Department of Biochemistry and Structural Biology

Protein Stability, Folding and Design: Stabilization of scFv Protein through its reconstitution using a split GFP biosensor in vivo

Master Thesis

Kalyani Sanagavarapu

2012/09

Supervisor: Sara Snogerup Linse
Co-Supervisors: Tanja Weiffert, Jonas Persson
Examinator: Sindra Peterson Årsköld

Table of Contents

1. ABSTRACT

General rules for rational design as well as prediction of tertiary structure and functionality of a protein can be described by investigating the interactions and the role of particular amino acids in protein structure. Mutagenesis has been used commonly to generate stable variants, with an ultimate goal to unravel the rules of protein stability and folding. Besides, reconstitution of dissected proteins has been used as well as an approach to find variants of particular proteins with increased affinity which could lead ultimately to enhancement of stability. In this project a random library of a hapten specific scFv, Anti Fluoroscien IsoThioCyanate (scFv) dissected into the fragments 1-124 (Heavy chain) and 125-246 (Light chain) was interrogated in order to find variants with improved affinity to be tested in further studies for stability enhancement of the corresponding intact protein variants. The split GFP system, a genetically codified biosensor, was used as a method to detect in vivo reconstitution of scFv (Heavy chain 1-124 + Light chain 125-246).

Firstly, reconstitution of a single chain antibody (scFv) fragment 1-124 (Heavy chain) and 125-246 (Light chain) was detected. Secondly, a random library of the Light chain fragment 125-246 cloned into the GFP system was screened to find variants with higher fluorescence intensity than WT Light chain. An increase in fluorescence is suggested to arise from increased affinity which in turn could be used to select for stabilized intact variants. However we failed to detect green fluorescence. This may be due to problems in the expression of one of the partners (heavy chain-CGFP) or steric constraints and hence we were not able to screen any high affinity mutants. Various suggestions for improving the expression of the protein or relieving steric constraints are discussed here. If these problems are solved, libraries will be screened for the possible stabilizing role of the found substitutions. This can be discussed in terms of establishment of favorable hydrophobic interactions, stabilization of secondary structure and indirectly destabilization of the unfolded structure. The insight into the interactions and roles played by specific amino acids can be used to understand protein design of other proteins.

2. INTRODUCTION

Proteins are highly complex systems performing or supporting nearly all the mechanisms and reactions present in the life phenomena. The function of a protein is reliant on its tertiary structure, which is composed of particular folded structures in space in regular patterns. The folding is guided by the primary structure i.e., the sequence of amino acids, which is ultimately established by the DNA belonging to the particular cell through the genetic code, and various splicing events.

The development of molecular biology techniques such as DNA recombinant technology have allowed us to study in more detail the properties of proteins through modification of the sequence of particular genes and further cloning and expression.

In this project, the split GFP system, a biosensor is used as a method to detect in vivo reconstitution of dissected single chain antibody fragment (scFv) obtained from a synthetic library, n-CoDeR (Söderlind, 2000). The variable regions of heavy and light chains fused together via a linker constitute a scFv protein. A random library of Light chain 125-246, one of the dissected fragments of scFv will be interrogated in order to find high affinity variants that may serve as templates for producing more stable variants of intact protein.

2.1 Protein Interactions

Understanding and solving the fundamentals of protein interactions is crucial for the future development of life sciences and biotechnology because proteins play many important roles in cell processes and biotechnological applications. It is a delicate balance of non-covalent inter- and intra-molecular interactions such as van der Waals, columbic, hydrophobic effect and hydrogen bonding that ultimately defines the biological and technological task of the proteins. Intra-molecular protein interactions are established between amino acids and govern important properties such as folding mechanisms as well as structure, and stability. Resolving the interactions taking place during these processes might help us to rationally design and modify of protein. It will also help us to predict the tertiary structure and functionality of a protein (Malakauskas, 1998). In addition, inter-molecular interactions are in operation in ligand binding. Intermolecular interactions modulate kinetic properties, lead to inactivation of proteins, drive substrate channeling, form new binding sites and regulate substrate specificity.

Also, following the advances toward deciphering the entire proteome, it is essential to discover the importance of inter-molecular interactions for the functionality of thousands of sequences of proteins codified in the genomes and to know the roles that all these interactions play in diseases as cancer and Alzheimer and biological processes such as DNA replication,

transcription, translation etc. Getting insight in protein-protein interactions will allow us to understand the life phenomena and our ability to propose therapies to diseases due to misfunctional interactions of proteins [Berggård, 2007; Kerppola, 2008].

2.2 Methods to Detect Protein Interactions

In order to study and characterize the impact of interactions on the functionality of a protein, there is a need for a reliable way to identify protein-protein interactions. Several methods, for example, analytical ultracentrifugation, NMR and optical spectroscopy, isothermal titration calorimetry, surface plasmon resonance, affinity chromatography, immunoprecipitation, affinity blotting, cross linking, phage display, bimolecular complementation and yeast two-hybrid system [Hebert, 2006; Lalonde, 2008; Song, 1989] are in use. Recently several fluorescence-based techniques such as Bimolecular Fluorescence complementation [Kerppola, 2008; Villalobos, 2007] have been introduced.

2.3 Bimolecular Fluorescence Complementation

Bimolecular fluorescent complementation (BiFC) is a fluorescent-based technique that allows detection of protein-protein interactions in living cells. It can be used to determine sub-cellular localization of the interacting proteins, and if it changes over time, without requiring addition of external agents. BiFC is based upon reconstitution of split non-fluorescent GFP variants, to form a fluorescent fluorophore [Ghosh 2000; Hu 2002]. The technique has become increasingly popular due to its simplicity, ease of use, and the capability to carry out experiments with regular epifluorescence or confocal laser scanning microscopes (CLSMs). Other complementation assays need exogenous fluorogenic or chromogenic agents, potentially perturbing the cells. This was first demonstrated for subtilisin [Johnsson, 1994], and was later accomplished for other proteins like β-Galactosidase [Rossi, 1997] and dihydropholate reductase [Pelletier, 1998]. BiFC can also be used to derive kinetic information and further characterize the protein-protein interaction by visualizing multiple variants of a library at the same time. Moreover the method can be used to detect interactions between sub populations of each protein and multicolor analysis allows simultaneous visualization of multiple protein complexes formed with a shared component. Other visualization methods as FRET require high levels of protein expression and placing the two fluorophores within 100 Å. Some of the disadvantages of BiFC are that it does not allow real time detection of interactions in part because there is a delay between the interaction of the fused proteins and fluorescence due to low rate of fluorophore formation. Apart from that, BiFC is useful to probe interactions of many structurally different proteins inside many different types of cells. One of the most

popular BiFC methods that was used in this thesis was the split Green Fluorescence Protein (split GFP) [(Magliery, 2005)].

2.4 Green Fluorescent protein

GFP was first described by Shimomura [(Shimomura, 1962)] and later Chalfie cloned and used it as reporter of protein expression patterns [(Chalfie, 1994)]. The GFP gene has all the information necessary for development of a mature and functional chromophore after posttranslational synthesis and thus no further processes by enzymes are needed. GFP is 238 amino acid residues in length. It is composed of 11 β-strands folded into a β-barrel with an α-helix running up through the axis of the barrel (Figure 1). The chromophore is located in the α-helix and is formed by the residues Ser-Tyr-Gly 65-67 in the native protein. It is matured after a series of reactions involving O_2 from air and ready to be excited and emit its inherent fluorescence. The wild type GFP has a major excitation peak at 395 nm, a minor peak at 475 nm, and emission at 508 nm. The ability to fluoresce without any additional chemical or substrate has made GFP a perfect tracer of proteins [(Tsien, 1997)].

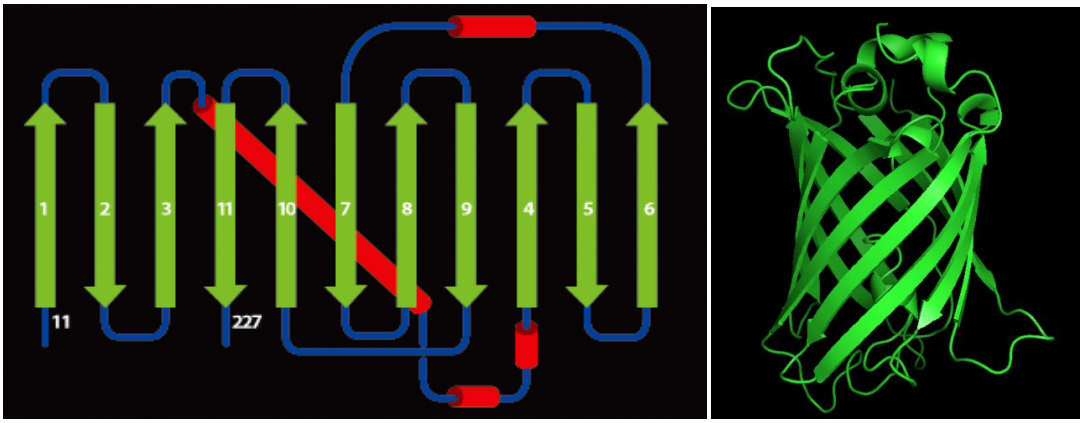


Figure 1: Topology of the green fluorescent protein (left) and tertiary structure of GFP using PDF file 1EMB (right).

2.5 Split Green Fluorescent Protein System

Split GFP system is part of the BiFC methods and was first described by the group of Lynn Reagan. Split GFP method is based on, at the genetic level, the dissection of GFP between residues 157 and 158 to generate two fragments. The affinity between GFP fragments seems to be too low for reconstitution and chromophore maturation under normal conditions which is observed upon co-expression of the fragments. However, these two fragments of GFP can be fused separately to two different proteins (or other molecules) or to fractions of one protein whose interaction promotes assembly and folding of GFP and thereby chromophore

development. Complex formation between the fusion partners raises the effective concentration of the GFP fragments for one another favoring their association and folding and the fluorescence emission is reached. Hence the method can be used to detect protein-protein interactions or protein fragment-fragment recomplementation [Ghosh, 2000, Magliery, 2005] (Figure 2).

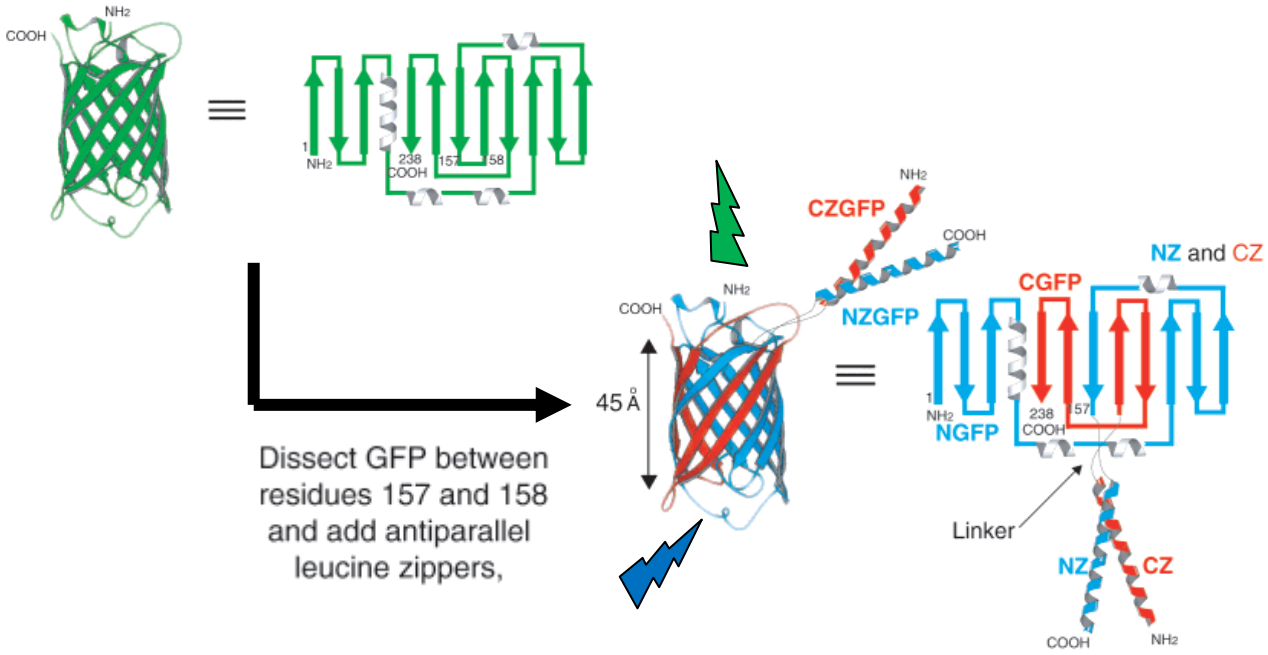


Figure 2: Application of the Split GFP system to detect reassembly of an antiparallel Leucine zipper by Regan group (adapted from Ghosh, et. al., 2000)

The most important advantages of the split GFP system are that no special exogenous reagents have to be added to detect fluorescence. Also, GFP can express and mature in almost every cell type and subcellular structure. Furthermore, it can be used in bacteria, nuclear importation of the proteins is not required and most cells do not have significant fluorescent background at the excitation wavelength used. The method has been used to detect and trap transient protein-protein interactions, to identify unknown interactive protein partners and protein localization in subcellular structures. This has been proved in bacteria or multicellular organisms such as yeast and plants [Wilson, 2004, Magliery, 2005; Barnard, 2008; Magliery, 2008; Sarkar, 2008, Lindman, 2009]. The fluorescence acquisition is a slow process involving GFP maturation with rate-limiting oxidation of the chromophore and generally is detected after two or three days. The reassembly multistep process is affected by properties of the fused proteins such as solubility and expression levels. However, since the reassembly process is irreversible it can be used to detect weak and transient interactions.

