# Development of a multi-copy bead display approach and its application in *in vitro* evolution of proteins and nucleic acids

DISSERTATION

submitted to the

Faculty of Chemistry and Biochemistry

Ruhr University Bochum

for the Degree

of Doctor of Natural Science

By

**Siddhartha Paul**

from Kamrup, India


Bochum, October 2012

# Acknowledgements

"*There is a single light of science, and to brighten it anywhere is to brighten it everywhere"*
*Issac Ashimov*

*Science delves into mystery and shows the path of light through understanding and comprehensibility. A lot of people have been instrumental in guiding me through my PhD years and I believe my research would not have been fruitful without them. I would like to take the opportunity to thank them all.*

*This work was inspired by one of the best mentors I have known, Prof. Klaus Ueberla. I would like to extremely thank him for believing in me, for his excellant scientific guidance and support and for giving me the freedom to implement my ideas.*

*I would like to thank Prof. Christian Herrmann for accepting my modest request to be my second referee.*

*A profound acknowledgement to Dr. Alexander Stang for his assistance and constructive criticism throughout my work.*

*I extend my gratitude to Prof. Matthias Tenbusch for aquainting me with FACS. I also deeply appreciate Dr. Ghulam Nabi for his support in these years.*

*I appreciated all the support and friendly work environment I had from my current fellow graduate students and colleagues Bastian Grewe, Thomas Niezold, Michael Storcksdieckgenannt Bonsmann, Bianca Hoffmann, Camilla Hrycak, Dr. Thomas Grunwald, Dr. Vladimir Temchura, Christoph Dedden, Kristin Wellbrockand past members Maik Blissenbach, Lena Jordan, Donatian Kamdem, Ruth Leitz, Wibke Bayer and other members of the department.*

*I specially acknowledege technical support of Klaus Sure and Klaus Lennartz from Uni Klinikum, University of Essen in various PCR related matters and FACS sorting respectively which has been invaluable for the success of this work.*

*My gratitude goes to Regina Bütermann, Alexa Ambrosat, Bettina Tippler, Ulrich Schumacher, Andreia Fernandes, Rosemarie Bohr and Heike Seidenstücker for their support during the course of this study.*

*I also thank Gundula Talbot of the GSCB, Chemistry department for all her help throughout my graduate years.*

*I am thankful to my friends in Bochum Sanjeev, Elisa, Ashish, Durga, Ram, Suvendu, Avinash, Divya, Adi, Soujanya, who accompanied me throughout and made my stay in Bochum worthwhile.*

*I am indebted to my family members and my in laws for their unconditional love, support and inspiration.*

*Finally, I would like to thank the special person in my life-Shankha, a long time friend and now wife. She had been an ardent listener, a critic, a constant source of inspiration and support throughout this endeavor. Her belief in me during weaker moments gave me the courage to persevere. I would not have made it without her by my side.*

**To my Parents.....**

# Contents

# List of Figures

# List of Tables

# Abbreviatons

| | | | |
|---|---|---|---|
| % | Percentage. | IVTT | *In vitro transcription translation* |
| °C | Degree celcius | kb | Kilobase pairs |
| µg | Microgram | kD | Kilo dalton |
| µl | Microliter | l | Liter |
| µM | Micromolar | LB | Luria Bertani |
| A | Adenine | M | Molar |
| Amp | ampicillin | mA | Milliamperes |
| BG | Benzyl guanine | mg | Millligram |
| BSA | Bovine serum albumin | min | Minute (s) |
| C | Cytosine | ml | Mililiter |
| dNTP | Deoxynucleotide triphosphate | mM | Millimolar |
| DNA | Deoxyribonucleic acid | mRNA | Messenger ribonucleic acid |
| ECL | *enhanced chemiluminiscence* | ng | Nanogram |
| EDTA | Ethylenediaminetetraacetic acid | nM | Nanomolar |
| EtBr | Ethidium bromide | O.D | Optical density |
| E. coli | *Eschericha coli* | ORF | Open reading frame |
| em | emulsion | PBS | Phosphate buffered saline |
| FACS | *fluorescence activated cell sorter* | PCR | Polymerase chain reaction |
| Fig. | Figure | Pol | Polymerase |
| G | Guanine | pM | Picomolar |
| h | Hour (s) | RLU | Relative light units |
| HIV | *human immunodeficiency virus* | RNA | Ribonucleic acid |
| IgG | Immunoglobulin G | RNAse | Ribonuclease |

| | |
|---|---|
| Rpm | *revolutions per minute* |
| RT | Reverse transcriptase |
| SDS-PAGE | Sodium dodecyl sulphate-polyacrylamide gel electrophoresis |
| sec | Second (s) |
| T | Thymine |
| TAE | Tris- acetate - EDTA |
| TBE | Tris- acetate- Borate |
| TE | Tris/EDTA |
| v/v | Volume per volume |
| w/o | water in oil |
| w/v | Weight per volume |
| WB | Western blot |

# Summary

Generation of a large number of variants and high-throughput selection of the best variants during re-iterative rounds has proven to be a successful strategy to improve biological and physical properties of nucleic acids and proteins. This approach generally depends on the maintenance of a stable linkage between the genotype and the phenotype during the selection procedure and on efficient determination of the genotype encoding the selected phenotype. Phage display technology has been the classical and most successful platform for in vitro evolution strategies till date. However with the advent of exclusively *in vitro* systems like ribosome display or covalent DNA display, there has been a tendency to cell free evolution systems because of their advantages like increased complexity of library to be displayed or unrestrictive selection conditions. However, some of these new approaches can be applied only for a narrow range of applications and also most of these systems are not monoclonal in their approach. Another practical difficulty of these DNA displays is that the genotype responsible for the selected phenotype has to be determined from single DNA molecules. To address these hitches, a novel multicopy bead display approach was developed for the *in vitro* evolution of nucleic acids and proteins.

In this display approach, a single bead is simultaneously linked to multiple copies of a clonal gene variant and to multiple molecules of the protein encoded by the bead´s gene variant. The beads then serve as the unit of selection. This display approach involves two steps of compartmentalization using water in oil emulsions. In the first step, an emulsion PCR is performed, in which amplicons from single template molecules are amplified on beads present in the same picoliter reactor. Beads are then transferred to a second emulsion for an *in vitro* transcription translation reaction, in which the protein encoded by each bead´s amplicon covalently binds to the bead present in the same picoliter reactor. The beads can then be isolated from the emulsion and used in subsequent selection procedures.

As a proof of concept, a randomized library of the T7 promoter was successfully screened for high expression levels by flow cytometry and identified a T7 promoter variant from this library which has a tenfold higher *in vitro* transcriptional activity compared to wild type T7 and results in twofold higher protein expression levels in *in vitro* transcription translation reactions confirming the applicability of multicopy bead display approach as an alternative platform for *in vitro* evolution.

As exemplified by the successful T7 promoter screen, the multi-copy bead display approach combines a number of advantages that should provide broad applicability and robustness. Microbeads used as the unit of selection, are rather inert, they can be transferred from one compartmentalized reaction to the next. A practical advantage of having multiple copies of DNA on each bead is that the genotype of single beads as units of selection can be easily determined by a one-step PCR and subsequent sequence analysis of the amplicons without the need for any additional cloning steps. Due to the advantages discussed above, the multi-copy bead display approach should be a useful addition to existing *in vitro* evolution methodologies with broad applicability in protein engineering and synthetic biology.

# 1. Introduction

## 1.1 In vitro evolution

Evolution is a continuous process to instill any change across successive generations in the inherited characteristics of biological populations. Evolutionary processes give rise to diversity at every level of biological organization including species, individual organisms and molecules such as DNA, RNA and proteins. Charles Darwin was the first scientist to formulate a compelling theory on evolution by natural selection in his famous book "*On the origin of species*". There are three important chemical processes that make up Darwinian evolution which work in a concerted fashion both at molecular as well as at the level of an organism [1] : (i) Amplification or genetic replication in progeny, (ii) Mutation or creating genetic diversity in progeny, and (iii) Selection of traits, which are desirable of the evolved progeny and are systematically favored. The third property is related to the phenotype rather than genotype. Understanding and utilizing these three important features of evolution, scientists have tried to recreate evolution in laboratory. Sol Speigelman in 1960 was the first scientist to conceptualize this idea and implement to the evolution of a self- replicating RNA molecule from Qβ bateriophage completely *in vitro* [2,3]. Through many steps of incubation and transfer, he could isolate Qβ bateriophage RNA molecules which were amplified 15 fold faster than at the beginning. This was the first successful experiment of *in vitro* evolution. However, it was not before 1990 in two different studies independantly, where *in vitro* evolution was fully applied to *in vitro* selection or SELEX (systemic evolution of ligands through exponential enrichment), in which novel RNA ligands were selected through different rounds by either binding to RNA polymerase or

different dyes [4,5]. These two reports and other molecular evolution studies allowed researchers to understand the scenario that is believed to exist during origin of life or the RNA world [6–8].

