western Blot and Dot blot against both *cmyc* epitope and His-tag (Figure 15), the protein of interest could not be detected using those same expression conditions.

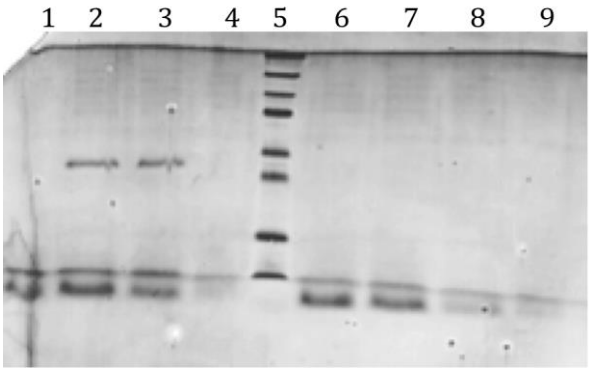


Figure 14 SDS-PAGE of Overexpression of control protein prolactin and empty vector pPICZaB. Lane1 Prolactin 0 hr, Lane2 Prolactin 24 hr, Lane3 Prolactin 48 hr, Lane4 Prolactin 72 hr, Lane5 prestained protein marker, Lane6 pPICZaB 0 hr, Lane7 pPICZaB 24 hr, Lane8 pPICZaB 48 hr, Lane9 pPICZaB 72 hr

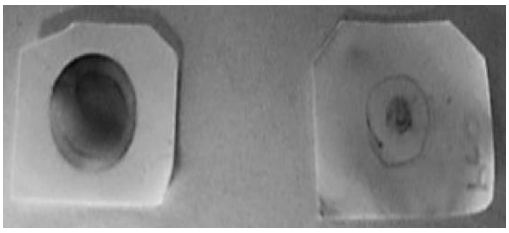


Figure 15 Dot Blot against His-tag: left pos. control His-Rd-FnIII.1, right Prolactin broth 48hr

Subcloning, expression and purification of the individual WAP and FnIII.1 domain in E.coli

Overexpression and purification of the WAP domain

Of all the recombinant constructs only the WAP domain expressed in the soluble fraction in *E.coli*. In figure 16 a 15% SDS-PAGE of the overexpression of the protein of interest is shown. The WAP domain was successfully overexpressed (Figure 16, lane3). Its expected molecular weight is about 33kDa, which combines the GST-tag (26kDa) and WAP (7kDa). Lane 5 depicts that after cell lysis, the protein is mainly found in its soluble form in the supernatant. The fusion protein was purified to homogeneity (Figure 17, lane3) and completely cleaved by thrombin. After on-column cleavage, WAP eluted in the flow through (lane7) yielding approximately 4 mg of the 7 kDa protein of interest per 1 liter culture.

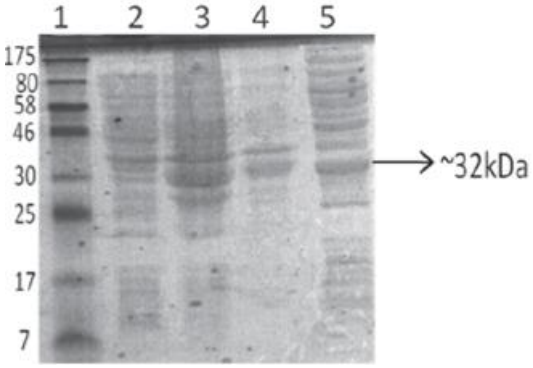


Figure 16 Lane1 Pre-stained proteinmarker, Lane2 uninduced, Lane3 induced, Lane4 pellet after Lysis, Lane5 supernatant after lysis

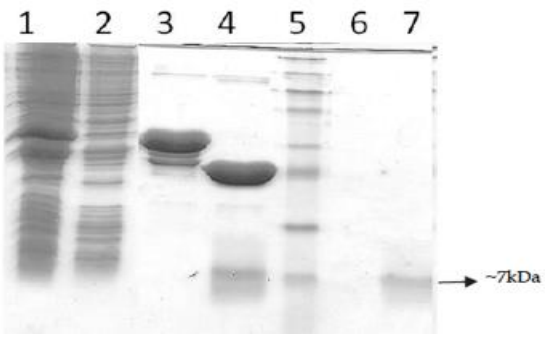


Figure 17 Lane1 supernatant, Lane2 flow through, Lane3 fusion protein, Lane4 cleaved fusion protein, Lane5 pre-stained protein marker, Lane6 empty, Lane7 on-column cleaved WAP