The GFP variant used in this study has the six mutations F64L, S65C, Q80R, Y151L, I116T and K238N in comparison to wild type GFP. These mutations give the protein a single excitation peak at 475 nm and emission at 505 nm. The green bright fluorescence is arising

from the mature chromophore core comprising the residues C65-Y66-G67 (Topell, 1999; Abedi, 1998).

2.6 In vivo protein stabilization based on fragment complementation and split-GFP system

Protein stabilization can be achieved through in vivo screening based on the thermodynamic linkage between protein folding and fragment complementation. Several approaches have been used to stabilize proteins like formulation, screening methods employing a combination of phage display and increased protease resistance, directed evolution and rational design etc (Jorgensen L, 2009; Sola RJ, 2009; Walle CF, 2009; Wunderlich, 2005; Eijsink, 2005; Eijsink, 2004).The split GFP system is found suitable to derive protein variants with enhanced stability via the correlation between effects of mutations on the stability of the intact chain and the effects of the same mutations on the affinity between fragments of the chain (Lindman, 2010; Bergard T, 2001; Sanz R J, 1995; Xue WF, 2006; Carey J, 2007). This implies that proteins may be stabilized using a method based on protein reconstitution from fragments and screening of a fragment library for enhanced affinity.

2.7 FITC8

FITC8 is a single chain antibody fragment (scFv) which binds to hapten, Fluorescein isothiocyanate. A single-chain antibody is a fusion protein of the variable regions of the heavy (V_H) and light chains (V_L) of immunoglobulins, connected with a short linker peptide of 10 to about 25 amino acids (Figure 4A). The linker is usually rich in glycine for flexibility, as well as serine or threonine for solubility, and can either connect the N-terminus of the V_H with the C-terminus of the V_L, or vice versa (Tsumoto K, 1994) (Figure 3). FITC8 has the linker connecting N-terminus of the V_H with the C-terminus of the V_L. These proteins retain the specificity of the original immunoglobulin, despite removal of the constant regions and the introduction of the linker. (Bird, 1988)

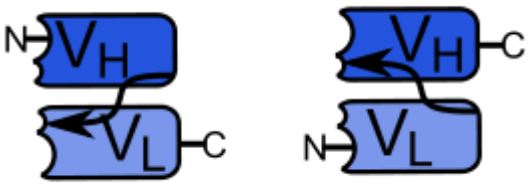


Figure 3: Possible conformations of a scFv molecule. Linker connects the N-terminus of the VH with the C-terminus of the VL, or vice versa.

FITC8, a single chain antibody fragment serves as a model of high affinity binder recognizing a hapten. FITC8 was originally obtained from a synthetic library, n-CoDeR [(Söderlind, 2000)] that had been constructed by shuffling human complementarity-determining region (CDR) sequences of different origins into a single framework consisting of the human IGHV3-23 and IGLV1-47 genes. Based on this framework 14 clones including FITC8 have been shown to be highly stable, well expressed in Escherichia coli and displayed easily on phage. [(Griffiths, A. D, 1994; Ewert S, 2003; Steinhauer C, 2002)] FITC8, like hapten-binders in general, displays a cavity in its antigen-binding site into which the antigen binds.

The solved structure of FITC8 is known (personal communication with Lena Danielsson) although it is not published. However, a scFv molecule is composed of two chains- a heavy chain and a light chain. The antigen binding site (paratope) is largely composed of residues in the complementarity determining region (CDR) loops, three each in the heavy and light chains. A beta sheet framework in the form of an immunoglobin fold brings the six CDRs together enabling them to interact jointly with the antigen (Figure 4B). CDRs determine the protein affinity and specificity for specific antigens [(Pandaln, 1994)]. FITC8 is formed of relatively short fragment of 246 residues.

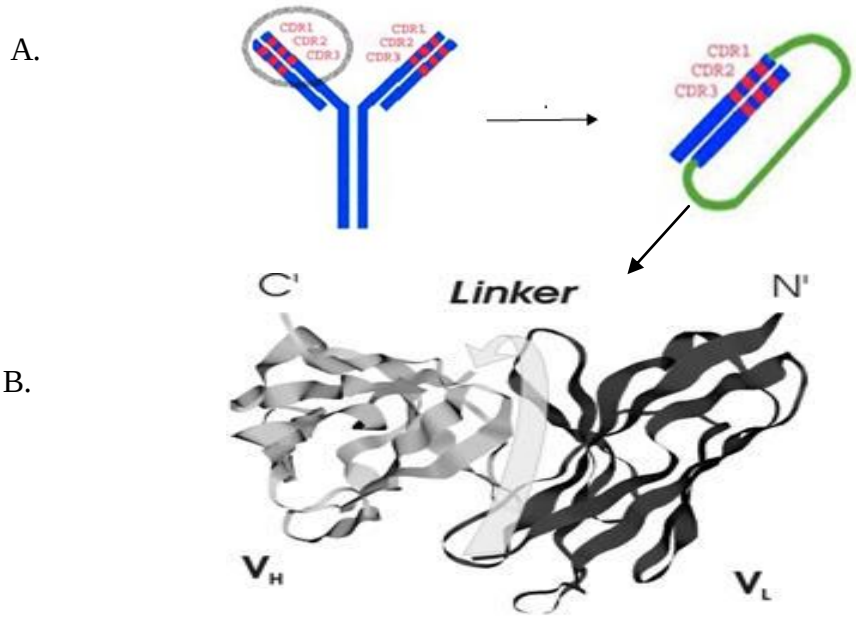


Figure 4 A: scFv, a fusion protein derived from the whole antibody; B: Structure of a single-chain Fv fragment of antibodies. Variable fragments over a flexible and soluble peptide linker.

2.8 Protein Reconstitution

In order to examine the role of the amino acid sequence in protein structure and folding several methods have been developed. Protein reconstitution or protein reassembly is one such method. Protein reconstitution is the spontaneous reassembly of fragments of a dissected

protein into the native-like tertiary structure, even when the isolated fragments show little or no similarity to the native structure [(Carey, 2007)]. The important idea behind protein reconstitution is that the chain connectivity is not necessary for correct folding in some cases. Thus it is an interesting way to probe the relationship between protein sequence on the one hand and structure or folding on the other. This can ultimately be used to understand factors behind reconstitution of proteins into their functional complexes. Some classical examples of reconstituted proteins are ribonuclease A, β-galactosidase, cytochrome C, protein G β1 and monellin [(Richards, 1958; Ullman, 1967; Hantgan, 1977; kobayashi, 1995, Xue, 2004).]

3 MATERIALS AND METHODS

3.1 Agarose gel electrophoresis

DNA Samples were analyzed using agarose gel electrophoresis according to the standard procedure.
1µL sample was mixed with 1µL 1X loading dye (10 mM Tris-HCl (pH 7.6),
0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 60 mM EDTA) and made upto 10 µL with sterile milliQ water. The prepared samples are loaded onto agarose gel. The gel was run for approximately 1 hour at 80V in TBE buffer (5X stock solution: 2 M Tris base, 2 M Boric acid, 0.5 M EDTA (pH 8.0) solution). 0.9% agarose gel was used for plasmid DNA and 1.5% agarose gel was used for PCR DNA.Agarose mix consists of 3 parts of Metaphor, 2 parts of Seakem and 1 part of NuSieve and all the components are purchased from Lonza. Either 1kb plus or 100bp plus DNA Ladders purchased from Fermentas were used as standard molecular weight markers.

3.2 SDS-PAGE

Samples were analyzed using tricine gel electrophoresis as originally described by H. v. J. G. Schägger.

Samples were mixed with 2x tricine sample buffer (450 mM Tris HCl pH 8.45, 12% glycerol, 4% SDS, 0.0025% Coomassie blue, 0.0025% phenol red), incubated for 10 min at room temperature and then loaded on a 10-20% tricine gel (Invitrogen). The gel was run for 3 h at 80 V in 1x running buffer (100 mM Tris, 100 mM Tricine, 0.1% SDS, pH 8.3), stained in staining solution (0.25% Coomassie Brilliant blue, 40% ethanol, 10% acetic acid) for at least 4 h, destained in destaining solution (30% ethanol, 7% acetic acid) and finally scanned. BenchMark™ Low-range unstained and prestained protein standard were from Novagen.

3.3 Media, stock solutions, bacterial strains and plasmids

All stock solutions and media were sterilized for 20 min at 121°C prior to use.

Chemicals used for preparation of growth media were purchased from BD:

- Media:

Luria broth medium (LB): 1% tryptone, 1% NaCl, 0.5% yeast extract
LB agar: LB medium with 1.5% agar

- Stock solutions for auto-inducing medium:

10 mM of Isopropyl β-D-1 galactopyranoside (IPTG). LB agar plates with a final concentration of 10 µM IPTG and a selective antibiotic resistance is used as auto-inducing medium.

- Stock solutions for antibiotics:

All antibiotics were prepared as 1000x stock solutions, filter sterilized and stored at -20°C.
50 mg/mL ampicillin (Amp) in sterile MilliQ water
30 mg/mL chloramphenicol (Cam) in 95% ethanol
35 mg/mL kanamycin (Kan) in sterile MilliQ water

- *E.coli* strains

E.coli ER2566
E.coli BL21*pLysS(DE3)
E.coli BL21 GOLD
DH5 alpha

Fresh cultures of each strain were made chemically competent. The *E.coli* cell culture was spread onto LB plates. Single colony was taken from the plate, amplified in 5ml culture. 1% of this culture is inoculated into 250 ml of LB medium. At an OD of 0.3 – 0.6, the cells are harvested at 6000 rpm for 5 minutes at 40°C. Harvested cells are re-suspended in 50 ml of 100 mM MgCl2 and centrifuged at 4000 rpm for 5 minutes at 40°C. The pellet is re-suspended in 100 ml of 100 mM CaCl2 and incubated on ice for 20 to 30 minutes followed by centrifugation at 4000 rpm for 5 minutes at 40°C. Under sterile conditions, the pellet is re-suspended in 3-4 ml of 85 mM CaCl2/15% glycerol mixture and stored as 40 µL aliquots at -80°C. The pLysS vector in *E.coli* BL21*pLysS carries the resistance gene for Cam. Hence preparation of competent cells was carried out at final concentrations of 30 µg/mL of Cam. No antibiotic resistance for *E.coli* ER2566 and *E.coli* BL21 GOLD. The preparation of the competent cells, *E.coli* ER2566 harboring a pET9a-CGFP construct was carried out at a final concentration of 100 µg/mL of kanamycin.

- Plasmids:

pQLinkN subcloned scFv fragments plasmids were synthesized by Genscript company. pET11a and pET9a plasmids were supplied by Regan L and Genscript respectively. Each plasmid was amplified in *E.coli* ER2566 and purified utilizing GeneJET™ Plasmid Miniprep Kit (Fermentas).

3.4 DNA and protein sequences of scFv

Fluoroscien isothiocyanate, FITC8 is the scFv used in this part of project. The wild type nucleotide sequence of FITC8 is available from the reference, Helena Persson, J. Mol. Bio 2006. The sequence is modified to contain *E.coli* preferred codons for a better expression in *E.coli* cells.

3.5 Cloning of Dissected GFP-scFv Wild Type Fragments into pQLinkN plasmids

3.5.1 Design of Split GFP fused with scFv fragments subcloned into pQLinkN plasmids

The idea of the design is to split scFv into two chains and fuse the fragments to the two halves of GFP plasmid.

scFv conformation N'-Heavy chain—(LINKER) $_{15\ \text{aminoacids}}$—Light chain-C' has been split to two chains

a) Heavy chain-(Linker) $_{7\ \text{amino acids}}$ = Heavy chain 1-124 = HC
b) (Linker) $_{8\ \text{amino acids}}$-Light chain = Light chain 125-246 = LC

Each chain is fused to split GFP plasmid (Figure 5). Heavy chain is fused to CGFP (GFP 158 - 238) via the linker residues, GGGGSGG (HC-CGFP) and Light chain is linked to NGFP (GFP 1- 157) after the linker GGSGGGGS (NGFP-LC). The fusion constructs are subcloned into pQLinkN plasmids between BamHI and HindIII sites available in multiple cloning site flanked by two LINK sequences.