A lot of research has also been done to understand the molecular evolution of proteins [9–11] in parallel to evolution of nucleic acids. Molecular evolution of proteins mostly consist of two features of evolution - mutation and selection, and does not have the amplification step [12]. Mutation occurs at the level of genes thereby resonates in the quality of the proteins and selection depends on the particular property of the protein. However in case of proteins, the selection step is the toughest especially when applied to high throughput settings. Selection criteria in *in vitro* evolution of proteins always have to be designed according to the function of the protein [13–15]. *In vitro* evolution strategies provide proteins either with improved physical and/or chemical properties or proteins with novel functions. *In vitro* evolution of proteins has been successfully applied in diverse fields of medical [16], industrial [17] and basic science [18].

One of the important steps in *in vitro* evolution of proteins is formation of a genotype-phenotype linkage. Proteins selected through *in vitro* evolution do not have the ability to amplify themselves, so their function is always dependent on the nature of the gene encoding them. Selection criteria based on the properties of a protein also necessitate the importance of selecting the genes encoding the particular phenotype for further characterization of the evolved phenotype. This is an important difference between *in vitro* evolution of nucleic acids and *in vitro* evolution of proteins. Therefore the linkage between a gene and its encoding protein has to be generated physically for the success of the *in vitro* evolution of proteins. With this knowledge, different display technology platforms have been developed to aid *in vitro* evolution of proteins. Most of these systems are partially *in vitro* and partially *in vivo*[19–21]. However in recent times a number of exclusively *in vitro* systems have been developed[22–24]. Each of these strategies will be discussed

by its principle, application potential and its advantages and limitations in the following sections.

## 1.2 Phage Display Technology

Phage display approach is the most classical and extensively utilized *in vitro* evolution strategy till date. It was developed by George P. Smith, who demonstrated that foreign DNA fragment can be inserted into gene III of a filamentous phage which codes for the coat protein and protein encoded by the DNA is displayed on the virion surface without any loss of its function [19]. It has been highly successful in a broad range of applications [25–28] including display of large antibody libraries for the discovery of novel therapeutics [29,30]. As illustrated in the figure 1.1, the basic principle of phage display involves a DNA library encoding for a ligand [peptides, proteins or antibody fragments] that is batch cloned into the phage genome as a fusion protein with one of the genes encoding for phage coat proteins [pIII [31] , pIV [32], pVIII [33]]. These coat fusion proteins are incorporated in a phage particle through assembly in a bacterium. Upon maturation, the ligands are displayed on the phage surface with the genetic material encoding for them being present within the phage. This genotype-phenotype linkage then allows the enrichment of a specific phage population through a repeated process called bio panning. It involves selection of phages carrying the protein with desired properties on an immobilized target (such as antigen etc). Phage particles that display a relevant ligand will be retained, while non-adherent phages will be washed away. Bound phage can be recovered from the surface, reinfected into bacteria and grown for further enrichment, or can be used for characterization. The most common phages used in phage display systems are M13 [34] and fd filamentous phages [35]. Also T4 [36], T7 [37] and λ [37] phages have been featured in phage display systems.

The correct assembly of the phage particles inside the bacteria is an important step in phage display. For this purpose bacteria with the F pilus (fertility factor) is

employed [19]. Once transformed in the *E.coli* containing F plasmid, phage particles containing natural filamentous phage genome attach to the tip of the F pilus that is encoded by genes on this F plasmid, and the phage genome as a circular single-stranded DNA molecule, is translocated into the cytoplasm. The genome then replicates involving both phage and host derived proteins, and is packaged by the infected cell into a rod-shaped particle which is released into the media. This will lead to the presence of heterologous proteins in all the coat proteins of the phage. Another system is to use a phagemid which is a simplified display vector (as reviewd by Lowman et. al. [38]) (containing only pIII coat protein) under the influence of a weak promoter, phage particles will not be released from the *E. coli* cells until they are co-infected with a helper phage, which enables packaging of the phage DNA and assembly of the mature virions with the relevant protein fragment as part of their outer coat protein (mostly one coat protein). Natural filamentous phages generally give rise to polyvalent display of peptides (proteins displayed on all the coat proteins of one phage) whereas phagemid give rise to monovalent display (one peptide is displayed on the coat protein of one phage). This display valency is important factor to discriminate between binders of different affinities. It has been reported that polyvalent display could prevent selection of individual clones with high affinity because multivalency could confer high binding avidity on weak-binding clones [39]. Therefore, monovalent display is sometimes more preferred than polyvalent display. Conversely using a phagemid containing a weak promoter can lead to less transformation efficiency in bacteria [38]. The choice for polyvalent or monovalent display completely depends on the protein that is in investigation. There are also frequent experimental strategies which involve initial polyvalent display and then shifting to monovalent display such as affinity maturation strategies [29,30].

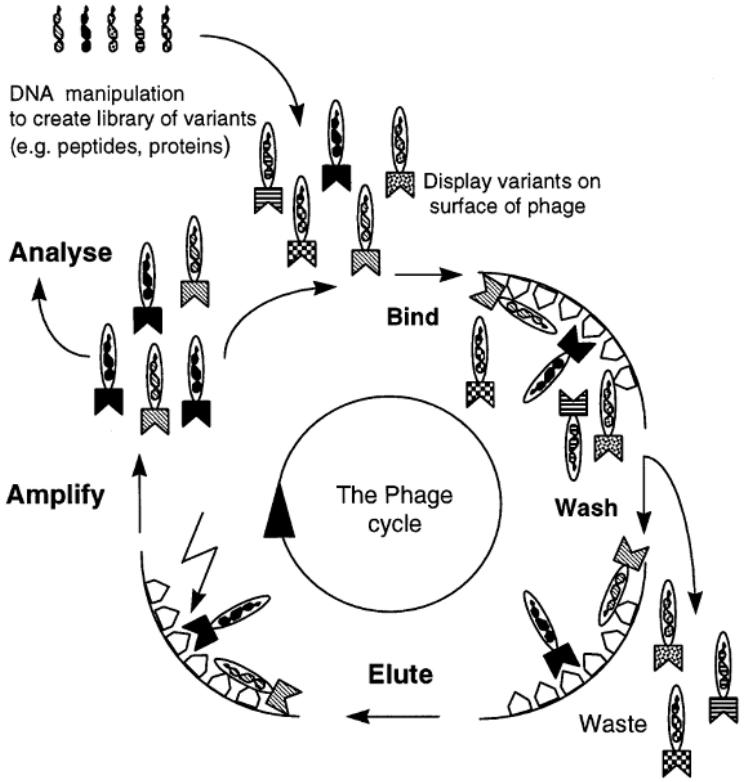



**Figure1.1 Schematic representation of the basic principle of phage**. DNA library encoding variants of certain ligands (e.g. peptides, proteins or antibody fragments) on the phage coat proteins (pIII, pVI or pVIII). Libraries of large diversity with phage can be obtained by cloning into *E.coli*. From these phage repertoires, phages expressing specific ligands can be isolated by repeated cycle of display and selection on antigen through binding, washing, eluting and amplifcation. Figure adapted from [40].

Development of phage display revolutionized research in different fields of biology especially immunology. A typical phage display system can display a libray with a complexity of $10^8$-$10^9$. Antibodies were the first proteins to be successfully displayed on the surface of a phage [41]. Different antibody libraries depending on the source of immunoglobulin genes either from immunized mice or immune donors [42], non immunized or naive [43] and also synthetic [44] have been successfully displayed through phage display library. As an example the first fully human antibody to TNF-α, HUMIRA (Abbott laboratories) was developed and is used as drug currently, key components of which were found through display human antibody fragments on phages [45]. Phage display has also been applied to evolution of receptor agonists [27] and antagonists [28] to target proteins. The huge

potential of phage display also helped researchers in drug discovery. But high throughput screening techniques such as flow cytometry etc cannot be applied in phage display system, making it very laborious and time consuming [20,21]. There are more limitations of this system which will be pointed out in later sections.