Cloning of Rd-FnIII.1

Due to its tendency to form inclusion bodies when expressed in *E.coli* an attempt was made to clone FnIII.1 as a fusion protein with His-tagged Rubredoxin (Rd). Rd has a molecular weight of 7.2 kDa and is known to be able to keep its native fold at extreme temperatures. It is expected to extend its thermostability to the protein of interest by stabilizing FnIII.1 in the soluble fraction for subsequent purification.

Both pGEX-KG-FnIII.1 and pET22b-Rd-D2 were double digested with *BamHI* and *Xho* to release FnIII.1 and D2 respectively. FnIII.1 with the size of 315 bp and pET22b-Rd, 5.6 kb, were excised (Figure 18), ligated, and transformed into DH5α competent cells. Both colonies were analyzed by double digestion of their plasmid DNA and show vector and insert bands at the expected sizes, 5.6 kb and 315 bp respectively (Figure 19). In case of the colony PCR experiment, the T7 promoter and terminator were used. Therefore, the positive controls Rd-D2 (Figure 20, lane 3) are expected to migrate with a size of 600 bp and Rd-FGF (Figure 20, lane 4)

with 300 bp. The two colonies, which were analyzed *via* double digestion (Figure 20, lane 5 and 6), show the expected size of 715bp. Therefore, they were confirmed to be positive clones.

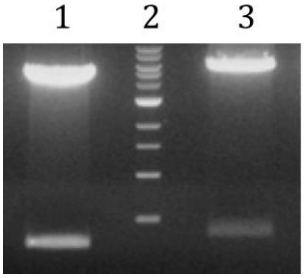


Figure 18 Double digestion products of pGEX-KG-FnIII.1 (Lane1) and pET-22b-Rd-D2 (Lane2)

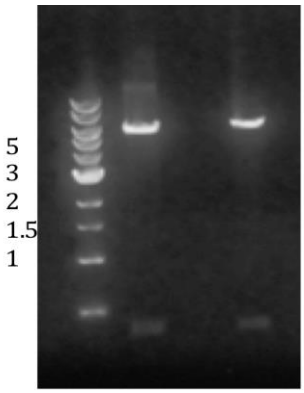


Figure 19 Double digestion of colonies to verify authenticity of Rd-FnIII.1

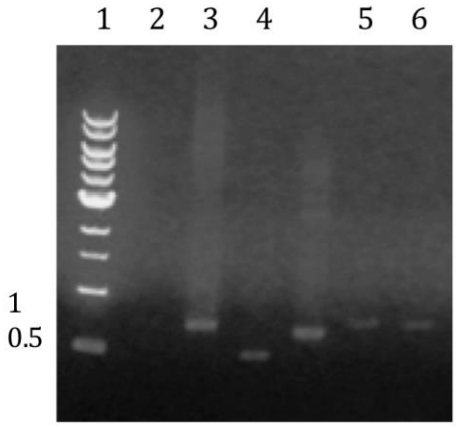


Figure 20 Colony PCR to verify Rd-FnIII.1 clones. Lane1 1kb ladder, Lane2 neg. control, Lane3 and 4 pos. controls, Lane 5 and 6 colony 1and 2

The purification procedure was monitored by SDS-PAGE (Figure 21A) and the protein of interest was detected by Western Blotting using antibodies against the His-tag (Figure 21B). A portion of the expressed Rd-FnII.1 is found in the soluble fraction after cell lysis (Figure 21B, lane 3). Significant amount of protein is still found in the pellet (Figure 21B, lane 2). The target protein mainly elutes at an imidazole concentration of 250 mM and is also detected in the 500 mM fraction (Figure 21B, lane 8 and 9). Nevertheless, some protein is lost as it precipitates on the column and therefore elutes with guanidine hydrochloride (Figure 21B, lane 10). Moreover, the SDS-PAGE shows that both the 250 mM as well as the 500 mM imidazole fraction do not contain pure protein, but have high molecular contaminants. This can be due to inappropriate buffer conditions and pH differences in the elution buffers. Moreover, the yield of Rd-FnIII.1 is

still considerably low. Careful cell lysis is imperative in order to keep the majority of the protein of interest in the solution.