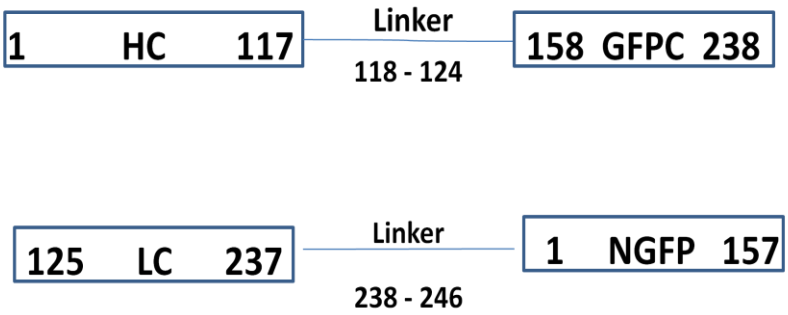


Figure 5: Cartoon representing the design of dissected scFv fragments (HC and LC) fused to split GFP system

pQLink plasmids are vectors for co expression of unrestricted number of proteins (C Scheich, 2007). The vectors contain two LINK sequences (Alexandrov A, 2004) that flank the expression cassette of promoter, multiple cloning site and a transcriptional terminator. LINK1 contains a PacI restriction site and LINK2 has a SwaI and a PacI site (Figure 6). The LINK sequences allow insertion of a PacI fragment of one plasmid at the SwaI site of another plasmid by Ligation

independent cloning (LIC). Thus a co expression plasmid is facilitated by LIC. The problem of imbalanced expression levels of two different plasmids transformed into a single bacterial cell can be avoided by using a co expression plasmid which can express multiple constructs in a single vector. Here we used pQLinkN plasmid which serves the purpose of accommodating and expressing the two constructs (HC-CGFP and NGFP-LC) in a single plasmid.

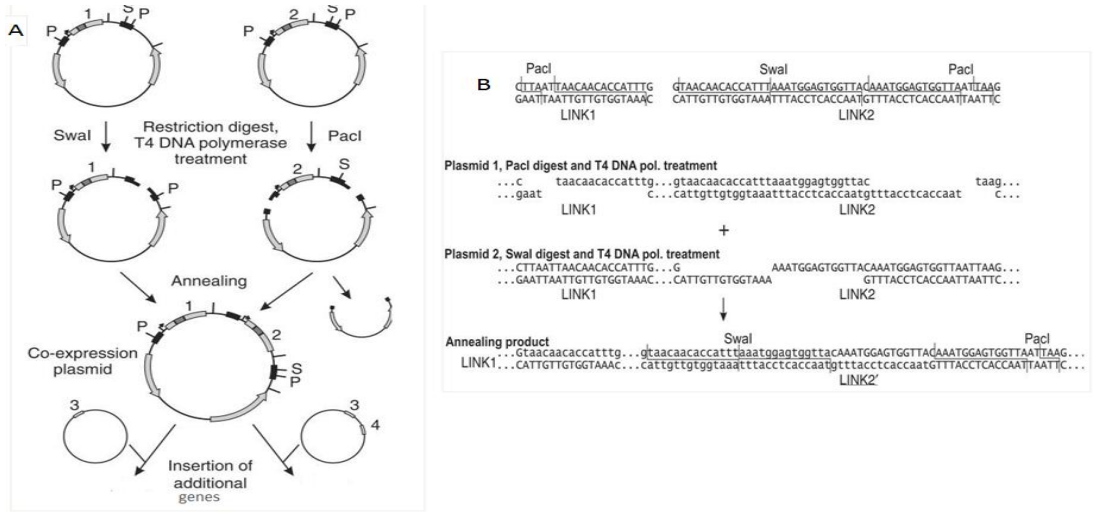


Figure 6 A) Construction of a co-expression plasmid from two pQLink plasmids with two different cDNA inserts labeled 1 and 2. The resulting plasmid can accept additional inserts, labeled 3 and 4. S = SwaI, P = PacI. B) The Link sequences and their digestion and annealing. The lines indicate overhangs generated by restriction digest and T4 DNA polymerase treatment. Two plasmids, identified by upper and lower case nucleotide codes, are digested and annealed, leading to a product with a LINK1 and a LINK2 sequence, slightly larger than the original LINK2 sequence (C Schiech, 2007).

3.5.2 Construction of co expression plasmid, pQLinkN-NGFP-Light chain-Linker-Heavy chain-CGFP plasmid:

The plasmid pQLinkN-NGFP-LC is digested with SwaI (Fermentas) restriction enzyme for 16 h at 25ºC. The plasmid pQLinkN-HC-CGFP is digested with PacI (Fermentas) for 16 h at 37ºC. SwaI cut plasmid and PacI digested HC-CGFP fragment are purified by GFX PCR, DNA and gel band purification Kit (GE healthcare, Uppsala, Sweden) hereafter called "the GFX kit". 5 µL of each digested product is treated individually with 2 µL of 1.25U LIC qualified T4 DNA polymerase, 2 µL of T4 DNA polymerase buffer, 1µL DTT, 2 µL of dCTP for PacI digested product and dGTP for SwaI digested product. The whole reaction mixture is made to 20 µL with sterile milliQ water and kept for incubation at 25ºC for 30 minutes. Later the tubes are kept at 65ºC for 20 minutes to inactivate the LIC qualified T4 DNA polymerase. PacI and SwaI treated tubes are mixed together and incubated at 16ºC for one hour. In the half way of incubation period, 4 µL of 25mM EDTA is added to the mixture. Later the mixture is heated to 75ºC and kept at 75ºC for one minute. The mixture is cooled to room temperature slowly by moving the tube to the bench. This step helps in annealing the PacI and SwaI

digested products. Hence the multi gene plasmid was constructed (Figure 7). After cooling to room temperature, the LIC mixture is used for transformation.

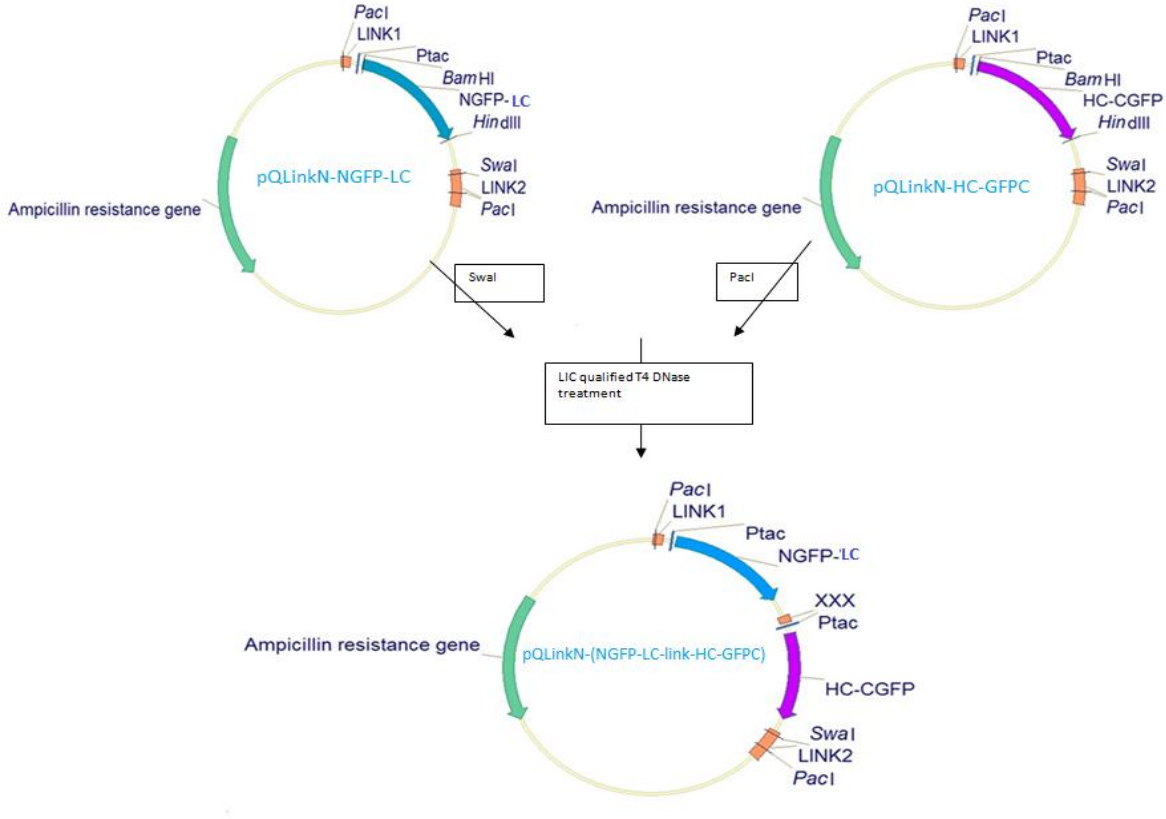


Figure: 7 Construction of (NGFP-LC-link-HC- CGFP)- pQLinkN plasmid.

Individually sub-cloned NGFP-LC (a) and HC-CGFP (b) were digested with SwaI and PacI restriction enzymes respectively. The digested products were mixed together and treated with ligation independent cloning qualified T4 DNase. LIC qualified T4 DNase treatment facilitates ligating the PacI digested fragment into the SwaI digested plasmid. Hence the co-expression plasmid (c).

LIC product is transformed into calcium competent *E.coli* ER2566 cells and plated on LB/agar plates with 50 µg/mL ampicillin. Several colonies obtained on the plate were amplified and purified. Purified plasmids were sequenced to confirm the presence of the whole construct (NGFP-LC-Linker-HC-CGFP). Sequencing is done using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) with the primers: 5'-c aac att ctg gga cac aaa ttg g-3'and 5'-tgc tag ttg aac gct tcc atc ttc-3'.

3.5.3 Preparation of Library of Light chain 125-246 Fragment and Cloning into Vector

In personal communication with Jonas Persson, it was learnt that mutations in light chain are less likely to change the specificity of scFv towards its hapten. However, a change in affinity for the hapten might happen. Hence we wanted to focus on light chain library rather than heavy chain.

A LC random library (LC-lib) was designed and constructed by overlapping PCR. The original design includes splitting of light chain into six degenerate oligo nucleotides (Figure 8) with desired restriction sites flanking the start and stop primers. The degenerate oligo nucleotides for library generation included the incorporation of random mutations at a theoretical mutation rate of 2 percent variants per base distributed in stabilizing and destabilizing mutations (Table 1). Overlapping PCR is carried out by mixing all the six degenerate primers in a single tube. The PCR was carried out using an Expand High Fidelity PCR system (Roche Diagnostics, Indianapolis, IN) in the cycle: 94ºC, 2 min; 10x (94ºC, 15 s; 72ºC, 30 s; 72ºC, 45s); 72ºC, 7 min. The start and stop primers are maintained at 30 µM concentration. The other four primers are tested at four varied concentrations: 2 µM, 5 µM, 12 µM and 30 µM. The PCR products were investigated by agarose gel electrophoresis and purified using GFX kit.

The overlapping PCR product of the DNA library of LC-lib was first digested with SacI for16 h at 37ºC. Later KpnI was added and digested for 16 h at 37ºC. pQLinkN-MNA-NGFP (from Genscript) was also digested in a similar way. Double digested product was treated with alkaline phosphatase. The digestions are followed by purification using GFX kit.

Ligation is carried out at 1:2 molar ratio of vector to insert. 3 µL of ligation product is added to 40 µL of calcium competent ER2566 cells and left on ice for approximately 30 minutes. The mixture is heated at 42ºC for 45 seconds and kept on ice for 10 min before it is plated on LB/agar plate with 50 µg/mL ampicillin and cultured overnight at 37ºC. Several colonies were picked, amplified and plasmids were purified. The plasmids bearing the light chain library were cut with SwaI and a co expression plasmid pQLinkN- NGFP-LC-lib-link-HC-CGFP is constructed by LIC, in a similar way to wild type.

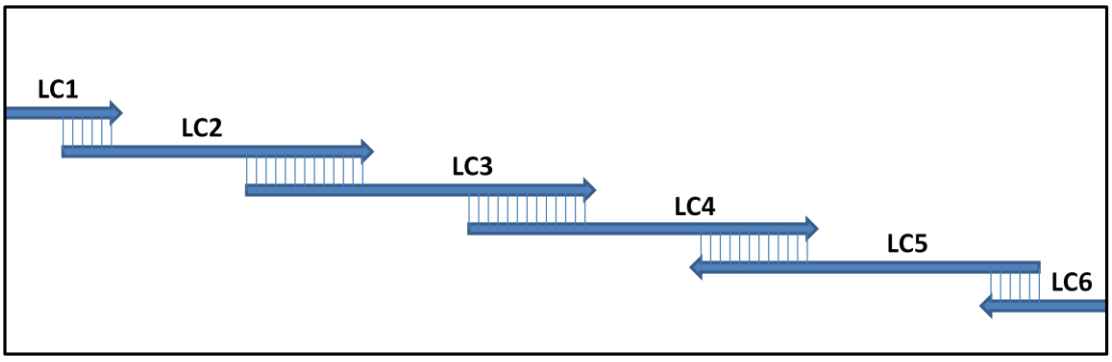


Figure 8: Cartoon representing the design of degenerate oligo nucleotides. Light chain is divided into six overlapping fragments, LC1 – LC6. Overlapping regions between adjacent fragments is marked with lines. PCR of the overlapping degenerate nucleotides generates the random mutant library of LC.