## 1.3 Cell surface display

Cell surface display is a different strategy, in which living cells are used for linking of phenotype and genotype. This method also involves fusing the protein of interest to the cell membrane protein so that it is presented on the cell surface. They generally can display a library complexity between $10^6$-$10^7$. The most important cell surface display methods are bacterial display and yeast display.

### 1.3.1 Bacterial display system

This display system was pioneered by Georgoiu *et al* [20] where they succesfully displayed single chain antibody fragments on the surface of gram negative bacteria and were able to sort them through flow cytometry. Different types of bacteria can be modified for the purpose of display. Outer membrane protein A (OmpA) of gram negative bacteria *E. coli* has been reported to successfully display peptides of different length [46,47]. The second, third and fourth loop of this membrane protein could be exploited for the best display of peptides in this system [47,48]. The different loops of OmpA are also depeicted in the figure 1.2. Another outer membrane protein LamB was subsequently shown to display peptide libraries and peptide vaccines [49,50]. Along with it, other outer membrane proteins of gram negative bacteria like Pho E, invasin etc have also been employed to display miscellanous proteins [51,52].

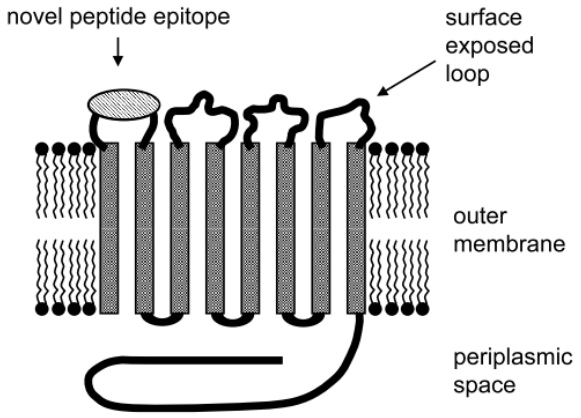



**Figure 1.2 Schematic presentation of E. coli OmpA,** which is one of the many membrane proteins used in bacterial surface display with a novel epitope expressed in a surface-exposed loop. Figure adapted from [53].

Gram negative bacteria are widely studied organisms and therefore they can be critically controlled during display of special proteins such as an enzyme [54,55]
Gram positive bacteria have also been shown as a promising alternative to gram negative bacteria [56–60]. Gram positive bacteria have added advantages such as, translocation only through one single membrane is required for proper display of heterologous proteins, which is different in gram-negative systems where a first translocation through the cytoplasmic membrane and then correct integration into the outer membrane are required for proper display of proteins and thereby making the display process more complicated in gram negative bacteria. Moreover, gram positive bacteria have a thicker cell wall, which would contribute to easier handling in laboratory [56]. However, gram positive bacteria have low transformation efficiency in contrast to gram negative bacteria which would compromise library complextity to be displayed on their surfaces. Nevertheless, cell wall bound proteins like SPA[57] , M6 [56], cell membrane proteins like DppE [58], CwbA [58] or different cell surface associated proteins [59,60] have been successfully exploited to display heterologous proteins in gram positive bacteria. Bacterial surface display of proteins has also been applied in different fields of research. In immunology, live bacterial vaccine delivery[49,58,60] and epitope mapping of immunological active

proteins [61,62]) is shown as a possibility through bacterial surface display. Different enzymes for industrial applications were also evolved through this approach [55,63]. Bacterial surface display has also found its role in the display of antibody fragments and also various binding proteins [20,50].

### 1.3.2 Yeast surface display

Broder and Wittrup [21]were the first to display a combinatorial peptide library on surface proteins of yeast . Yeast display has certain advantages over bacterial display system [21,64]. Yeast system has more homology to a mammalian system concerning protein folding and secretion. Yeast cells can be cultured as single cell microbe and thereby suited in library based settings. Moreover, yeast has stronger cell wall than bacteria making them rigid to undergo various laboratory processes without lysis. The principle of yeast display mainly consists of two important protein systems [64]. The first protein which is used for display in yeast is agglutinin protein. In this system, the mating specific agglutinins which mediate cell adhesion during mating between different cell types are present on the outermost surface and are exploited for display. Two types of agglutinins- α- agglutinin from α-cells and a-agglutinin - which are mating specific are equally utilized. . Heterologous proteins are targeted to the outermost glycoprotein layer on the cell wall through these agglutinins. The peptide library can be fused to the C- terminus of the agglutinins. In case of α- agglutinin [65] the library is anchored to yeast cell surface through GPI anchor segment and displayed on the cell surface as shown in figure 1.3 A.In the second type, a-Agglutinin consists of a core subunit encoded by AGA1, which is involed in secretion and is linked through disulfide bridges to a small binding subunit encoded by AGA2. Both these subunits also contains secretion-signal region and also GPI-anchor-attachment signal.The peptide library remains fused in the c terminus of Aga2p as fusion protein. The Aga2p fusion protein and Aga1p associate within the secretory pathway, are exported to the cell

surface and covalently linked to the cell wall and thereby displaying the heterologous protein [21]as shown in the figure 1.3 B.

The second protein which is used for display on yeast surface is Flocculin. It is encoded by FLO1 gene. It is a lectin like protein, which is involved in flocculation [66]. This protein is considered to adhere non covalently to α mannan carbohydrate residues on the cell wall which causes cells to flocs [67]. Again two types of display are possible with this protein. The library is again fused to the C- terminus of FLO1 gene and the protein then expressed gets fused to the cell wall through GPI anchor segment [68] as shown in figure 1.3 C. Another way is to fuse the target protein in the N-terminus of the flocculation signal domain between proteins namely FS and FL [69].The protein then produced non covalently interacts through its flocculation functional domain and the mannan chain of the cell wall and is thereby displayed on the surface as shown in the figure 1.3 D

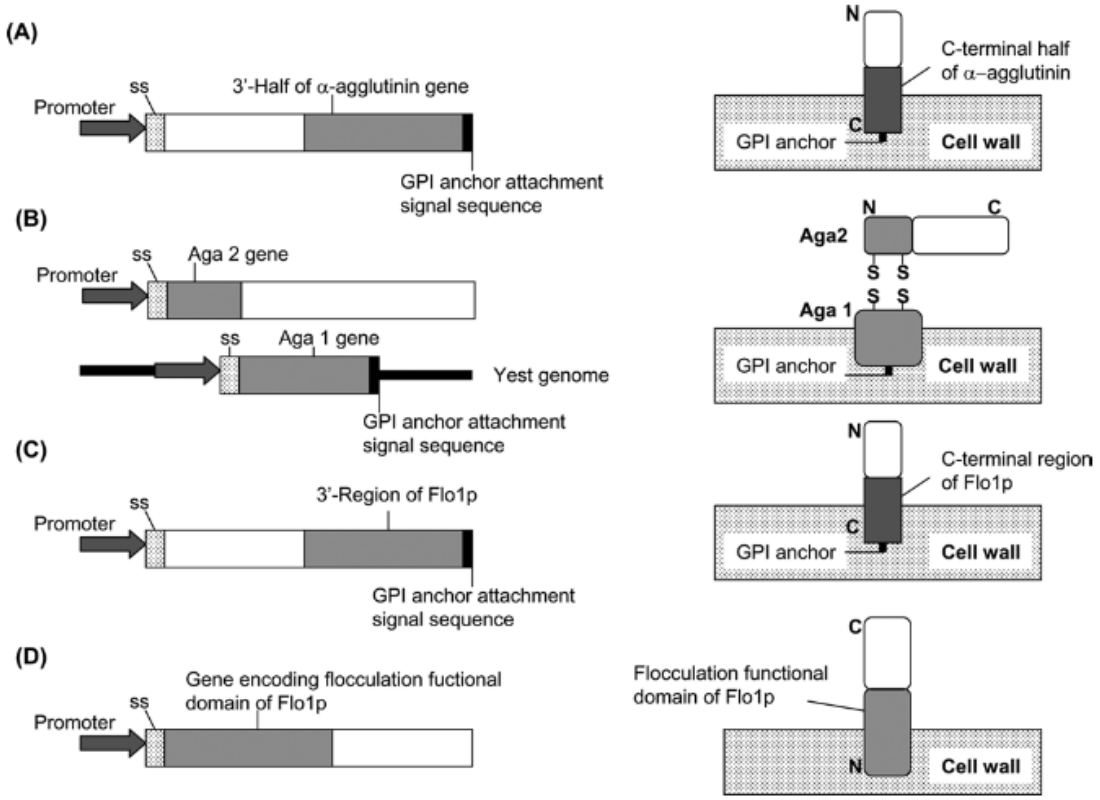



**Figure 1.3 Schematic presentation of different protein systems used in yeast cell surface display** -(A) α-agglutinin, (B) a-agglutinin, (C) C-terminus region of Flo1p, and (D) N-terminus flocculation functional domain of Flo1p. Figure adapted from [64].