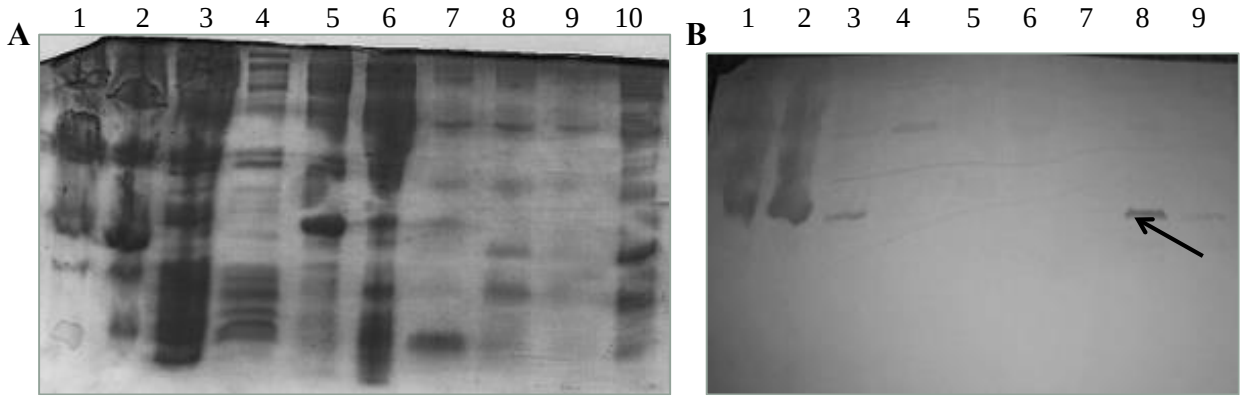


Figure 21 Panel A) SDS-PAGE depicting the purification of His-Rd-FnIII.1 and Panel B) corresponding Western Blot; Lane1 positive control for Western Blot, Lane2 pellet after lysis, Lane3 supernatant after lysis, Lane4 flow through, Lane5 20 mM imidazole, Lane6 50 mM imidazole, Lane7 100 mM imidazole, Lane8 250 mM imidazole, Lane9 500 mM imidazole, Lane10 6M GdnHCl

Isolation and refolding of FnIII.1 from inclusion bodies

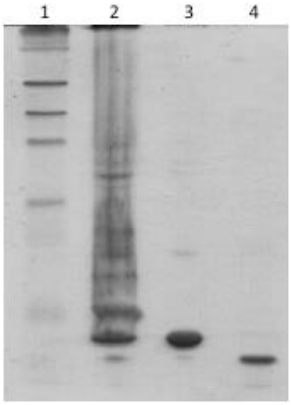


Figure 22 SDS-PAGE depicting the isolation and refolding of FnIII.1. Lane1 prestained protein marker, Lane2 pellet post deoxycholate cleaning, Lane3 pellet after refolding, Lane4 supernatant post refolding

Due to poor yields using the Rd-FnII.1 clone, a protocol was optimized using a combination of three approaches: cleaning, solubilizing, and refolding of inclusion bodies. Firstly, the inclusion bodies were isolated and solubilized from cell debris and other contaminating proteins using various buffers containing either chaotropic reagents or detergents. To further isolate FnIII.1, the inclusion bodies are treated with deoxycholate (Figure 22, lane 2) and then refolded via flash dilution (ratio 1:10) into a refolding buffer containing both oxidized and reduced glutathione (Figure 22, lane 4). Figure 22 depicts the SDS-PAGE monitoring the

procedure that was optimized to clean, solubilize, and refold FnIII.1. The bands in lane 3 and 4 appear to be clean isolated protein. As both the bands move very close to the expected molecular weight of FnIII.1, a Western Blot was performed identifying both bands as the target protein. FnIII.1 is also present in the pellet lane 3. Probably one can make further attempts in the future to refold this protein by further diluting to avoid aggregation and consequently increasing the yield.