Table 1:

The degenerate oligonucleotides designed for library generation.

Primer LC1 : 5'*GGAACAGT GAG CTC cag tct gtg* 3'

Primer LC2:

5' *GGAACAGT GAG CTC cag tct gtg* ctg act cag cca ccg tca gcg tct ggg acc ccg ggg cag cgc gtc acc atc tct tgc act ggg agc agc tcc aac *atc ggg gca ggt tat gat gta cac* 3'

Primer LC3:

5' *atc ggg gca ggt tat gat gta cac* tgg tat cag cag ctc cca gga acg gcc ccg aaa ctc ctc atc tat *ggc aac aac aat cgg ccg tca gg* 3'

Primer LC4:

5' *ggc aac aac aat cgg ccg tca ggg gtc cct gac cgt ttc tct ggc tcc aag tct ggc acc tca gcc tcc ctg gcc atc agt ggg ctc cgg tcc gag gat gag gct gat tat tac tgc gca gcc tgg gac gac agc*3'

Primer LC5:

5' *GCATATCCGGTACCCTATTA tag gac cgt cag ctt ggt tcc tcc gcc gaa* tac gcg agt tcc act cag *gct gtc gtc cca ggc tgc gca*3'

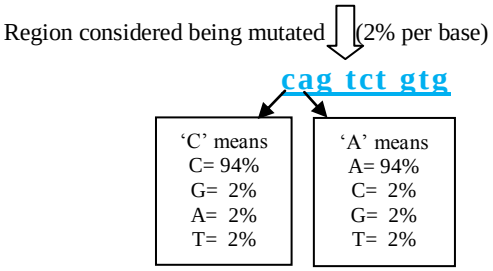
Primer LC6:

5' *GCATATCCGGTACCCTATTA tag ga* 3'

- The underlined regions with blue colored residues of the fragment are considered to have 94% correct and 2% of each other base.

Eg: Primer LC1

5' GGAACAGT GAG CTC cag tct gtg 3'



Representation of mutation frequency is shown for the first two bases of the sequence. Similar mutation rate was followed for all the bases in the underlined regions!!

Letters in light blue highlight and underlined were the variable regions considered for including 2% mutation. Letters in small case black color were the CDR regions. CDR regions were kept constant without any mutation. Letters in upper case black color were the extra nucleotides included for introducing flanking restriction sites which help for sub-cloning the library fragments into pQLinkN plasmid. Italics represent the overlapping regions between the adjacent primers.

3.5.4 Detecting in vivo reassembly of GFP-scFv Wild Type fragments

The multi gene plasmid coding for both the wild type (WT) fused genes NGFP-LC and HC-CGFP was transformed by thermal shock adding 1 µL of each plasmid of the pair to 50 µL of calcium competent *E. coli* ER2566.

The transformed mixture was spread onto plates LB plates containing 100 µg/mL ampicillin and 10 µM of IPTG. The plates were incubated for 16 h at 37ºC and then taken out and left at RT during fifteen days. The appearance and maturation of green fluorescence in the colonies over time was detected and followed placing the plates on a Dark Reader DR45M non-UV blue light transilluminator (Clare Chemical Research, Inc, Dolores, Co) with light emission between 420 and 500 nm. Avoiding any external light, pictures were taken using a digital CCD camera.

3.5.5 Detecting in vivo reassembly of GFP-scFv Light chain lib fragments and screening of mutants

The multi gene plasmids coding for the fused gene NGFP-LC-lib fragments and WT HC-CGFP was transformed by thermal shock to 50 µL of chemical competent *E. coli* ER2566. The transformed mixture was spread onto several LB plates with 50 µg/mL of ampicillin, 10 µM IPTG. The plates were incubated 16 h at 37ºC and then taken out and left at RT during five days. The appearance and maturation of green fluorescence in the colonies over time was detected and followed placing the plates on a Dark Reader to rank the colonies manually.

3.5.6 Expression of scFv in other *E.coli* strains

There could be a limitation for expression of protein in a particular strain. As no green colonies were observed in *E.coli* ER2566, we tried to express GFP fused scFv protein in two other strains. The two strains were BL21*PLysS and BL21 Gold. The wild type and library carrying multi gene pQLinkN plasmid was transformed into two strains and screened in a similar way to *E.coli* ER2566.

3.5.7 SDS-PAGE analysis of expression levels

SDS-PAGE was used to evaluate whether the absence of green fluorescence is due to poor expression or poor reconstitution of fusion protein, due to steric constraints or other problems.

The co expression plasmid containing NGFP-LC and HC-CGFP was transformed into *E.coli* ER2566 and BL21* PLysS. LB/ampicillin plates were used for plating *E.coli* ER2566 transformed cells. LB/ampicillin/chloramphenicol plates were used for BL21* pLysS transformed culture. Colonies obtained were checked for the presence of NGFP-LC-linker-HC-CGFP construct by BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The five plasmids that were confirmed for the presence of construct were transformed into two different cell types selected. An overnight culture from each strain was grown in 5 ml LB in 15 ml falcon tubes for 16 h. 1 ml was inoculated into 100 ml of LB medium in a 500 ml culture flask and cultured in shaking incubator. At an OD of 0.6, the bacterial cells were induced by 0.4 mM IPTG during 4 h of incubation. Seven other culture flasks of LB medium were set up, five with un-induced transformed cells where no IPTG was added, one with un-induced empty cells (cells without the plasmid and no IPTG) and the other with induced empty cells (cells without plasmid induced with IPTG). 1 ml of the medium was collected in an eppendorf tube and cells were harvested at 13000 rpm for 1 min. The rest of the medium is collected in a centrifuge tube and harvested. Pellet is stored in 15 ml falcon tube and stored at -4ºC. Cells harvested in 1.5 ml eppendorf tube were disrupted in 100 µL of 8 M urea and diluted with 200 µL Tris buffer pH 8.0. 10 µL of the sample is mixed with 20 µL of 2X loading buffer. About 10 µL of sample was loaded onto the SDS-PAGE gel and run as described above (**4.2**)

Band pattern obtained was compared among induced empty cells, induced transformed cells, un-induced empty cells and un-induced transformed cells of the two strains. SDS gel was examined for the presence of distinct bands of approximately 23 kDa and 30 kDa Mw in induced transformed cells.

3.5.8 Plating of transformed cells onto LB-Amp plates with varying concentrations of IPTG

As SDS-PAGE gel gives a hint in problems with expression profile of GFP fused scFv fragments, the transformed bacterial cells were checked for its expression at different concentrations of IPTG (1.25, 2.5, 5, 10, 20, 40 and 80 µM). Three IPTG concentrations above and below the standard concentration (10 µM) were tested.

3.6 Cloning and expression of Dissected GFP-scFv Wild Type Fragments into pET plasmids

3.6.1 Constructing pET11a-NGFP-LC and pET9a-HC-CGFP plasmid

A single chain antibody fragment expression and reconstitution was not successful with pQLinkN plasmid for a reason which is still unclear. So we tried to clone the scFv fragments into pET plasmids, an old and successful method employed before.

DNA fragments LC and HC-CGFP were amplified separately through PCR from vector pQLinkN-NGFP-LC plasmid and pQLinkN-HC-CGFP (GenScript USA Inc.). Primers used for amplification of both scFv fragments (Table 2) were particularly designed in order to have specific restriction sites flanking the gene segment for subsequent cloning into vectors pET11a-link-GFP, carrying the residues 1 to 157 (GFPN) of dissected GFP gene (Ghosh, 2000) and pET9a plasmids respectively (Figure 9). The possible spatial conformation of reassembly interaction tested was GFPN-LC + HC-CGFP

Table 2: Primers used to amplify the scFv fragments and cloning into pET plasmids

Construct	Primers	Restriction site
NGFP_LC-125-246 start	5′-GCAGAACAAT*CTCGAG*CCAGTCTGTGCTGACTCAGC-3′	XhoI
NGFP_LC-125-246 stop	5′-GGTAATAT*GGATCC*TTATAGGACCGTCAGCTTGGTTC-3′	BamHI
CGFP-(HC-1-125) start	5′-GCAGAACAAT*CATATG*GAGGTGCAGCTGTTGGAG-3′	NdeI
CGFP-(HC-1-125) stop	5′-GGTAATAT*GGATCC*TTAGTTGTACAGTTCATCCATGCC-3′	BamHI

Underlined bases are the annealing sequence; italics are the restriction sites and red is the sequence lost after enzyme digestion.

The PCR was carried out using an Expand High Fidelity PCR system (Roche Diagnostics, Indianapolis, IN) in the cycle: 94ºC, 2 min; 25x (94ºC, 15 s; 58ºC for NGFP-LC/62^{0}C for HC-

CGFP, 30 s; 72ºC, 45s); 72ºC, 7 min. The PCR products were investigated in agarose gel electrophoresis and purified using the GFX kit.

The plasmid pET11a-link-NGFP was first digested with XhoI (Fermentas) restriction enzyme for 16 hours. Later BamHI (Fermentas) was added and the plasmid was digested overnight at 37ºC. Plasmid pET9a- Calmodulin-CGFP was digested first with NdeI (Fermentas). After 16 h BamHI (Fermentas) was added and the double digestion occurred overnight at 37ºC [Lindman, 2009]. Doubly digested plasmids were treated with alkaline phosphatase for half an hour and then inactivated at 75ºC for 5 min. The amplified and purified PCR products NGFP-LC and HC-CGFP were digested in the same conditions according with the restriction sites flanking the gene sequence. No alkaline phosphatase treatment for PCR products. Each digestion is followed by GFX kit. Doubly digested vectors and scFv fragments were purified by gel electrophoresis and GFX kit. The double digestion was repeated once more in order to increase the fraction of digested plasmids available for ligation.

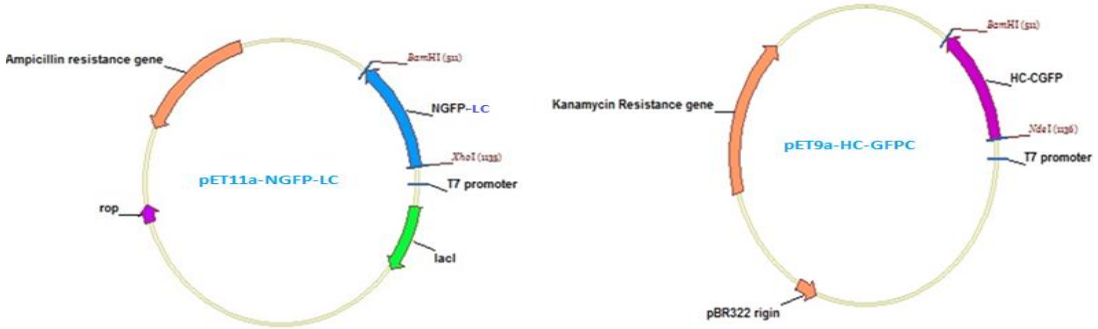


Figure 9: Plasmid maps used for cloning scFv fragments into split GFP system

The doubly digested scFv PCR fragments were ligated into the appropriate vector using T4 DNA ligase (Fermentas) during 16 h at 16ºC. The ligations were carried out at 1:1, 1:2 and 1:3 molar ratio of vector to insert. 3 µL of each ligation product was mixed with 40 µL of *E. coli* ER2566 calcium competent cells and kept in ice bath at least 30 min, incubated at 42ºC during 45 s and incubated again 10 min in ice before approximately 40 µL of solution was plated on LB/agar. The pET11a-NGFP-Light chain ligation products were plated in medium containing 50 µg/mL of ampicillin; while pET9a-Heavy chain-CGFP was added to medium containing 100 µg/mL of kanamycin. Several colonies from the plates grown overnight at 37ºC were picked and amplified during at least 8 h at 37ºC in 1.5-2 mL of LB medium containing appropriate antibiotic. The bacteria were amplified in a 15 mL Falcon tubes laid down for maximal surface exposed to air during agitation at 130 rpm in an incubator. Amplified plasmids were recovered and purified using illustra plasmidPrep Mini Spin Kit (GE Healthcare Life Sciences, Uppsala, Sweden). To confirm whether the plasmids contain the correct gene, DNA was sequenced with the primers included in Table 2 were used for

sequences harbored in the pET11a plasmid and for the pET9a plasmid together with BigDye Terminator v1.1 Cycle Sequencing Kit.