Yeast display approach has been successully applied to the display and screening of a number of different industrially important proteins like amyloytic enzymes [70], cellulolytic enzymes [71] etc. Yeast cell-surface display systems have been found to be effective for the display of single-chain antibody (scFv) and the development of antibodies with enhanced affinity and stability [21,72]. Even a combinatorial protein library was successfully generated and displayed through yeast surface display [73].

Cell surface display approaches are comparable to phage display. However, they have certain advantages over the phage system. The most important advantage is that, it is possible to use flow cytometry for screening [12]. This allows for screening of more than 50,000 events per second making the display process very fast. Fluorescence labeled antigens could bind to high affinity displayed proteins on *E. coli.* Also examples of flow cytometric based screening of enzymes displayed by cell surface has been reported [74]. Morevover, the absence of precise screening methods such as flow cytometry makes phage display susceptible to huge leakage of false positive during screening steps. Taken together, cell surface display systems can circumvent these disadvantages of phage display system.

### 1.3.3 Other cell surface display

Apart from above mentioned highly successful display systems, other display systems with living cells were developed which are worth mentioning. Autographa californica nuclear polyhedrosis virus (AcNPV), a eukaryotic virus, whose surface glycoprotein was modified to display HIV gp120 in its functional form [75]. Human Rhinovirus was also exploited to display a mimotope (a peptide molecule which mimics the structure of an epitope) library from membrane proximal external region (MPER) of HIV to elicit broadly neutralizing antibodies to HIV [76]. Another study showed that the surface of self inactivating lentivirus could be employed to display stable single chain human antibodies [77]. Mammilian cell surface system has also been developed and reported to display peptides and

antibodies [78]. One such mammalian display system is Lymphocyte display where, T-cells were engineered to display single chain antibodies, is an essential advancement in cell surface display [79]. Although, a lot of different approaches have been developed and shown to be successful in displaying proteins, the most important constrain remains the complexity of the library displayed through these approaches. The library size displayed is much lesser than phage display making these approaches less attractive [79].

These partially *in vivo* systems including phage display have certain disadvantages. Modulating and manipulating living organism is laborious and time consuming and also selection criteria are restrictive [80]. For example it can be highly difficult to select proteins with extreme pH or to utilize un-natural amino acids in these systems. It has also been reported that proteins over-expressed in E.coli can cause severe growth defects and also growth inhibitions in the host [81]. Other studies too claimed that antibody fragments displayed through microbial cell surface display cannot be reformatted easily to produce well expressed IgGs [82] and they also tend to precipitate in the cytoplasm [83]. Moreover, the most crucial fact is that the library size can be strongly limited due to transformation or transfection inefficiencies in live organisms [21,23,84,85] Most of these issues have been addressed by the development of exclusive *in vitro* display systems such as ribosome display [22], mRNA display [86] or *In vitro* compartmentalization [24] etc.

## 1.4 Ribosome display

Ribosome display approach was the first exclusive *in vitro* system developed for the *in vitro* evolution of proteins. The concept was first introduced by Mattheakis and co-workers where they could select a short peptide from a library by using polysomes [87]. This was further developed by Hanes *et al,* where they could evolve single chain antibodies to haemagglutinin [22]). They demonstrated that after five rounds of display an $10^8$ fold enrichment of these novel antibodies can be achieved. The principle of ribosome display begins with the transcription of naive DNA

libraries encoding polypeptides. Each of these DNA sequences is transcribed and translated to polypeptides through a cell free expression system - both prokryotic (*E.coli* S30 extract) [87] or eukaryotic (rat reticulocyte extract) [88,89]. This leads to the formation of a protein-ribosome-mRNA complex (PRM). The nascent polypeptide remains linked to its corresponding mRNA non-covalently and this permits the selection of the genetic material through binding property of its encoded protein. This mRNA can then be extracted and amplified by RT- PCR and can further be mutated, cycled or cloned. The general principle of ribosome display is depicted in figure 1.4 which shows the display of antibody libraries.

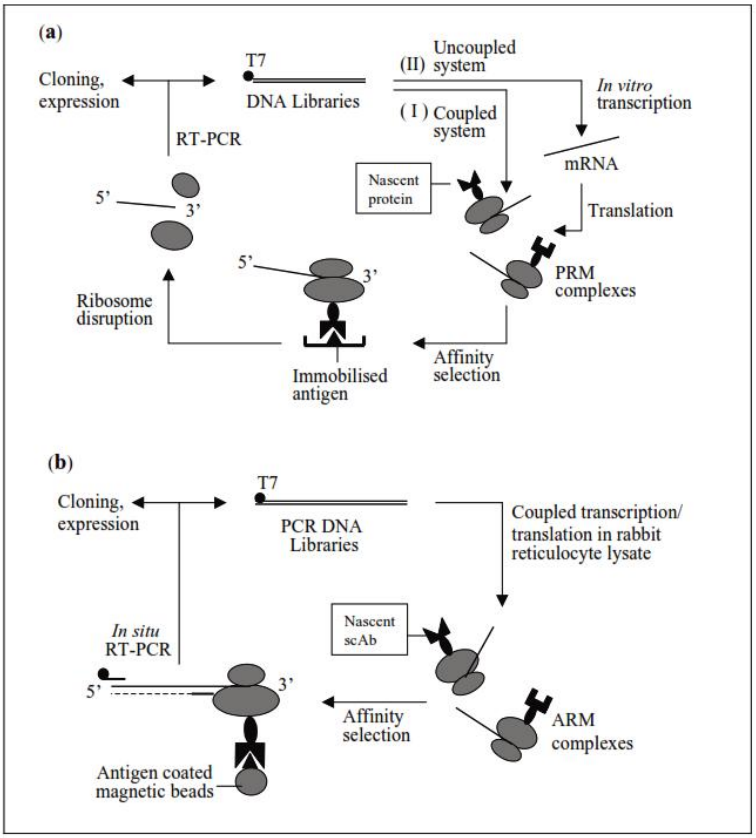



**Figure 1.4** Schematic presentationpresentations of Prokaryoticand eukaryotic ribosome display and selection cycles of antibody libraries. (A) E. coli ribosome display: (I) using coupled (transcription and translation coupled in one reaction) E. coli S30 extract and (II) using uncoupled (transcription and translation reactions are separate) E. coli S30 extract. (B) ARM ribosome display for single-chain antibodies. T7: T7 promoter; PRM: protein–ribosome– mRNA; RT-PCR: coupled reverse transcription polymerase chain reaction; scAb: single chain antibody fragment; ARM: antibody–ribosome. Figure adapted from [90].

The crucial feature of this system is to stall the ribosome so that the nascent protein and mRNA remain associated in the PRM complex. This can be either achieved through (i) halting the translation of ribosome at random with the use of antibiotics such as rifampicin and chloramphenicol (for prokaryotic ribosomes) or cyclohexamide (for eukaryotic ribosomes) [87,91] or (ii) deletion of the stop codon, which is normally recognized to trigger detachment of the nascent polypeptide at the 3' end of mRNA [22,88]. Ribosome display being the first completely *in vitro* system offers certain advantages over *in vivo* display approaches [90]. Due to the lack of transformation, large libraries with diversity ($>10^{12}$) can be displayed. Mutagenic PCR can be employed to further enhance the diversity of libraries [92,93]. It also allows modified amino acids such as chemically labeled or unnatural amino acids to be incorporated into the protein at defined positions which is not possible in cell based display approaches [94]. Moreover, eukaryotic cell free expression systems, which are used in this approach are capable of a variety of post translational modifications, which expands the possibility to study proteins with regulatory functions. In addition to the above mentioned applications, ribosome display has also been employed for the development of the first Human synthetic antibody library-HuCAL (by a Germany based company MorphoSys) [93] from which many members are already in clinical trials against cancer. Ribosome displayed antibody libraries have also been reported to produce antibodies with higher affinity and stability [22,95]. It has also been applied to high throughput proteomics based applications [96]. However, there are certain disadvantages of ribosome display as it will be pointed in the next section.