5.4.2 Identification and Characterization of the Heparin-Binding regions in Anosmin-1

It has been reported that Anosmin-1, especially its FnIII.1 domain and WAP domain, bind to heparin. A more detailed characterization in terms of stability, structure, and functionality is presented on the WAP domain.

Binding affinity of the WAP domain to Heparin

The binding affinity of the WAP domain to heparin was observed and measured by ITC. Moderate binding affinity of WAP to heparin was detected, with a K_d value of 590 μM. Moreover, it was revealed that binding disappeared in the presence of 500 mM NaCl, confirming that the interaction of WAP with Heparin is of electrostatic nature (Figure 23).

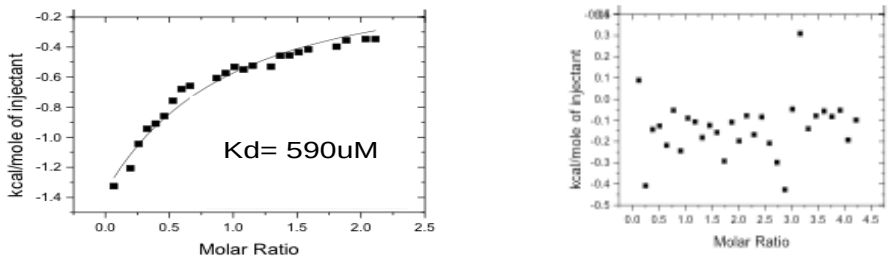


Figure 23 ITC Profiles of 0.1 mM WAP vs. 2 mM Heparin in the absence (left) and presence (right) of 500 mM NaCl

Studies on the Structure of the WAP domain of Anosmin-1

Secondary Structure of the WAP domain

Circular Dichroism was employed in order to investigate the secondary structure and possible structural changes in the protein upon binding to its interaction partner. As one can

observe in Figure 24, WAP's far UV spectrum displays a minimum at 205 nm and therefore exhibits a mix of secondary structural motifs characteristic for a random coil and α-helix. Upon binding to heparin, the target protein's far UV spectrum shifts. The minimum is now observed at 209 nm, concluding the WAP gains α-helical character when interacting with heparin. Moreover, the addition of the strong reducing agent DTT (dithiolthreitol) disturbs the network of disulfide bonds, upon which WAP's structure transitions to a random coil. This was detected by the shift of the minimum to 200 nm, even further than was shown for apo WAP. This implies that these disulfide bonds are important to maintain the structure of the protein.

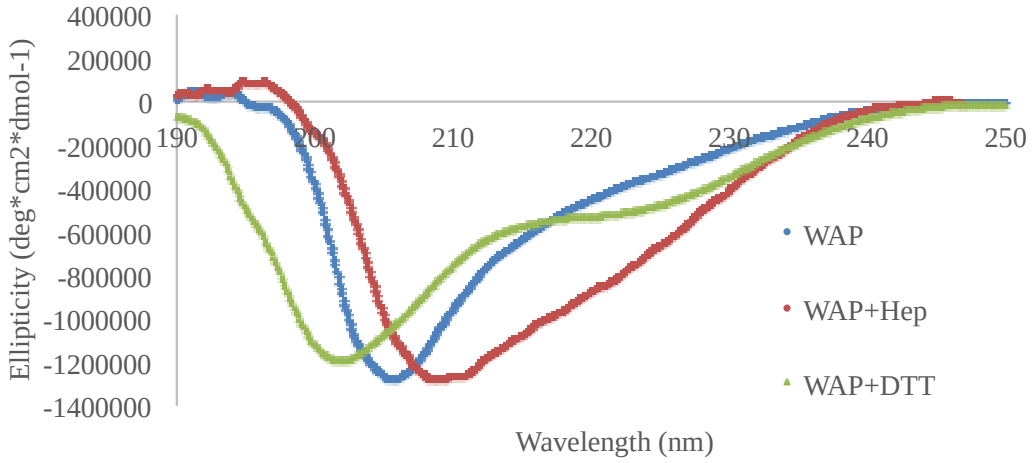


Figure 24 Far-UV CD Spectra of WAP in the presence and absence of 5 mM DTT or heparin