3.6.2 Detecting *in vivo* reassembly of GFP-scFv Wild Type fragments

To make competent cells harboring pET9a-HC-CGFP plasmid, the plasmids coding for the wild type (WT) fused genes HC 1-124-CGFP previously constructed was co-transformed by thermal shock adding 1 µL of each plasmid of the pair to 40 µL of calcium competent *E. coli* ER2566 (4.3).
pET11a-NGFP-LC ligation product was transformed to *E.coli* ER2566 competent cells harboring pET9a-HC-CGFP. The transformed mixture was added to LB plates containing 50 µg/mL of ampicillin, 100 µg/mL of kanamycin and 10 µM of IPTG. The plates were incubated five days at RT or 16 h at 37ºC and then taken out and left at RT during five days. The appearance and maturation of green fluorescence in the colonies over time was detected and followed as above (4.5.4). The construct GFP-Calbindin D_{9K} was used as positive fluorescent reassembly.

3.6.3 Detecting *in vivo* reassembly of GFP-scFv Light chain lib fragments and screening of mutants

The plasmids coding for the fused gene NGFP-LC-lib was transformed by thermal shock to 50 µL of chemical competent *E. coli* ER2566 containing pET9a-HC-CGFP plasmid. The transformed mixture was spread onto several LB plates with 50 µg/mL of ampicillin, 100 µg/mL of kanamycin, 10 µM of IPTG. The plates were incubated 16 h at 37ºC and then taken out and left at RT during five days. The appearance and maturation of green fluorescence in the colonies over time was detected and followed placing the plates on the Dark Reader for manual ranking.

3.6.4 SDS-PAGE Analysis

As transformation of both the pET plasmids gave no green colonies, again a test of expression was done using SDS-PAGE analysis. Expression and lysis of cells was done similarly as described before for co- expressing pQLinkN plasmid. A comparison was done between induced and un-induced empty and transformed *E.coli* ER2566 cells. Transformed cells tested include the cells having plasmids, pET11a-NGFP-LC, pET11a-NGFP-peptide lib, pET9a-HC-CGFP and pET-9a-CaM-CGFP.

3.6.5 Inducing expression of co-transformed *E.coli* ER2566 at different temperatures

While doing SDS-PAGE analysis, a simultaneous trial was setup for inducing expression of co-transformed *E.coli* ER2566 cells at different temperatures. Transformation and plating was done as described before for pQLink plasmids. Plates were incubated for 16 h at 37ºC. Few of these plates were left for incubation at 37ºC, few at room temperature and rest at 4ºC. Incubation of plates was done for 10 days and monitored daily for green fluorescent colonies.

4. RESULTS

4.1 DNA and Protein sequences of scFv

The wild type DNA sequence of FITC8, a single chain antibody was derived from the reference H.Persson, 2007 and the Pubmed id ABL14156. The nucleotide sequence was modified to *E.coli* preferred codons.

WT FITC8

EVQLLESGGGLVQPGGSLRLSCAASGFTFSNYWMSWVRQAPGKGLEWVSGISGNGGYTYFADSVKDRFTI
SRDNSKNTLYLQMNSLRAEDTAVYYCAGGDGSGWSFWGQGTLVTVSSGGGGSGGGGSGGGGSQSVLTQPP
SASGTPGQRVTISCTGSSSNIGAGYDVHWYQQLPGTAPKLLIYGNNNRPSGVPDRFSGSKSGTSASLAIS
GLRSEDEADYYCAAWDDSLSGTRVFGGGTKLTVL

E.coli compatible FITC8 nucleotide sequence

gag gtg cag ctg ttg gag tct ggg gga ggc ttg gta cag cct ggg ggg tcc ctg cgc ctc tcc tgt gca gcc tct gga ttc acc ttt agt aac tat tgg atg agc tgg gtc cgc cag gct cca ggg aag gga ctg gag tgg gtc tca ggg att agt ggt aac ggt ggt tac aca tac ttc gca gac tcc gtg aag gac cgg ttc acc atc tcc cgt gac aat tcc aag aac acg ctg tat ctg caa atg aac agc ctg cgc gcc gag gac acg gct gtg tat tac tgt gcg gga ggt gat ggc agt ggc tgg tcc ttc tgg ggc caa ggt aca ctg gtc acc gtg agc agc ggt gga ggc ggt tca ggt gga ggt ggc tcc ggc ggt ggc gga tcg cag tct gtg ctg act cag cca ccg tca gcg tct ggg acc ccg ggg cag cgc gtc acc atc tct tgc act ggg agc agc tcc aac atc ggg gca ggt tat gat gta cac tgg tat cag cag ctc cca gga acg gcc ccg aaa ctc ctc atc tat ggt aac aac aat cgg ccg tca ggg gtc cct gac cgt ttc tct ggc tcc aag tct ggc acc tca gcc tcc ctg gcc atc agt ggg ctc cgg tcc gag gat gag gct gat tat tac tgc gcg gcc tgg gac gac agc ctg agt gga act cgc gta ttc ggc gga gga acc aag ctg acg gtc cta

Letters in red: Heavy chain; Letters in Green: Linker sequence; Letters in Blue: Light chain, Letters in yellow highlight represents the E.coli compatible codons which were modified for the original codons available in the WT sequence.

4.2 Cloning and expression of GFP-fused scFv fragments sub-cloned into pQLinkN plasmid

4.2.1 Construction of co-expression plasmid, pQLinkN-NGFP-LC-linker-HC-CGFP

In order to detect interactions between the two wt fragments of scFv (LC and HC) with the GFP split system, a co-expression plasmid was constructed by ligation independent cloning (LIC) of two different pQLinkN plasmids carrying either wt NGFP-LC or wt HC-CGFP. Different concentrations of EDTA (25 mM – 100 mM), different incubation times (10 min, 1 h) and temperatures (room temperature and 16ºC) were evaluated. We found successful LIC if 100 mM EDTA was added half way in the incubation time of mixture of cut plasmids at 16ºC for 1 h. The LIC product was transformed into *E.coli* ER2566. There were some colonies on LB/agar plates without antibiotics which might be due to the re-ligation of SwaI plasmid. For an unknown reason, LIC qualified T4 DNA polymerase was not working if the digested products were treated with alkaline phosphatase. There lies no specific marker for the co-

expression pQLinkN plasmid apart from an antibiotic resistant marker, ampicillin, which was also present in individual pQLinkN plasmids carrying a single fragment of scFv. Individual colonies were picked from the LB/agar plates with ampicillin, cultured ON in 2 mL LB + ampicillin and plasmids prepared. Sequencing confirmed the complete construct NGFP-LC-linker-HC-CGFP (using primers referred in 4.5.2) to be in-frame with GFP gene fragment and with the adequate linker region (Figure 10 and 11).

Figure 10: Chromatogram of co-expressing pQLinkN plasmid. A and B represents NGFP-LC-Linker-HC-CGFP sequence obtained with forward and reverse primer (4.5.2) respectively.

CAACATTCTGGGACACAAATTGGAATACAACTATAACTCACACAACGTTCCCATCATGGCAGACAAACAAggtggctccggcggtggc
ggatcgcagtctgtgctgactcagcc
accgtcagcgtctgggacccc ggggcagcgcgtcaccatctcttgcactgggagcagctccaacatcggggcaggttatgatgtacactggtatcagcagctcccagg
aacggccccgaaactcctcatctatggtaacaacaatcggccgtcaggggtccctgaccgtttctctggctccaagtctggcacctcagcctccctggccatcagtggg
ctccggtccgaggatgaggctgattattactgcgcggcctgggacgacagcctgagtggaactcgcgtattcggcggaggaaccaagctgacggtcctaTAAAA
GCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGT
TCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGTAACAACACCATTTGTCGAGAAATCATA
AAAAATTTATTTGCTTTGTGAGCGGATAAC
AATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACT
ATGGGATCCatggaggtgcagctgttggagtctgggggaggcttggtacagcctgggggtccctgcgcctctcctgtgcagcctctggattcacctttagtaact
attggatgagctgggtccgccaggctccagggaagggactggagtgggtctcagggattagtggtaacggtggttacacatacttcgcagactccgtgaaggaccgg
ttcaccatctcccgtgacaattccaagaacacgctgtatctgcaaatgaacagcctgcgcgccgaggacacggctgtgtattactgtgcgggaggtgatggcagtggc
tggtccttctggggccaaggtacactggtcaccgtgagcagcggtggaggcggttcaggtggaAAGAATGGAATCAAAGTGAACTTCAAGACCCGCC
ACAACATTGAAGATGGAAGCGTTCAACTA

Figure 11: Sequence supposed to be amplified to confirm the in-frame presence of LC (blue letters) and HC (green letters). Text in yellow highlight represent the sequence identified from chromatogram (Figure 10) obtained individually either from forward or reverse primer. Text between the yellow highlight represents the overlapping sequence amplified by both the forward and reverse primers. Thus prove the presence of both the genes in- frame in pQLink plasmid. Linker residues shown in red letters, NGFP in black letters and CGFP in grey letters.

4.2.2 Preparation of Library of Light chain 125-246 fragment

To generate high affinity variants of scFv, a mutation library of light chain was constructed using degenerate oligo-nucleotides having 94% of correct base and 6% of any other three bases. Overlapping PCR of degenerate nucleotides was setup in different tubes with four concentrations of LC2-LC5 degenerate oligo- nucleotides. The size of PCR band obtained was tested by agarose gel electrophoresis. Complete light chain corresponds to a band size of approximately 390 bp. A band of right size was achieved at all the four different concentrations of intermediate primers used. However a difference in the intensity of band size was observed. Lowest intensity was noticed at 30 µM concentration, whereas 2 – 12 µM gave reasonable yield (Figure 12). Band was cut from 2, 5 and 12 µM lanes for further experiment.

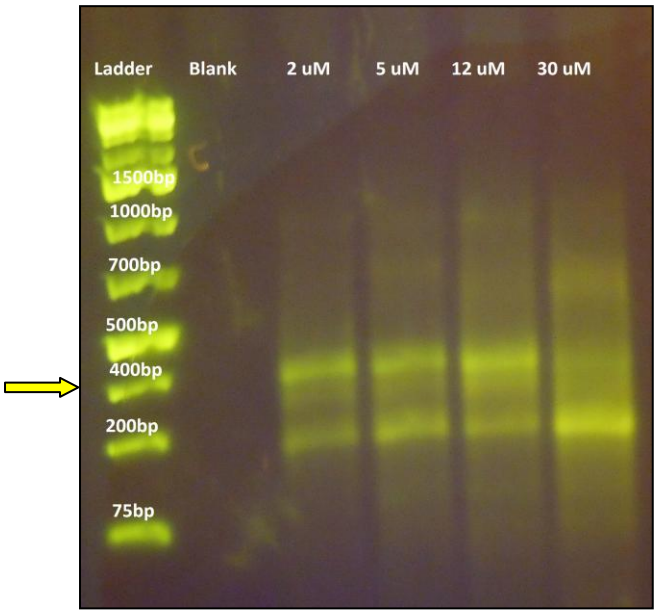


Figure 12: Agarose gel pic showing bands obtained by overlapping PCR of degenerate oligo nucleotides. Band pattern achieved for four different concentrations (2 µM, 5 µM, 12 µM and 30 µM) of LC2-LC6 and blank is labeled. Complete light chain corresponds to a band size of 391bp.

4.2.3 Detecting in vivo reassembly of scFv wild type fragments sub-cloned into pQLinkN plasmid

The NGFP-LC-link-HC-CGFP-pQLinkN plasmid was plated on LB/ampicilin/IPTG plate and incubated at 37ºC for 16 h. Later the plates were incubated at room temperature and screened during fifteen days for presence of green fluorescent colonies on a daily basis. No green colored colonies were obtained.

4.2.4 Detecting in vivo reassembly of scFv Light chain library fragments sub-cloned into pQLinkN plasmid and screening of mutants

As the wild type plasmid did not show any green colonies, we evaluated whether higher affinity variants would yield successful reconstitution using a mutant library of light chain. The NGFP-LC lib-link-HC-CGFP-pQLinkN plasmid was screened in the same manner as wild type plasmid. No green colonies were observed (Figure 13).

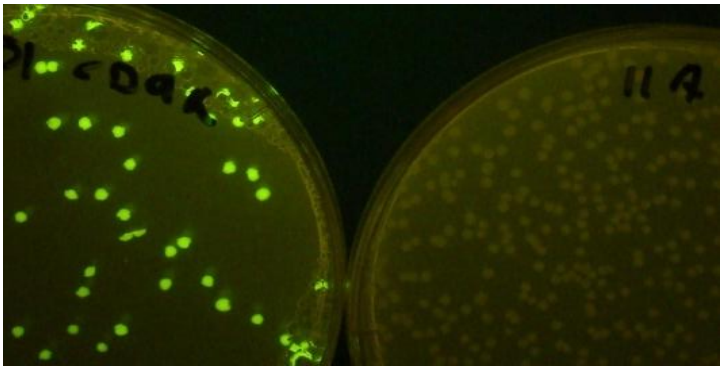


Figure13: Induced expression of split GFP fused protein fragments. The photograph includes a positive control, CD9K fragments fused to split GFP showing the green fluorescent colonies and scFv fragments fused to split GFP which have no green colonies.