## 1.5 mRNA display

mRNA display system was developed during same time as ribosome display. It was developed by R W Roberts and J W Szostak in 1997. They showed in a model selection, that a mRNA fused with c-Myc peptide epitope can be enriched 20- to 40-fold from the mixed population by immunoprecipitation [86]. Using this system they could also evolve functional ATP binding protein from a random library of

the size of 80 amino acids following 18 rounds of display and selection [97]. The first real application was reported by Fukuda *et al* in 2006 where they displayed functional single chain antibodies [98]. The principle of mRNA display is similar to ribosome display. Like ribosome display, it can also display large libraries with complexity of more than $10^{13}$. In this approach, encoded peptide and protein libraries are covalently fused to their own mRNA. Fusion synthesis is possible because the messenger RNA can act as both template and peptide acceptor if it contains a 3'-puromycin molecule. As translation proceeds, the ribosome moves along the mRNA template, and once it reaches the 3' end of the template, the fused puromycin enters ribosome's A site and get incorporated into the nascent peptide. Puromycin is an analogue of the 3' end of a tyrosyl-tRNA with a part of its structure mimicing a molecule of adenosine, and the other part mimicing a molecule of tyrosine. Compared to the cleavable ester bond in a tyrosyl-tRNA, puromycin has a non-hydrolysable amide bond. As a result, puromycin interferes with translation, and causes premature release of translation products. The genotype-phenotype linkage is made through a linkage between the puromycin and the last amino acid of the polypeptide. This mRNA-polypeptide fusion complex is thus released from the ribosome. Also a spacer has to be included in the mRNA template between 40- 60 nucleotide long to provide flexibility to puromycin for its proper entry in ribosome [99] . The mRNA-polypeptide fusion is selected through ligand binding on an immobilized target. This mRNA can then be extracted and amplified by RT- PCR and can further be mutated, cycled or cloned. The principle of mRNA-display and selection cycles is shown in the figure 1.5.

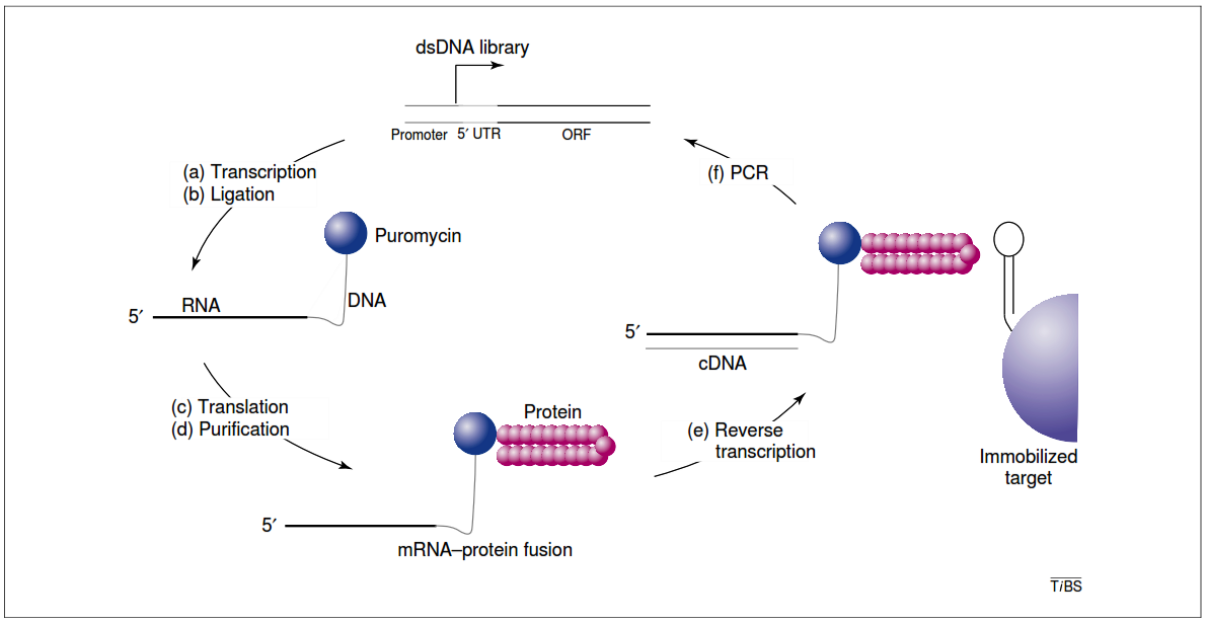



**Figure 1.5 Schematic representation of a typical mRNA-display and selection cycle**. (a) A library of dsDNA sequences is transcribed to generate mRNA. (b) The mRNA is ligated to a puromycin oligonucleotide (blue) and this mRNA is used as a template in an *in vitro* translation reaction.(c). (d) cDNA synthesis is performed (d,e) cDNA/mRNA–protein fusion is sieved using the target of interest.(f) PCR is used to regenerate the full-length DNA construct. Figure adapted from[23].

Although mRNA display and ribosome display work on a similar principle, there are certain advantages of mRNA over ribosome display [100]. In ribosome display the stable stalling of PRM in a complex is the most critical point. This is the disadvantage of this technology, which would affect the selection stringency because of only non covalent interaction within the PRM complex. This might cause unspecific interaction with ribosomal parts, giving rise to high background binding during the selection cycle. The covalent binding of the polypeptide to its coding mRNA and the absence of ribosomes is the most important advantage in mRNA display. Again, the nascent protein is attached to an enormous complex of mRNA and ribosome (a ribosome has a molecular weight of 2000 kDa). This might lead to unpredictable interactions between protein displayed and the ribosome and this may lead to loss of potential binders during the selection cycle. Inclusion of a spacer in mRNA display which is much smaller than a ribosome and would less likely interact with the target protein and thereby would offer the results with less

biases. Apart from its application in evolution diverse peptide aptamers (aptamers are nucleic acid or protein molecules which act as ligands and bind to specific target molecules) such as ATP aptamers [100], streptavidin aptamers [101], TNFα aptamers [102] etc, mRNA display has been applied to some unique applications. mRNA- protein fusion from mRNA display can used in high throughput screening strategies such as developing protein chips with this complex and thereby converting the system into self assembling protein micro arrays [103] . One of the most interesting applications of mRNA display was the incorporation of unnatural amino acids. Liu *et al.* demonstrated that the suppressor tRNA strategy for incorporating unnatural amino acids [104] can be used to create mRNA-display libraries bearing an unnatural residue [105]. However, one of the major limitations of mRNA display and ribosome display technologies is that mRNA is used as the library-encoding nucleic acid, which may be prone to rapid degradation. Also, the mRNA selected through this display requires a reverse transcription step before amplification of the selected sequence, which might lead to the formation of unstable RNA-DNA hybrids and this could hamper subsequent characterizations steps. Moreover, mRNA display is highly demanding and tedious and not easy to replicate under usual laboratory conditions and requires a huge amount of expertise in this system [100]. To address these limitations, DNA display systems were developed to further improve *in vitro* evolution strategies.

## 1.6 CIS and CAD display

DNA-based display systems were developed to offer speed and stability over RNA templates as the DNA template is less sensitive to degradation [106,107]. Furthermore, libraries can be quickly generated by standard PCR procedures. One such system is CIS display which exploits the ability of a bacterial replication initiator protein, RepA, which exclusively binds back to its encoding DNA (termed *cis*-activity, hence the term- CIS display). This was developed by the McGregor group where they demonstrated recovery of a specific binder from a pool of non-binding members present at a ratio of 1 in $10^{10}$ [106]. The activity of RepA protein

is dependent upon two non-coding regions 3' to the *repA* sequence that are essential for *cis*-activity. These regions are termed *CIS* and *ORI*. The *CIS* element is believed to be involved in fusing and loading of the protein in *cis* whereas *ori* contains the binding region for RepA. The mechanism is unknown but it is believed that the RNA polymerase is stalled during transcription at the *CIS* element which allows the nascent RepA protein, which has been translated concurrently, to non-covalently attach to its binding site in the *ori* region of its own template. The DNA can be selected on the basis of the property of the protein, amplified and characterized or re used in the second round of display. The principle of CIS display is shown in figure 1.6.