3D solution structure of the WAP domain

Multidimensional nuclear magnetic resonance spectroscopy (NMR) is a useful technique to elucidate the 3D solution structure and backbone dynamics at atomic resolution. Figure 25 depicts the ^{1}H-^{15}N HSQC spectrum of WAP. The cross-peaks are spread-out and well dispersed indicating that the WAP domain is structured. Nevertheless, there are more peaks visible than the protein of interest has residues. As each peak represents one residue in a particular backbone

conformation of WAP, this observation suggests the there are multiple populations of WAP are present in the analyzed sample. Moreover, the possibility of contaminants can be ruled out as the SDS-PAGE confirmed a pure preparation of the WAP-sample.

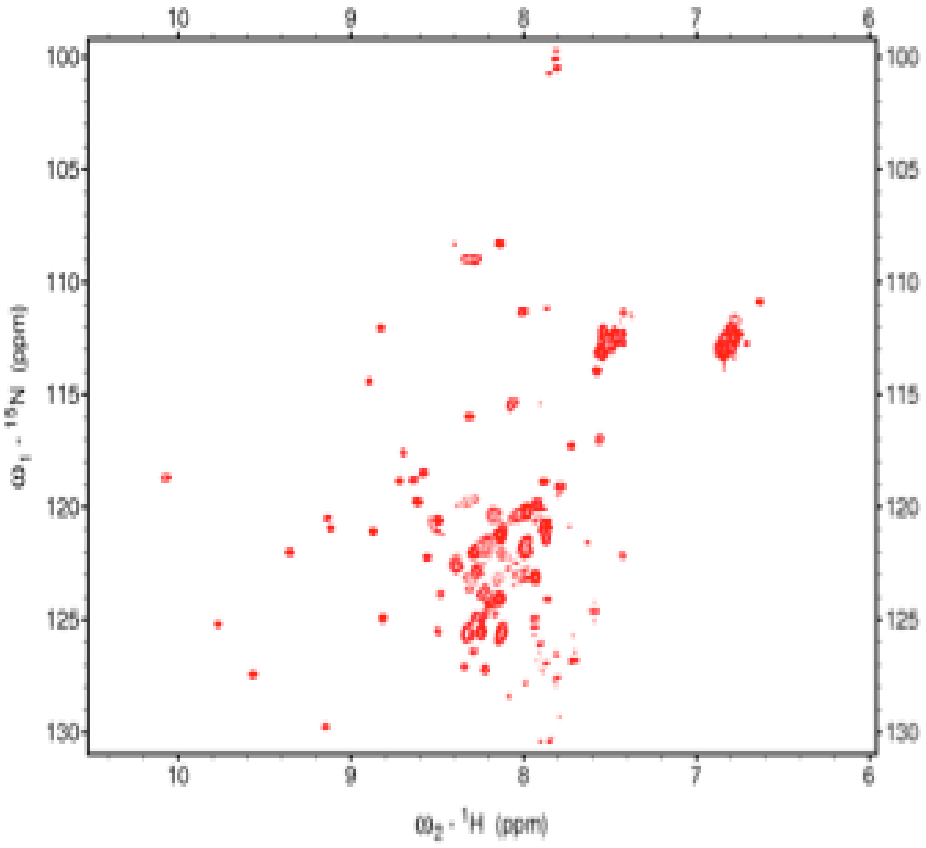


Figure 25 ^{1}H-^{15}N HSQC spectrum of 0.5 mMWAP

Studies on the Stability of the WAP domain of Anosmin-1

DSC experiments are able to measure and compare the thermal stability of the WAP domain in the presence and absence of heparin. The profiles in Figure 26 depict the melting temperatures (Tm), the temperature at which 50% of the protein population exits in its folded conformation while the rest is unfolded, of apo WAP and heparin bound WAP. WAP is only marginally stabilized in the presence of heparin, which can be observed in the slight increased T_m of 3 °C.