4.2.5 Expression of scFv in other *E.coli* strains

Other strains of *E.coli*, BL21*PLysS and BL21 Gold were used to determine if the ER2566 strain had any limitations for expression of scFv protein. The two strains were transformed with (NGFP-LC/NGFP-LC-lib)-link-HC-CGFP-pQLinkN plasmids and screened similar to *E.coli* ER2566. No green colonies were identified.

4.2.6 SDS-PAGE analysis

As no green fluorescence was observed with scFv wild type or light chain library, a question was raised whether it is because of low expression of fusion proteins or it is because of steric problems in reconstitution. Hence we set up a SDS-PAGE to compare the protein band pattern of induced and un-induced empty and NGFP-LC-link-HC-CGFP-pQLinkN transformed cells of three different strains.

We found no difference in the band pattern of induced and un-induced empty and transformed cells. We could not find a band specific for NGFP-LC (30kDa) and HC-CGFP (23kDa). This conveyed us that there could be a problem with the expression of protein which might be hidden in the promoter sequence. We tried to activate the promoter by using different concentrations of IPTG.

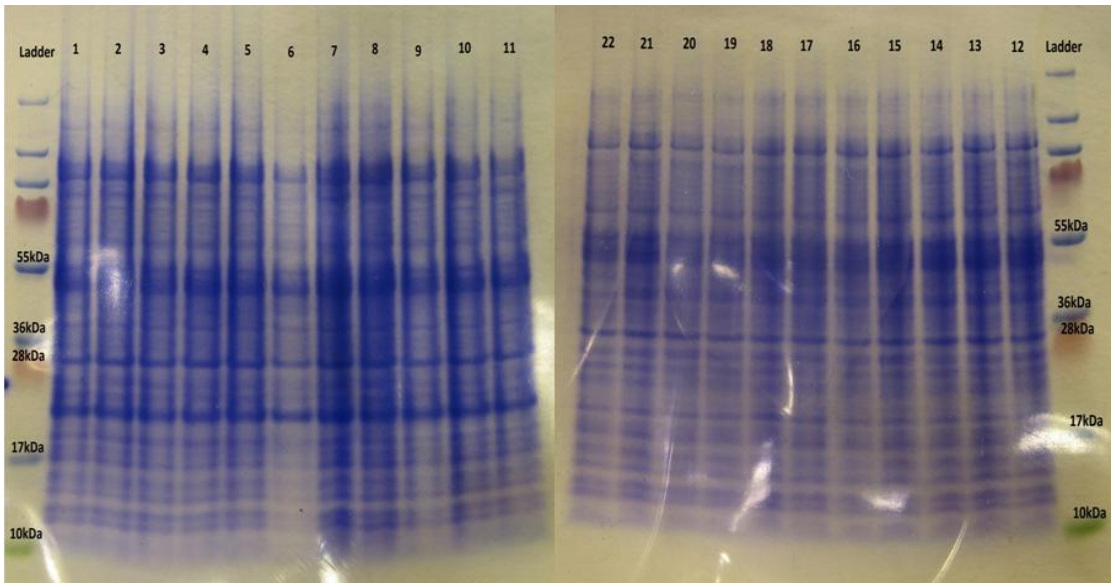


Figure 14: SDS-PAGE test to check expression pattern of induced and un-induced E.coli ER2566 and BL21*PLysS cells. Lanes 1: Induced E.coli ER2566 empty cells; Lanes 2-6: Transformed and un-induced E.coli ER2566; Lanes 7-11: Transformed and induced E.coli transformed cells; Lanes 12: Induced empty BL21* PLysS cells; Lanes 13-17: Transformed and un-induced BL21*PLysS; Lanes 18-22: Transformed and induced BL21*PLysS.

4.2.7 Plating of transformed cells onto LB-Amp plates with varying concentrations of IPTG

SDS-PAGE analysis conveyed that problems could be with the expression of protein. LB/agar/ampicillin plates with varying concentrations of IPTG (1.25 - 80 µM) were evaluated for activation of the promoter. However no green colonies were observed.

4.3 Cloning and expression of Dissected GFP-scFv Wild Type Fragments into pET plasmids

4.3.1 Constructing pET11a sub-cloned Light chain125-246 and pET9a sub-cloned Heavy chain 1-125-CGFP plasmid

To clone the scFv fragments into pET plasmids, the HC-CGFP and NGFP-LC constructs were extracted from pQLink plasmids by using PCR. Agarose gel analysis confirmed the amplified bands of right size (Figure 14). The extracted fragments were ligated into respective pET plasmids followed by transformation, plasmid preparation and DNA sequencing. This confirmed the presence of right construct.

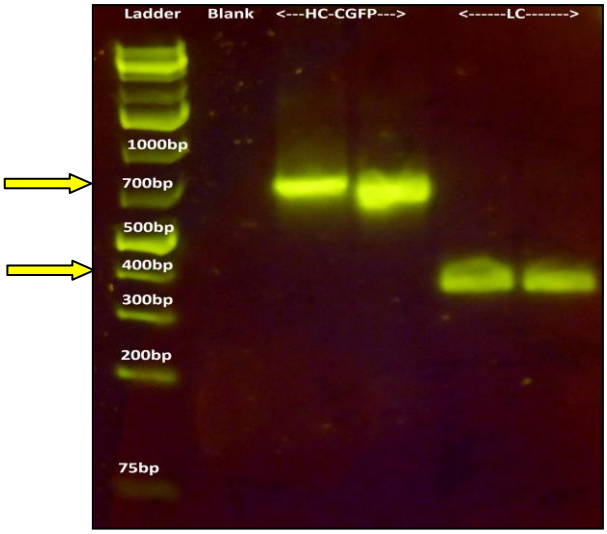


Figure 15: Agarose gel analysis confirming the right size of extracted fragments, HC-CGFP: Heavy chain fused to CGFP; LC: Light chain

4.3.2 Detecting in vivo reassembly of wild type GFP fused scFv fragments

The *E.coli* ER2566 cell co-transformed with wild type scFv sub-cloned pET plasmids was plated on LB/agar/ampicillin/kanamycin/IPTG plate and incubated at 37ºC for 16 h. Later the plates were incubated at room temperature and screened during fifteen days for presence of green fluorescent colonies on a daily basis. No green colored colonies were obtained.

4.3.3 Detecting in vivo reassembly of GFP-fused wt heavy chain and light chain library fragments and screening of mutants

The *E.coli* ER2566 cells co-transformed with pET9a-HC-CGFP and pET11a-NGFP-LC-lib were plated on LB/Ampicillin/Kanamycin/IPTG plate and incubated at 37ºC for 16 h. Later the plates were incubated at room temperature and screened during fifteen days for presence of green fluorescent colonies on a daily basis. No green colored colonies were obtained.

4.3.4 SDS-PAGE analysis

The absence of green colonies might be due to either poor expression of scFv or poor reconstitution of split scFv fragments. To distinguish between these possibilities SDS-PAGE was used to evaluate expression. pET11a-NGFP-LC was expressed only in induced cells while no expression was observed with pET9a-HC-CGFP plasmid. pET11a-NGFP-peptide was expressed in both induced and un-induced cells. pET9a-CaM-CGFP was not expressed. None of the constructs seemed to be expressed in the cells harboring both the NGFP-LC and HC-CGFP constructs (Figure 16).

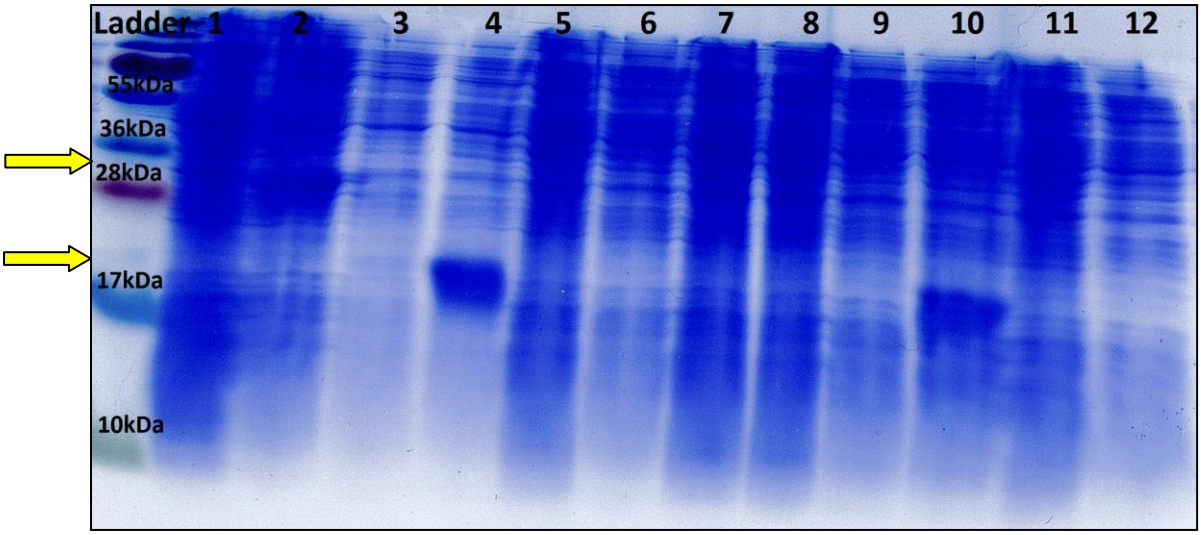


Figure 16: SDS-PAGE analysis of pET plasmid sub-cloned constructs transformed into E.coli ER2566 cells. Lanes 1-6 show induced colonies and Lanes 7-12 represent un-induced colonies. The order of constructs in the lanes is Empty cells (1, 7), pET11a-NGFP-LC (2, 8; Mw: 30.6kDa), pET9a-HC-CGFP (3, 9; Mw: 22.8kDa), pET11a-NGFP-Peptide (4, 10; Mw: 19.9kDa), pET9a-CaM-CGFP (5, 11; Mw: 25kDa), pET11a-NGFP-LC and pET9a-HC-CGFP co-transformed into E.coli ER2566 (6, 12).

4.3.5 Inducing expression of co-transformed *E.coli* ER2566 at different temperatures

To check if temperature was a limitation for expression of protein, the co-transformed *E.coli* ER2566 cells with wild type and mutant library fragments were incubated at three different temperatures (37ºC, room temperature and at 4ºC) for 10 days. No green colonies were observed.

5. DISCUSSION

The here presented work aims at finding high stability variants of scFv, a single chain antibody fragment. The project was based on the idea that the intensity of green fluorescent colonies formed by reconstitution of split GFP fused dissected protein fragments is correlated to the stability of intact protein variants. The interaction between the protein fragments promotes the assembly and folding of GFP and hence the chromophore development. This process is usually slow with split GFP alone. The idea evidently worked with a good number of proteins (Lindman, 2009, 2010). Hence we believed that more stable variants can be extracted using split GFP system.

If the scFv is soluble and expressed in cytoplasm, this should result in reconstitution and activity of GFP. In this we failed to detect green fluorescence. Therefore we will propose a number of strategies for further development of the scFv reconstitution in the split-GFP system.

5.1 DNA and protein sequence of scFv

The major interest in selecting a scFv construct is because of its good number of applications. The recombinant antibodies, single chain fragments (scFv) have been used in both therapy (Chester, 1995) and diagnoses, and have shown advantages over conventional and monoclonal antibodies (Lorimer, 1996; Turner, 1997; Wintlow, 1991). Such molecules were proven to have rapid blood circulation, since patients treated with scFv presented good localization of the molecules only one hour after their injection (Georg, 1996). They also present good penetration in tissues, low immunogenicity in theory, low retention in the kidneys and other non-target organs, better penetration in target tumors, are easily constructed, have low commercial cost in large-scale production (Wintlow, 1991) and, moreover, it is possible to restructure them in order to improve their activity and production (Turner *et al.*, 1997).

As no scFv molecule of medical importance is conveniently available to work with, we thought to continue with a model scFv protein having a considerable stability to work with and here comes the FITC8. The properties studied with FITC8 can be extended to other scFv proteins of medical importance which could be of any help. The wild type FITC8 protein has been proved to be highly stabile and well expressed in *E.coli* (Griffiths, 1994; Ewert S, 2003; Steinhauer, 2002). In order to enhance the production of functional scFv, codon optimization compatible to *E.coli* was done (Tiwari A, 2010).