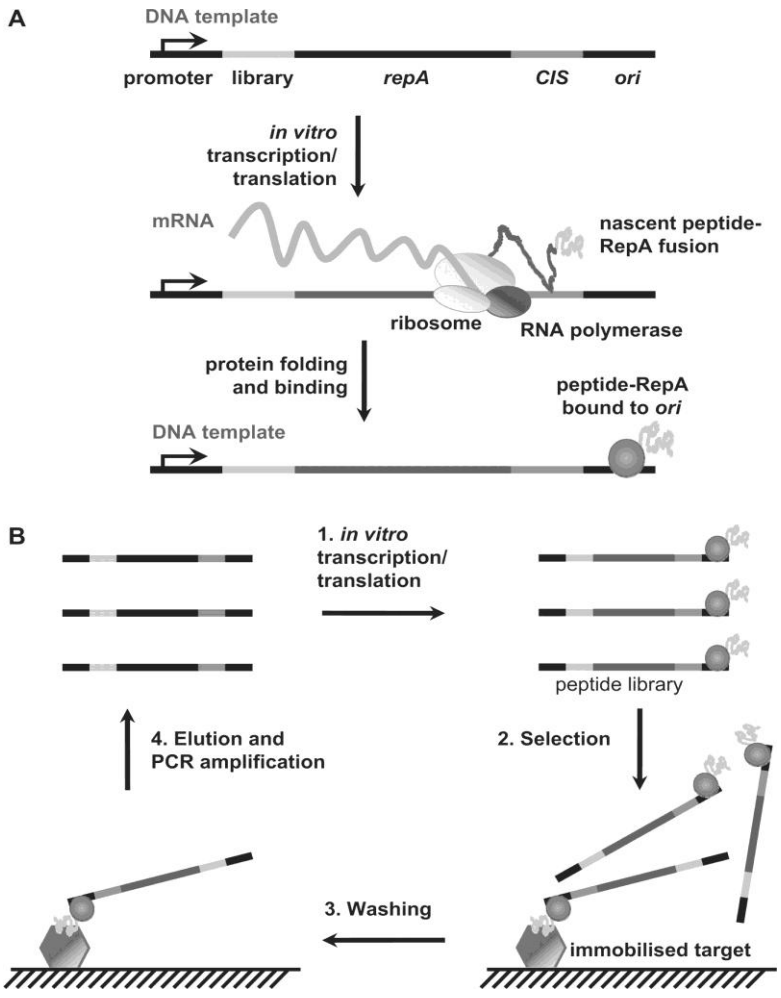


**Figure 1.6 *(continued)* Schematic presentation of principle of CIS display system**. (A) Transcription of template DNA encoding an N-terminal library fused to the RepA sequence is paused when the RNA polymerase reaches the CIS element. Translation produces the RepA protein, which transiently interacts with the CIS element, thereby forcing its subsequent binding to the adjacent ori sequence establishing linkage between a template DNA and the expressed polypeptide that it encodes for. (B) CIS display requires transcription/translation of a double-stranded DNA template, encoding peptide libraries, to generate peptide- DNA fusions (Step 1). These are then selected against a target (Step 2) and non-binders are removed by washing (Step 3). PCR is then used to amplify DNA from the bound peptide- DNA complexes, ready for the next cycle of CIS display (Step 4). Figure and caption adapted from[106].

Another system, covalent antibody display (CAD), uses a *cis*-acting DNA binding protein (bacteriophage endonuclease P2A) that covalently links to its own coding strand through the activity of a catalytic tyrosine within its sequence[108]. This system has been used to select for tetanus toxin binders from an immune human library with enrichment rates between 14- and 300-fold [108].

The research and development in the field of *in vitro* evolution of proteins has been unprecedented. Most of these systems although an improvement over their competing approaches, however most of them are difficult to replicate in the normal laboratory settings or some of them are limited in their application (such as DNA display). These systems further require being fast enough to be able to screen large libraries in short time. Flow cytometry based screening can speed up the screening process; however it cannot be applied to the above discussed *in vitro* display strategies. The combination of *in vitro* compartmentalization with DNA display has the possibility to address these limitations.

## 1.7 In vitro compartmentalization

*In vitro* compartmentalization (IVC) was first developed by Tawfik *et al,* [24] which is essentially a water in oil (w/o) emulsion based system containing millions of picoliter reactors. This was introduced with an idea to mimic cellular compartmentalization. Each of these reactors have a mean diameter of 2-5 micrometer and containing reagents of femto liter range [24,109–112]. IVC was first successfully applied to *in vitro* evolution studies where Tawfik *et al* specifically

isolated a DNA binding enzyme HaeIII methyl transferase gene from a mixture with containing another gene encoding fol A enzyme which was present in $10^7$ fold high amount compared to HaeIII methyl transferase [113] . They could further apply IVC to evolution of a protein phospodiesterase which showed a 63 times higher Kcat than the wild type enzyme. They generated a library of $10^7$ variants which were screened and could show that catalytic properties of an enzyme can be exploited in IVC. They also demonstrated the amount product formed could be screened with anti-product antibody in flow cytometry thereby emphasizing the applicability of flow cytometry to *in vitro* evolution [113] with IVC. In addition to *in vitro* evolution, IVC has been applied successfully to the development of diagnostic strategies such as BEAMing which is applied for the detection of mutations in plasma of cancer patients [114,115]. IVC has also revolutionized next generation sequencing technology [116].

The principle of IVC in *in vitro* evolution strategies involves the DNA library either linked to a solid surface such as a microbead [110] or free, encoding DNA binding proteins [24], or DNA with certain modifications such as biotin [111], BG [117], or an enzyme substrate[113] etc are emulsified in a water in oil emulsion in such a way that not more than one copy of DNA is present in each of these droplets. These droplets also contain reagents required for translation of the gene. After translation, the protein binds to its encoding gene, through either its DNA binding domain [110], through a chemical linking such as biotin-streptavidin [111], etc or of a substrate through a catalytic domain [113]. These DNA protein conjugate can be extracted from these droplets and selected depending upon the catalytic [24,118–120], binding [109,111,121,122] or regulation [123] property of the encoded protein. After selection, the DNA can be amplified and it subsequently either undergoes another cycle of IVC for enrichment of the positive sequence or is characterized by sequencing. A typical example of the principle of IVC in *in vitro* evolution of proteins is depicted in the figure 1.7.

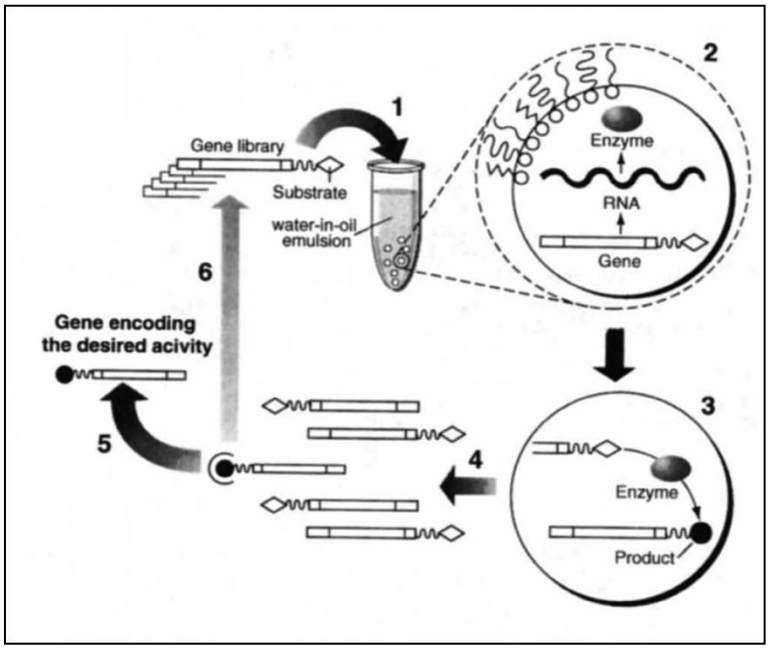



**Figure 1.7 Schematic representation of principle of IVC**. Step 1: Gene library in *in vitro* transcription/translation reaction mixture is emulsified in water in oil emulsion. The gene product also contains a substrate to capture the protein products. Step 2: The gene transcribes and translates to respective proteins. Step 3: the protein converts the substrate on the DNA through its catalytic property and remains linked to the DNA. Step 4: The DNA-protein complex is extracted from the emulsion. Step 5: The genes linked to proteins which are selectively enriched are amplified and characterized or linked again to the substrate and cycled for another round of IVC (Step 6). Figure adapted from [24].