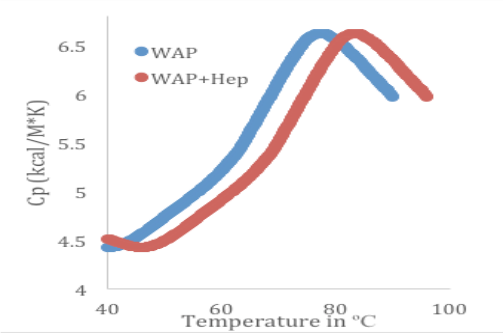


Figure 26 DSC Profiles of WAP in the presence and absence of heparin

The serine protease trypsin cleaves the peptide bond at the carboxyl side of the amino acids lysine and arginine. As indicated by the performed ITC experiments, WAP's interaction with heparin is of electrostatic nature. Therefore, the positively charged residues arginine and lysine of WAP are assumed to bind to the negatively charged heparin. Consequently, they are masked by heparin and protected from the proteolytic degradation by trypsin. As expected, heparin shields the trypsin digestion sites resulting in a faster digestion of WAP in the absence of heparin than in its presence (Figure 27).

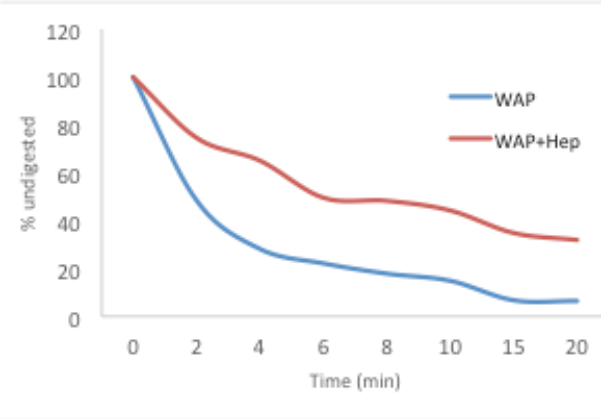


Figure 27 Limited Trypsin Digestion of WAP in the presence and absence of heparin

5.5. Conclusion

Unfortunately, the heterologous expression of full-length Anosmin-1 as well as it shortened constructs in *Pichia pastoris* was unsuccessful. A potential pitfall in the method could be that it is not known, if the linearized DNA encoding for Anosmin-1 was integrated into the yeast genome. Even though colonies grew upon antibiotic selection pressure, it only confirms the presence of the plasmid in the cell. Furthermore, PCR experiments with gene specific primers for Anosmin-1 showed amplification, which confirms the presence of the DNA in the cell. In order to shed light on the question of proper incorporation of the gene of interest into the host genome, PCR experiments with primers that anneal up- and downstream of the Anosmin-1 gene within the yeast genome need to be performed.

In another attempt, collaborators at the Department of Plant Science, University of Arkansas, explored the possibility of expressing Anosmin-1 in tobacco plants. Again, the transient expression of the protein of interest was unsuccessful due to cloning issues. Although Anosmin-1 was cloned into the *E.coli* vector, the agro bacterium rejected the DNA. Possible reasons could be RNA or protein impurities, unfavorable secondary structures within the foreign plasmid DNA, or the growth phase of the competent agrobacterium cells that were prepared for electroporation[30,31].

Expression in mammalian cells could be the solution for the production of full-length Anosmin-1. Studies performed in the past employed Chinese Hamster ovary (CHO) cells[13], D2 Schneider cells[32] that both only yielded little amounts of protein (µg quantities). Therefore, the human embryonic kidney cell line HEK293, which has been shown to be a suitable expression host due to more effective transfection rates and high protein yields, might be the appropriate host for subsequent characterization studies[33].

5.6. References

1. Murcia-Belmonte, V.; Esteban, P. F.; Garcia-Gonzalez, D.; de Castro, F., Biochemical dissection of anosmin-1 interaction with FGFR1 and components of the extracellular matrix. *J. Neurochem.* **2010,** *115* (5), 1256-1265.

2. Kallmann, F.; Schoenfeld, W.; Barrera, S., The genetic aspects of primary eunuchoidism. *Am J Ment Defic* **1944,** *XIVIII*, 203-36.