5.2 Cloning and expression of split GFP-fused scFv fragments sub-cloned in pQLinkN plasmids

To minimize problems during screening of libraries due to varying expression of library members after co-transformation of separate plasmids, our first approach was to use a plasmid which can carry two different genes in a single plasmid and is able to express both the constructs, the pQLinkN plasmid. The LINK sequences are the interesting regions in pQLinkN. Ligation independent cloning (LIC) of the two LINK sequences establishes the creation of co-expressing plasmid and no PCR is required. LIC makes use of 3'-5' activity of T4 DNA polymerase. In principle this process is more efficient than T4 DNA ligase cloning as only desired product can form.

When we tried LIC of two pQLinkN plasmids with individual dissected constructs, we found that only few colonies had the co-expressing construct. The other colonies show a re-ligated SwaI cut plasmid product. To prevent re-ligation of cut plasmids we tried alkaline phosphatase treatment. However, LIC qualified T4 DNA polymerase had problems in its activity when we used alkaline phosphatase, for unknown reason. Gel extraction prior to purification of cut plasmids increased the ratio of colonies having co-expressing construct. While doing ligation independent cloning it is better to try for different concentrations of EDTA, different temperatures for incubation of T4 DNA polymerase treated cut plasmids and also different concentrations of plasmid. Depending on the constructs used, improvements might be needed to the protocol. The possible role of EDTA could be stabilizing the non-covalent interactions between the insert and vector till it gets repaired by bacterial host ligase enzyme.

Protein expression in general is controlled by varying multitude of factors known to influence the process, such as aeration, IPTG concentration, temperature, medium, strain, DNA sequence etc. Here we examined some of the factors that might inhibit the formation of green fluorescent colonies. We started to look into the primary sequence of the scFv protein and GFP. Big dye termination sequencing confirms that the LIC product obtained is proper in terms of the sequence, in frame with the GFP counterparts and also with the promoter as per the map provided in the reference Scheich.C, 2007. There might be strain limitation for expression of few proteins. Hence a switch from *E.coli* ER2566 to BL21*PLysS and BL21 DE3 Gold was done. BL21*PLysS strain has a highly efficient protein expression under the control of T7 promoter. It also contains a plasmid, pLysS, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG. BL21 Gold has high transformation efficiency. In addition, the

gene that encodes endonuclease I (*endA*), which rapidly degrades plasmid DNA isolated by most miniprep procedures, is inactivated. However these strains were not of any help. Later we looked into the promoter involved and its properties. As per the reference, pQLink plasmid has a *tac* promoter, a hybrid promoter. However we could find only T5 promoter sequence. T5 promoter can also be induced with IPTG and it should be able to express the fusion proteins. Different concentrations of IPTG tried to induce the expression were also of not any help to produce any protein. There is likely not a problem with enough oxygen transfer as we used 500 ml culture flasks and shaking at 130 rpm which gives very high level expression of many proteins.

5.3 Library generation of scFv Light chain fragment

Random mutant library can be generated by several ways. In our project, a mutant library was generated using PCR of overlapping degenerate oligo nucleotides having a designed mutant rate. A 30 µM concentration of start and end primers with a combination of range of low concentrated intermediated primers (2 µM, 5 µM and 12 µM) taken in a single tube gave a full length mutant light chain. Error-prone PCR, an usual way used to generate mutant variants is actually very random. The number of mutants generated by this method is not assured always. At times a mutant generated might end up to a silent mutation or null mutation. However as the expression strategy hadn't worked satisfactorily, we couldn't screen any mutant colonies for its high affinity towards the counter partner.

5.4 Cloning and expression of split GFP-fused scFv fragments sub-cloned in pET plasmids

As the co-expressing plasmids did not show any expression of the scFv constructs, we got back to work with pET plasmid system, a classical and well-expressing set of plasmids. Earlier pET11a-NGFP and pMRBAD-CGFP plasmids were used. This combination of plasmids have imbalance in the levels of expression in particular pMRBAD-CGFP has low expression. As mentioned before it is important to balance the expression of co-transformed plasmids. Therefore we tried pET9a plasmid which has high expression levels comparable to pET11a-link- NGFP vector. Extraction of LC and HC-CGFP construct from the pQLinkN plasmids was successful and cloning into pET plasmids was confirmed to be in-frame by the sequencing analysis. Transformed *E.coli* cells with calmodulin-CGFP and also HC-CGFP constructs had problems when kanamycin of 100 µg/ml concentration was used. A range of concentrations was tried to optimize the kanamycin concentration for the constructs used. Kanamycin of concentration 35ug/ml was found to be good as evaluated by increased number of the total number of colonies harboring the HC-CGFP-pET9a plasmid. As the co-

transformed pET 11a and pET9a plasmids have different antibiotic markers, colonies were confirmed to have both the plasmids. However we couldn't find any green colonies on inducing with IPTG. The SDS-PAGE analysis of the individually expressed NGFP-LC-pET11a and pET9a-HC-CGFP compared with the positive constructs show the problem could be with CGFP fused protein constructs. Strangely the positive control CGFP construct, CaM-CGFP did not show a band when treated in a similar way to all other plasmids. The HC-CGFP construct is either not expressing at all or it is getting lost or degraded. Also the un-induced transformed cells with positive NGFP-CaM peptide construct show a basal expression of the fusion protein. This could be because of the T7 promoter, a leaky promoter. T7 RNA polymerase is IPTG controlled and will be leaky. Media might have minute amounts of lactose. So the target gene is transcribed and translated already in the upgrowth. Usage of BL21 (DE3) pLysS or E product can switch off T7 RNA polymerase. Also addition of 1% glucose in medium controls the basal expression via catabolite repression (Robert Novy). Another thing to be noticed from SDS-PAGE gel was there is no expression of light chain in the cells that were co-transformed with both pET11a-NGFP-LC and pET9a-HC-CGFP plasmids. There could be some interference while the bacteria express both the constructs.

5.5 Expression of scFv protein

In general there are many evidences that put forward the problems involved in the expression of scFv proteins. One of the main problems associated with most scFvs is that they are not able to fold under the reducing conditions of the cell cytosol and nucleus, where most of the interesting targets are located. This is thought to be due to the limited stability of scFv's after the two conserved disulfide bonds are reduced, as occurs in the cell cytosol (Biocca S, 1995). Indeed, in vitro, most of the scFv's cannot be renatured under reducing conditions (Ramm K, 1999; Martineau P, 1999). To be an efficient intrabody a scFv must thus present a high in vitro stability (Worn A, 2000). Recent studies using either statistical analyses of scFv sequences (Visintin M, 2004) or an experimental approach (Ayf der Maur A, 2004) have shown that less than 1% of the scFv's are stable enough to be high quality intrabodies and that only about 10% have a "moderate chance" to be functional in vivo. In addition, even if a scFv protein is indeed stable enough in its reduced form to be expressed and active in vivo, other parameters such as protease susceptibility (Parsell DA, 1989) or folding kinetics (Martineau P, 1999) may also influence the final in vivo fate of the protein and are critical for intrabody expression and activity (Auf der Maur A, 2002).

FITC8 protein has two disulfide bridges which usually play an important role in protein structure and activity (Darby, 1995) and are required in some folding pathways (Schultz S, 1987) and for the stability of certain folded polypeptides (Vanhove, 1997; Stewart, 1998).

In E. coli, disulfide bonds occur almost in proteins secreted in the periplasm or the cell envelope (e.g. PhoA, OmpA) (Missiakkas, 1997; Raina, 1997; Rietsch, 1998). The disulfide oxidoreductases

and isomerases (Dsb chaperones) located in periplasm are involved in these processes. Therefore, when produced in the cytoplasm of E. coli wild-type cells, the scFv's were found reduced and unfolded. Oxidized and functional scFv can be produced efficiently in the cytoplasm *E. coli trxB and gor* strain FA 113 cells. This strain has a suppressor mutation in a gene coding for peroxiredoxin which makes the enzyme to act as disulfide reductase. Also the growth rates and biomass yields of the strain are close to those obtained by E. coli wild-type strains [(Jurado P, 2002)]. In this part of my project, we failed to express the scFv protein and it probably needs more experiments to really explain the problem.

6 FUTURE WORK

Here we propose a number of factors to evaluate for achieving successful reconstitution of FITC8 - scFv in the split GFP system.

1. Use different *E.coli* strain, which might help in a better expression of the scFv fragment. Most of the protocols for expression of a scFv protein used in several labs include *E.coli* K-12, *E.coli* XL-1 Blue, *E.coli* TG-1, *E.coli* MRE600, *E.coli* JM83, *E. coli trxB and gor* strain FA 113 cells [Philibert P, 2007, Su-Jun,1994; Freund C, 1993; Nieba L, 1997, Jurado P, 2002].
2. Use rich medium like SOC, YT etc which might improve the bacterial growth and expression of protein. Few protocols also had M9 medium, I*A medium to culture bacteria expressing a scFv protein [Nieba L, 1997, Su-Jun, 1994, Philibert P, 2007].
3. Check the expression of bacteria in liquid culture at lower temperatures.
4. Check expression in bacteria having the pET plasmid constructs on LB plates with varied concentrations (1-100 µM) of IPTG.
5. Also the expression of bacteria in liquid culture can be tested at various IPTG concentrations.
6. It could also be that the expression of CGFP constructs is very low to detect on SDS-PAGE. Western blot analysis with an antibody specific to CGFP can resolve the issue.
7. The failure to observe the reconstitution may be due to steric hindrance. This might be resolved by optimizing linker length and composition. Vary the linkers in NGFP-LC and HC-CGFP using quick change PCR.
8. Steric hindrance may also be relieved by changing the orientation of the fragments in the split-GFP system. Try also NGFP-HC and LC-CGFP.

7 CONCLUSIONS

- Construction of co-expressing multi gene pQLinkN plasmid is successful. The genes included, NGFP-LC and HC-CGFP of scFv were in-frame.

- Construction of a mutant library by overlapping PCR of degenerate oligo nucleotides having a certain mutation rate assures better variability of the library fragments developed. It is a simple and easy technique.

- Extraction of light chain and heavy chain fused CGFP fragments from pQLinkN plasmids is successful.

- Cloning of the LC IN pET11a and HC-CGFP in pET9a was successful and in-frame.

- SDS-PAGE analysis shows excellent expression of NGFP-LC from pET11a plasmid.

- SDS-PAGE analysis hints a basal expression activity of pET11a plasmid.

- The absence of green colonies may be due to problems with expression of protein or to steric hindrance which might prohibit successful reconstituion.

References

Abedi M R, Caponigro G, Kamb A (1998). "Green Fluorescent protein as a scaffold for intracellular presentation of peptides" Nucleic Acids Res 26: 623-630.

Alexandrov A, Vignali M, LaCount DJ, Quartley E, de Vries C, De Rosa D, Babulski J, Mitchell SF, Schoenfeld LW, Fields S, Hol WG, Dumont ME, Phizicky EM, Grayhack EJ (2004)."A facile method for high-throughput co-expression of protein pairs" Mol Cell Proteomics. 3(9):934-8.

Auf der Maur A, Tissot K, Barberis A (2004). "Antigen-independent selection of intracellular stable antibody frameworks" Methods, 34:215-224.

Auf der Maur A, Zahnd C, Fischer F, Spinelli S, Honegger A, Cambillau C, Escher D, Pluckthun A, Barberis A (2002). "Direct in vivo screening of intrabody libraries constructed on a highly stable single-chain framework" J Biol Chem, 277:45075-45085.

Barnard E, McFerran N V, Trudgett A, Nelson J, Timson D J (2008). "Detection and localization of protein-protein interactions in Sacharomyces cerevisiae using a split GFP method" Fungal Genet Biol 45: 597-604.

Berggård T, Linse S, James P (2007) Methods for the detection and analysis of protein-protein interactions. Proteomics 7: 2833-2842.

Biocca S, Ruberti F, Tafani M, Pierandrei-Amaldi P, Cattaneo A (1995). "Redox state of single chain Fv fragments targeted to the endoplasmic reticulum, cytosol and mitochondria" Biotechnology (N Y). 13(10):1110-5.

Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M (1988). "Single-chain antigen-binding proteins" Science 242(4877):423-6.

Baird G, Zacharias D A, Tsien R Y (1999). "Circular permutation and receptor insertion within green fluorescence proteins" Proc Natl Acad Sci 98: 11241-11246.
Carey J, Lindman S, Bauer M, Linse S (2007). "Protein reconstitution and three dimensional domain swaping: benefits and constrains of covalency" Prot Sci 16: 2317-2333.

Chalfie M, Tu Y, Euskirchen G, Ward W W, Prasher D C (1994). "Green fluorescent protein as a marker for gene expression" Science 263: 802-805.

Christoph Scheich, Daniel Kümmel, Dimitri Soumailakakis, Udo Heinemann,and Konrad Büssow (2007). "Vectors for co-expression of an unrestricted number of proteins" Nucleic Acids Res. 35(6): e43.

Chester KA, Hawkins RE (1995). "Clinical issues in antibody design". Trends Biotechnol.13 (8):294-300.