The combination of IVC and DNA display can at least theoretically address most of the limitations of other *in vitro* evolution systems. The most important feature of any evolution or display system is the generation of a genotype-phenotype linkage. This allows the tracking of the genes encoding the respective proteins. Different methods have been possible to create this both covalent and non covalent physical linkage in IVC. This was first realised in DNA display by a system called STABLE developed by Doi *et al.* [111], where they coupled the DNA library members with biotin. This DNA contains the streptavidin encoding partner, which when translated to protein, will bind to its coding DNA through streptavidin-biotin conjugation. Although, this conjugation is widely used in molecular biology, but this interaction is less specific and can lead to huge number of false positive. Moreover, streptavidin-biotin conjugates are not as stable as any covalent linkage. Different DNA binding proteins have been evolved and utilized for genotype and

phenotype linkage [24,110]. However, they are application specific and not universally usable. A covalent linkage between genotype and phenotype would provide a more stable linkage during breaking up of the emulsion droplets. Stein *et al* [117] extended the system developed by Jongsma *et al.* [124] where they coupled the DNA with benzyl guanine, which covalently binds to O6-alkylguanine-DNA-alkyltransferase (SNAP), to *in vitro* evolution process through IVC where coding DNA also encodes for SNAP protein and gets linked with BG moeities on encoding DNA for the creating genotype-phenotype linkage. The chemistry of SNAP-BG interaction is described later in this chapter.

From the time of its inception, IVC has found its applicability in the evolution of complex biological proteins [24,109,113,118–120,123,125]. It was demonstrated that IVC based selections can be performed by direct sorting of emulsion droplets through flow cytometry, which definitively expanded the scope of IVC in high throughput screening procedures [126]. Flow cytometry being a highly sensitive technique, enzyme libraries has been selected by expressing either in cell free expression [127], or one solid surface [113] or even on the surface of bacteria [126] respectively in IVC. This was further developed by the amalgamation of microfluidics to IVC. Highly monodisperse microplets at a rate of 10000 per second can be generated for both single emulsions or double emulsions [128–131]. Even cell free translation in these microdroplets in a microfluidic based setting has been demonstrated [132]. This advancement made in IVC based system makes it a highly attractive system in contrast to its competing display approaches.

IVC can display libraries of huge diversity ($\leq 10^{10}$) like other *in vitro* evolution strategies. Apart from this, IVC based *in vitro* evolution strategies have added advantages over other *in vitro* evolution strategies. It can provide greater control in selection process [80]. Also femto liter range of reagents is required in IVC and thereby it has the ability to further define conditions for selection such as reduced reaction time during an enzymatic reaction etc. Athough, IVC theoretically boasts of its capacity to display of huge library size ($>10^{13}$) it is yet to be completely

realized practically. Its main disadvantage is more of a technical issue [80]. The way the emulsions are generated can vary both in size and number between different batches of emulsions preparation and thereby library displaced might be compromised. Microfluidic based emulsion generation could solve this problem [133]. Also, most of these IVC based selection have a very narrow range of possible applications [24,109,111–113,117,120,122,134] and cannot be used universally. IVC based evolution therefore requires the development of a system which is more generic in its applicability and which would further expand the scope of IVC beyond evolution of enzymes only. A general comparison between different display systems is summarized in table 1.1.

**Table 1.1** Comparison of Different Display Systems

| Display Type | Phage Display | Cell Surface Display | Ribosome Display | mRNA Display | CIS/CAD | IVC+DNA display |
|---|---|---|---|---|---|---|
| **Genotype** | Phagemid | Plasmid | mRNA | mRNA | DNA | DNA |
| **Phenotype** | Antibody fragments proteins, peptides | Antibody fragments, Proteins, peptides, enzymes | Antibody fragments, Proteins, peptides, enzymes | Antibody fragments, Proteins, peptides, enzymes | Antibody fragments | Enzymes, antibody fragments |
| **Use of Live organism** | Yes | Yes | Yes | No | No | No |
| **Library complexity** | $10^7$–$10^9$ | $10^6$–$10^7$ | $10^{11}$–$10^{12}$ | $10^{11}$–$10^{12}$ | $10^{11}$–$10^{12}$ | $10^9$–$10^{10}$ |
| **Advantages** | Most popular, highly utilized | Popular, can used for FACS based screening | Large complex library | Large complex library | Medium to large complex library, DNA chemically more stable | Medium to large complex library, High throughput settings, easier to handle |
| **Limitations** | Library complexity compromised due to use of live organism, laborious, time consuming | Library complexity compromised due to use of live organism | mRNA prone to degradation Unpredicted interaction between target protein and ribosome. | mRNA prone to degradation, laborious | Limited application potential | Limited application potential, technical problems in IVC generation. |

## 1.7.1 Chemistry of SNAP-BG covalent interaction

The use of SNAP-BG interaction in biological applications was first demonstrated by Keppler *et al.* in 2003. SNAP-tag which is a variant of human DNA repair protein O6-alkylguanine-DNA alkyltransferase (hAGT) [135] irreversibly transfers the alkyl group from its substrate, O6-alkylguanine-DNA, to one of its cysteine residues. Due to its low substrate specificity, it can also react with the nucleobase O6-benzylguanine (BG) irreversibly as shown in figure 1.8. Characterization of fusion proteins containing SNAP as one of its partner has demonstrated with

different fluorophore containing BG substrate, that the SNAP domain is functional [136]. Oligonucleotides can be labeled with BG and used for protein arrays [124]. The use of SNAP-BG interaction for creating genotype- phenotype linkage in *in vitro* evolution has also been demonstrated [117].

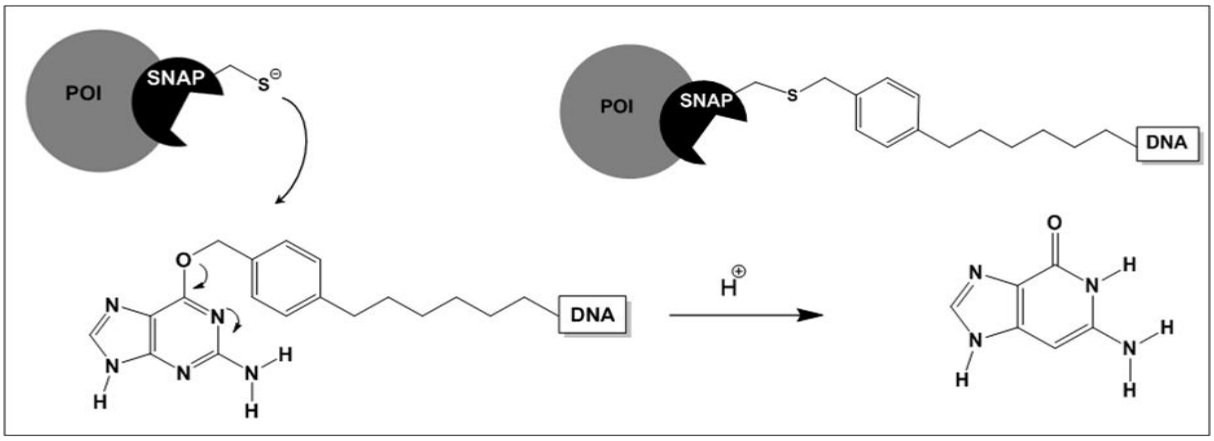



**Figure 1.8 Chemistry of SNAP-BG interaction.** Benzyl guanine can be chemically coupled to DNA or beads (shown here is DNA which contains a benzyl guanine (BG) residue at its free 5´end connected through a C6 linker). Fusion proteins between the protein of interest (POI) and modified O6-alkylguanine-DNA alkyltransferase (SNAP) are covalently coupled to the DNA by irreversible transfer of the alkyl group from BG to one of the cysteine residues of SNAP.

## 1.8 Aim of the Present Study

IVC has a number of limitations as mentioned in the earlier section. In addition to them there are some more drawbacks which require to be pointed out. Apart from catering only a narrow range of applications, their monoclonal continuity was lost during their selection as most of the selection was done outside the compartment [24,109,111–113,117,120,122,134]. Most of IVC involved in evolution strategy allow the presence of single DNA in picoliter reactors which transcribes and translates multiple copies of proteins [24,109,111–113,117,120,122,134]. This DNA then conjugates with only one of the vast excess of proteins that are produced. During the selection process, this can lead to strong competition for ligand binding between one DNA-protein conjugate and the huge amount of unconjugated

proteins. This in turn might lead to the loss of target sequence recovery and thereby make these systems less robust. Moreover, most of these systems do not have a common screening strategy which can be applied to the evolution of any functional nucleic acids and proteins.

The first aim of this present study is to develop a robust and versatile *in vitro* evolution platform, in which the screening reaction is independent from the *in vitro* transcription translation reaction and in which the unit of selection contains multiple copies of the DNA stably linked to multiple molecules of the encoded protein.