3. Choy, C.; Kim, S.-H., Biological actions and interactions of anosmin-1. *Front. Horm. Res.* **2010,** *39* (Kallmann Syndrome and Hypogonadotropic Hypogonadism), 78-93.

4. Villanueva, C.; de Roux, N., FGFR1 mutations in Kallmann syndrome. *Front. Horm. Res.* **2010,** *39* (Kallmann Syndrome and Hypogonadotropic Hypogonadism), 51-61.

5. Hu, Y.; Tanriverdi, F.; MacColl, G. S.; Bouloux, P.-M. G., Kallmann's syndrome: molecular pathogenesis. *Int. J. Biochem. Cell Biol.* **2003,** *35* (8), 1157-1162.

6. Dode, C.; Hardelin, J.-P., Kallmann syndrome: fibroblast growth factor signaling insufficiency? *J. Mol. Med. (Heidelberg, Ger.)* **2004,** *82* (11), 725-734.

7. Schwanzel-Fukuda, M.; Bick, D.; Pfaff, D. W., Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) syndrome. *Brain Res Mol Brain Res* **1989,** *6* (4), 311-26.

8. Chung, W. C. J.; Tsai, P.-S., Role of fibroblast growth factor signaling in gonadotropin-releasing hormone neuronal system development. *Front. Horm. Res.* **2010,** *39* (Kallmann Syndrome and Hypogonadotropic Hypogonadism), 37-50.

9. Gonzalez-Martinez, D.; Kim, S.-H.; Hu, Y.; Guimond, S.; Schofield, J.; Winyard, P.; Vannelli, G. B.; Turnbull, J.; Bouloux, P.-M., Anosmin-1 modulates fibroblast growth factor receptor 1 signaling in human gonadotropin-releasing hormone olfactory neuroblasts through a heparan sulfate-dependent mechanism. *J. Neurosci.* **2004,** *24* (46), 10384-10392.

10. Hu, Y.; Gonzalez-Martinez, D.; Kim, S.-H.; Bouloux, P. M. G., Cross-talk of anosmin-1, the protein implicated in X-linked Kallmann's syndrome, with heparan sulphate and urokinase-type plasminogen activator. *Biochem. J.* **2004,** *384* (3), 495-505.

11. Kim, S. H.; Hu, Y.; Cadman, S.; Bouloux, P., Diversity in fibroblast growth factor receptor 1 regulation: learning from the investigation of Kallmann syndrome. *J. Neuroendocrinol.* **2008,** *20* (2), 141-163.

12. Hu, Y.; Bouloux, P.-M., Novel insights in FGFR1 regulation: lessons from Kallmann syndrome. *Trends Endocrinol. Metab.* **2010,** *21* (6), 385-393.

13. Cariboni, A.; Pimpinelli, F.; Colamarino, S.; Zaninetti, R.; Piccolella, M.; Rumio, C.; Piva, F.; Rugarli, E. I.; Maggi, R., The product of X-linked Kallmann's syndrome gene (KAL1) affects the migratory activity of gonadotropin-releasing hormone (GnRH)-producing neurons. *Hum. Mol. Genet.* **2004,** *13* (22), 2781-2791.

14. Robertson, A.; MacColl, G. S.; Nash, J. A. B.; Boehm, M. K.; Perkins, S. J.; Bouloux, P.-M. G., Molecular modelling and experimental studies of mutation and cell-adhesion sites in the fibronectin type III and whey acidic protein domains of human anosmin-1. *Biochem. J.* **2001,** *357* (3), 647-659.

15. Bulow, H. E.; Berry, K. L.; Topper, L. H.; Peles, E.; Hobert, O., Heparan sulfate proteoglycan-dependent induction of axon branching and axon misrouting by the Kallmann syndrome gene kal-1. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (9), 6346-6351.

16. Dvorak, P.; Dvorakova, D.; Hampl, A., Fibroblast growth factor signaling in embryonic and cancer stem cells. *FEBS Lett.* **2006,** *580* (12), 2869-2874.