Christian Freund, Alfred Rossa, Birgith Guthb, Andreas Pluckthun and Tad A. Holak (1993). "Characterization of the linker peptide of the single-chain F, fragment of an antibody by NMR spectroscopy" FEBS 97-100.

Darby, N. J., Morin, P. E., Talbo, G. and Creighton, T. E. (1995). "Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates." J. Mol. Biol. 249, 463–477.

Eijsink VGH, Gåseidnes S, Borchert TV, van den Burg B (2005) “Directed evolution of enzyme stability” Biomol Eng 22:21–30.

Eijsink VGH, et al. (2004) “Rational engineering of enzyme stability” J Biotechnol 113:105–120.

Ewert, S., Honegger, A. & Pluckthun, A. (2003). Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. Biochemistry, 42, 1517–1528.

Fields S, Song O-K (1989). “A novel genetic system to detect protein-protein interactions” Nature 340: 245-246.

George A. J.T.; Huston, J.S.; Haber E (1996). “Exploring and exploiting the antibody and Ig superfamily combining sites” Nat. Biotechnol, v.14.

Ghosh I, Hamilton A D, Regan L (2000). “Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein” J Am Chem Soc 122: 5658-5659.

Griffiths, A. D., Williams, S. C., Hartley, O.,Tomlinson, I. M., Waterhouse, P., Crosby, W. L. et al.(1994). “Isolation of high affinity human antibodies directly from large synthetic repertoires” EMBO J. 13, 3245–3260.

Hantgan R R, Taniuchi H (1997). “Formation of a biologically active ordered complex from 2 overlapping fragments of cytochrome c” J Biol Chem 252: 1367-1374.

Helena Persson, Johan Lantto and Mats Ohlin (2006). “A focused antibody library for improved hapten recognition” J.Mol.Biol 357, 607-620.

Henert T, Gales C, Rebois R V (2006). “Detecting and imaging protein-protein interactions during G protein-mediated signal transduction in vivo and in situ by using fluorescence-based techniques” Cell Biochem Biophys 45: 85-109.

Hu C D, Chinenov Y, Kerppola T K (2002). “Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation” Mol Cell 9: 789-798.

Johnsson N, Varshavsky A (1994). “Split ubiquitin as a sensor of protein interactions in vivo” Proc Natl Acad Sci USA 91: 10340-10344.

Jorgensen L, Hostrup S, Moeller EH, Grohganz H (2009) “Recent trends in stabilizing peptides and proteins in pharmaceutical formulation-considerations in the choice of excipients” Expert Opin Drug Del 6:1219–1230.

Kerppola T K (2008). “Bimolecular Fluorescence Complementation (BiFC) analysis as a probe of protein interactions in living cells” Annu Rev Biophys 37: 465-487.

Kobayashi N, Honda S, Yoshii H, Uedaira H, Munekata E (1995). “Complement assembly of two fragments of the streptococcal protein G B1 domain in aqueous solution” FEBS lett 366: 99-103.

Lalonde S, Ehrhardt D W, Loqué D, Chen J, Rhee S Y, Frommer W B (2008). “Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations” Plant J 53: 610-635.

Lindman S, Johansson I, Thulin E, Linse S (2009). “Green fluorescence induced by EF-hand assembly in a split GFP system” Protein Sci 18: 1221-1229.

Lindman S, Garcia AH, Szczepankiewicz O, Frohm B, Linse S (2010). “In vivo protein stabilization based on fragment complementation and a split GFP system” PNAS 19826-19831.

Lorimer IA, Keppler-Hafkemeyer A, Beers RA, Pegram CN, Bigner DD, Pastan I (1996). “Recombinant immunotoxins specific for a mutant epidermal growth factor receptor: targeting with a single chain antibody variable domain isolated by phage display” Proc Natl Acad Sci U S A. 10; 93(25):14815-20.

Lars Nieba, Annemarie Honegger, Claus Krebber and Andreas Pluckthun (1997). “Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment” Protein Engineering, 435–444.

Martineau P, Betton JM (1999). “In vitro folding and thermodynamic stability of an antibody fragment selected in vivo for high expression levels in Escherichia coli cytoplasm” J Mol Biol 292:921-929.

Malakauskas S M, Mayo S L (1998). “Design, structure and stability of a hyperthermophilic protein variant” Nat Struct Biol 5: 470-475.

Magliery T J, Regan L (2006). “Reassembled GFP: Detecting protein-protein interactions and proteins expression paterns” Chapter in Green Fluorescent Protein: Perspectives, applications and protocols, 2nd Ed, John Wiley, pp 391-405.

Magliery T J, Wilson C G M, Pan W, Mishler D, Ghosh I, Hamilton A D, Reagan L (2005). “Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism” J Am Chem Soc 127: 146-157.

Martineau (2007). “A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm” BMC Biotechnology, 7:81.

Missiakas, D and Raina, S (1997). “Protein folding in the bacterial periplasm.” J. Bacteriol. 179, 2465–2471.

Paola Jurado, Daniel Ritz, Jon Beckwith, Victor de Lorenzo and Luis A. Fernandez (2002). “Production of Functional Single-Chain Fv Antibodies in the Cytoplasm of Escherichia coli” J. Mol. Biol.320, 1–10.

Parsell DA, Sauer RT (1989). “The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli” J Biol Chem, 264:7590-7595.

Pascal Philibert, Audrey Stoessel, Wei Wang, Annie-Paule Sibler, Nicole Bec, Christian Larroque, Jeffery G Saven, Jérôme Courtête, Etienne Weiss and Pierre

Pelletier J N, Campbell-Valois F X, Michnick S W (1998). “Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments” Proc Natl Acad Sci USA 95: 12141-12146.

Raina, S and Missiakas, D(1997). “Making and breaking disulfide bonds.” Annu. Rev. Microbiol. 51,179–202.

Ramm K, Gehrig P, Pluckthun A (1999). “Removal of the conserved disulfide bridges from the scFv fragment of an antibody: effects on folding kinetics and aggregation” J Mol Biol, 290:535-546.

Richards F M (1958). “On the enzymatic activity of subtilisin-modified ribonuclease” Proc Natl Acad Sci USA 44: 162-166.

Rietsch, A and Beckwith, J (1998). “The genetics of disulfide bond metabolism.” Annu. Rev. Genet. 32,
164–184.

Robert Novy and Barbara Morris. “Use of glucose to control basal expression in the pET System” Novagen, Inc.

Rossi F, Charlton C A, Blau H M (1997). “Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation” Proc Natl Acad Sci USA 94: 8405-8410.

Sarkar M, Magliery T J (2008). “Re-engineering a split-GFP reassembly screen to examine RING-domain interactions between BARD1 and BRCA1 mutants observed in cancer patients” Mol Biosyst 4: 599-605.

Schägger H., v. J. G. (1987). "Tricine-Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range from 1 to 100 kDa " Analytical Biochemistry 166: 12.

Shimomura O, Johnsson F H, Saiga Y (1962). “Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea” J Cell Comp Physiol. 59: 223-239.

Schultz, S. C., Dalbadie-McFarland, G., Neitzel, J. J. and Richards, J. H. (1987). “Stability of wild-type and mutant RTEM-1 beta-lactamases: effect of the disulfide bond.” Proteins: Struct. Funct. Genet. 2, 290–297.

Steinhauer, C., Wingren, C., Hager, A. C. & Borrebaeck, C. A. K. (2002). Single framework recombinant antibody fragments designed for protein chip applications. Biotechniques, Suppl., 38–45.

Stewart, E. J., A slund, F. & Beckwith, J. (1998). “Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins.” EMBO J. 17, 5543–5550.

Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Åberg, A. M. et al. (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. Nature Biotechnol.18, 852–856.

Solá RJ, Griebenow K (2009) “Effects of glycosylation on the stability of protein Pharmaceuticals” J Pharm Sci 98:1223–1245.

Su-jun Deng, C. Roger MacKenzie, Joanna SadowskaS, Joseph Michniewicz, N. Martin Young, David R. Bundle, and Saran A. Narang D (1984). “Selection of Antibody Single-chain Variable Fragments with Improved Carbohydrate Binding by Phage Display” The J. Biol. Chemistry, 9533-9538.

Tiwari A, Sankhyan A, Khanna N, Sinha S (2010). “Enhanced periplasmic expression of high affinity humanized scFv against Hepatitis B surface antigen by codon optimization” Protein Expr Purif. 74(2):272-9.

Topell S, Hennecke J, Glockshuber R (1999). “Circularly permuted variants of the green fluorescent protein” FEBS lett 457:283-289.

Tsien R (1997). “The green fluorescent protein” Annu Rev Biochem 67: 509-544.

Tsumoto K, Nakaoki Y, Ueda Y, Ogasahara K, Yutani K, Watanabe K, Kumagai I (1994). “Effect of the order of antibody variable regions on the expression of the single-chain HyHEL10 Fv fragment in E. coli and the thermodynamic analysis of its antigen-binding properties” Biochem Biophys Res Commun. 201(2):546-51.

Turner, D.J.; Ritter, M.A.; George, A.J.T (1997). “Importance of the linker in expression of single chain Fv antibody fragments-optimisation of peptide sequence using phage display technology”. J. Immunol. Methods, v.205.

Ullmann A, Jacob F, Monod J (1967). “Characterization by in vivo complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli” J Mol Biol 24: 339-343.

Vanhove, M., Guillaume, G., Ledent, P., Richards, J. H., Pain, R. H. & Frere, J. M. (1997). “Kinetic and thermodynamic consequences of the removal of the Cys-77-Cys-123 disulphide bond for the folding of TEM-1 beta-lactamase” Biochem. J. 321, 413–417.

van der Walle CF, Sharma G, Ravi Kumar MNV (2009) “Current approaches to stabilising and analysing proteins during microencapsulation in PLGA” Expert Opin Drug Del 6:177–186.

Villalobos V, Naik S, Pownika-Worms D (2007). “Current state of imaging protein-protein interactions in vivo with genetically encoded reporters” Annu Rev Biome Eng 9: 321-349.

Visintin M, Quondam M, Cattaneo A (2004). “The intracellular antibody capture technology: towards the high-throughput selection of functional intracellular antibodies for target validation” Methods, 34:200-214.

Wilson C G M, Magliery T J, Regan L (2004). “Detecting protein-protein interactions with GFP fragment reassembly” Nat Met 1:255-262.

Wintlow, M.; Filppula, D (1991). “Single-chain Fv proteins and their fusion proteins”. Methods, v.2.

Worn A, Auf der Maur A, Escher D, Honegger A, Barberis A, Pluckthun A (2000). “Correlation between in Vitro Stability and in Vivo Performance of Anti-GCN4 Intrabodies as Cytoplasmic Inhibitors” J Biol Chem 275:2795-2803.

Wunderlich M, Martin A, Staab CA, Schmid FX (2005) “Evolutionary protein stabilization in comparison with computational design” J Mol Biol 351:1160–1168.

Xue W F, Szczepankiewicza O, Bauer M, Thulin E, Linse S (2004). “Intra- versus Intermolecular Interactions in Monellin: Contribution of Surface Charges to Protein Assembly”. J Mol Biol 358: 1244-1255.

List of abbreviations

Amp	Ampicillin
AP	Alkaline phosphatase
BiFC	Bimolecular Fluorescent Complementation
BL21*	*E.coli* BL21*pLysS(DE3)
bp	base pairs
Cam	Chloramphenicol
CLSM	Confocal Laser Scanning Microscopy
dH2O	Sterile double distilled water
dH2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotidetriphosphate
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FITC	Fluorescein isothiocyanate
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
HC	Heavy chain
HC-CGFP	Heavy chain fused to C'-Green fluorescent protein fragment
Ig	Immunoglobulin G
IPTG	Isopropyl β-D-1 thiogalactopyranoside
Kan	Kanamycin
LB	Luria broth
LC	Light chain
LC-lib	Random mutant library of light chain
LIC	Ligation independent cloning
MW	Molecular weight
NGFP-LC	Light chain fused to N'-Green fluorescent protein fragment
OD	Optical density
PCR	Polymerase chain reaction
RT	Room temperature
scFv	Single chain antibody fragment
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
TBE	Tris/Borate/EDTA
Tris	tris(hydroxymethyl)aminomethane
Wt	Wild type
YT	Yeast extract-Tryptone

Acknowledgements

I thank my supervisor Sara Linse for giving me the chance to work in her group. My deepest gratitude for all the advice and support I got from you during the last months.

I also thank Tanja and Jonas for every brilliant idea you had during discussion of my results and for all your input on my work.

I thank Brigitta Frohm for taking the time to teach me in the lab and all her support.

Many thanks as well to my lovely colleagues Erik, Karin, Risto, Xiaoting, Anna and Irem for the good company in the lab.

Thanks to all the crew of the Biochemistry Department for the very good memories and all the help provided.

All my friends in Sweden thank you so much for encouraging me along the way and making me smile every day.

And last but not least the greatest thanks to my family and friends at home for being there for me, always.