17. Hung, K.-W.; Kumar, T. K. S.; Kathir, K. M.; Xu, P.; Ni, F.; Ji, H.-H.; Chen, M.-C.; Yang, C.-C.; Lin, F.-P.; Chiu, I.-M.; Yu, C., Solution Structure of the Ligand Binding Domain of the Fibroblast Growth Factor Receptor: Role of Heparin in the Activation of the Receptor. *Biochemistry* **2005,** *44* (48), 15787-15798.

18. Hu, Y.; Guimond, S. E.; Travers, P.; Cadman, S.; Hohenester, E.; Turnbull, J. E.; Kim, S.-H.; Bouloux, P.-M., Novel mechanisms of fibroblast growth factor receptor 1 regulation by extracellular matrix protein anosmin-1. *J. Biol. Chem.* **2009,** *284* (43), 29905-29920.

19. Ornitz, D. M., FGFs, heparan sulfate and FGFRs: complex interactions essential for development. *BioEssays* **2000,** *22* (2), 108-112.

20. Esteban, P. F.; Murcia-Belmonte, V.; Garcia-Gonzalez, D.; de Castro, F., The cysteine-rich region and the whey acidic protein domain are essential for anosmin-1 biological functions. *J. Neurochem.* **2013,** *124* (5&6), 708-720.

21. Cavagnero, S.; Zhou, Z. H.; Adams, M. W. W.; Chan, S. I., Unfolding Mechanism of Rubredoxin from Pyrococcus furiosus. *Biochemistry* **1998,** *37* (10), 3377-3385.

22. Eidsness, M. K.; Richie, K. A.; Burden, A. E.; Kurtz, D. M., Jr.; Scott, R. A., Dissecting contributions to the thermostability of Pyrococcus furiosus rubredoxin: β-sheet chimeras. *Biochemistry* **1997,** *36* (34), 10406-10413.

23. Lovenberg, W.; Sobel, B. E., Rubredoxin; a new electron-transfer protein from Clostridium pasteurianum. *Proc. Natl. Acad. Sci. U. S. A.* **1965,** *54* (1), 193-9.

24. Harper, S.; Speicher, D. W., Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2011,** *681* (Protein Chromatography), 259-280.

25. Garcia-Fruitos, E., Inclusion bodies: a new concept. *Microb. Cell Fact.* **2010,** *9*, No pp given.

26. Rodriguez-Carmona, E.; Cano-Garrido, O.; Seras-Franzoso, J.; Villaverde, A.; Garcia-Fruitos, E., Isolation of cell-free bacterial inclusion bodies. *Microb. Cell Fact.* **2010,** *9*, No pp given.

27. Singh, S. M.; Panda, A. K., Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **2005,** *99* (4), 303-310.

28. Anon, Protein purification: Inclusion bodies. *BioTechniques* **2013,** *54* (2), 75-76.

29. Tsumoto, K.; Ejima, D.; Kumagai, I.; Arakawa, T., Practical considerations in refolding proteins from inclusion bodies. *Protein Expression Purif.* **2003,** *28* (1), 1-8.

30. Glick, B. R.; Thompson, J. E., *Methods in Plant Molecular Biology and Biotechnology*. CRC Press, Inc: Boca Raton, Florida, 1993.

31. McCormac, A. C.; Elliott, M. C.; Chen, D. F., A simple method for the production of highly competent cells of Agrobacterium for transformation via electroporation. *Mol. Biotechnol.* **1998,** *9* (2), 155-159.

32. Hu, Y.; Sun, Z.; Eaton, J. T.; Bouloux, P. M. G.; Perkins, S. J., Extended and Flexible Domain Solution Structure of the Extracellular Matrix Protein Anosmin-1 by X-ray Scattering, Analytical Ultracentrifugation and Constrained Modelling. *J. Mol. Biol.* **2005,** *350* (3), 553-570.

33. Suen, K. F.; Turner, M. S.; Gao, F.; Liu, B.; Althage, A.; Slavin, A.; Ou, W.; Zuo, E.; Eckart, M.; Ogawa, T.; Yamada, M.; Tuntland, T.; Harris, J. L.; Trauger, J. W., Transient expression of an IL-23R extracellular domain Fc fusion protein in CHO vs. HEK cells results in improved plasma exposure. *Protein Expression Purif.* **2010,** *71* (1), 96-102.