The second aim of the study is the proof of the concept of this approach i.e to apply this novel *in vitro* evolution strategy in display and evolution of a novel biomolecule. To do this, a T7 promoter library is chosen to evolve a T7 promoter variant with a higher *in vitro* protein expression activity.

### 1.8.1 Concept of multi-copy bead display approach

The overall concept of the multi-copy bead display approach is outlined in figure 1.9. Linear template DNA encoding a fusion protein between the protein under selection and the O6-alkylguanine-DNA alkyltransferase (SNAP) that reacts specifically with benzylguanine (BG) is clonally amplified in picoliter reactors of an emulsion PCR. One of the PCR primers is covalently coupled to magnetic microbeads while the other contains a BG modification. Multiple BG-labeled copies of the same amplicon are therefore captured by beads that are present in the same picoliter reactor. If the template DNA consists of a pool of different variants, a library of beads is generated with each bead carrying multiple copies of the same DNA variant and different beads representing different variants. After the amplification the beads are purified from the emulsion and added to a second emulsion for an *in vitro* transcription translation reaction. The SNAP domain of the fusion proteins expressed in each picoliter reactor links the protein under selection to the BG moieties coupled to the encoding DNA by irreversible transfer of the

alkyl group from BG to one of its cysteine residues as described in 1.7.1. This results in beads carrying multiple copies of the same DNA variant and multiple copies of the protein encoded by the particular DNA variant. The beads can then be screened and selected by different methods including flow cytometric analysis with antibodies directed against the protein under selection

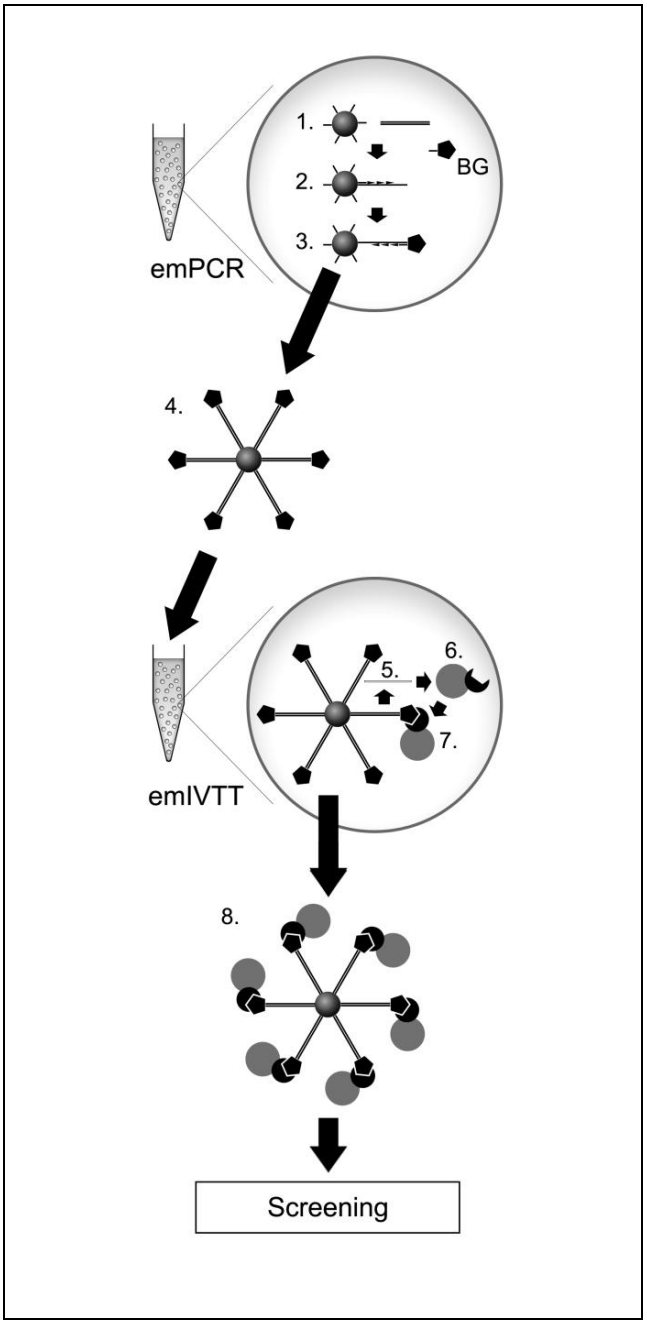


**Figure 1.9 *(continued)* Principal steps of the multi-copy bead display approach**. The bead display approach involvestwo steps of compartmentalization using water in oil emulsions. In the first step, single template molecules of a DNA library are amplified in each picoliter reactor of an emulsion PCR (emPCR) with a forward primer coupled to microbeads and a reverse primer covalently coupled with benzylguanine (BG) moieties (1-3). As a result copies of the same DNA each carrying a BG substrate are immobilized on the bead. These beads are then recovered from the water in oil emulsion (4) and used in a second step in an in vitro transcription translation reaction also performed in emulsion. The amplicons contain a phage promoter (e.g. T7) upstream of an open reading frame encoding a fusion protein between the protein under selection and the modified DNA repair protein O6-alkylguanine-DNA alkyltransferase (SNAP) that reacts specifically with BG mojeties incorporated into the amplicons. During the transcription translation reaction the fusion proteins synthesized in each picoliter reactor of the emulsion (5,6) are covalently coupled to the BG mojeties of the beads present in the same picoliter reactor (7). The beads can then be isolated from the emulsion (8) and used in subsequent screening assays.

# 2. Materials

## 2.1 Bacterial Strains

All the bacterial strains used are subtypes of *Escherichia .coli*

DH5α (Invitrogen)

*supE*44 Δ*lac*U169 (f80*lacZ*ΔM15) *hsd*R17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1

NEB5 α (New England Biolabs)

*fhuA2*Δ*(arg*F-*lacZ)U169 phoA glnV44* Φ*80* Δ*(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsd*R*17*

BL21De3 (Stratagene)

*F-dcm ompT hsdS(rB-mB-) gal λ(DE3)*

## 2.2 Nucleic acids

## 2.2.1 Plasmids

pCR TOPO 2.1 (Invitrogen)

It is used for direct cloning of PCR products through thymidine adenosine complementation.

pJET1.2/blunt (Fermantas)

It is used to clone PCR products through blunt end ligation.

pEGFP-AGTSM

The expression cassette GFP-SNAP is present pET 21 vector for bacterial expression

pMS2CP-AGTSM

The expression cassette MS2-SNAP is present pET 21 vector for bacterial expression

pCR-Luc

The luciferase expressing cassette is present TOPO vector.

## 2.2.2 Oligonucleotide

### 2.2.2.1 Oligonucleotides for multi- copy Bead display

**Table 2.1** List of oligonucleotides (multi-copy bead display)

| Name | Sequence | Binding Site* |
|---|---|---|
| T7-s2+TGSpacer | 5'-Amino-C12-TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAGCAACCGCACCTGTGG-3' | a |
| T7-s2 | 5'-AGCAACCGCACCTGTGG-3' | b |
| T7-consensus-s | 5'-AGCAACCGCACCTGTGGTAATACGACTCACTATAGGGAGATTGTGAGCGGATAACAATTCCCC-3' | c |
| T7-wobble-s | 5'-AGCAACCGCACCTGTGGTAATACGACTCACNNNNNNNNNNTTGTGAGCGGATAACAATTCCCC-3' | d |
| Luc_ov_Rev | 5'-ATGTTTTTGGCGTCTTCCATGGTATATCTCCTTCTAAAGTTAAACAAA-3' | e |
| T7pro_RP_seq2 | 5'-GCCGTAGGTCAGGGTGGT-3' | f |
| T7Term-a | 5'-GGATATAGTTCCTCCTTTCAGC-3' | g |
| T7Term-a (BG)[1] | 5'-GGATATAGTTCCTCCTTTCAGC-3' | h |
| Ovrlprim_NNB_T7 | 5'-GACCACCCTGACCTACGGC-3' | |
| Luc_ov_Fwd | 5'-ATGGAAGACGCCAAAAACAT-3' | |
| Luciferase_RP | 5'-CATTTAGGTGACACTATAGAATAGGG-3' | |

**Table 2.1** List of oligonucleotides (multi-copy bead display)

| | |
|---|---|
| Luc-seq | 5'-AGCTTCTGCCAACCGAAC-3' |
| Luc-RT-FP1 | 5'-GGAAGTCGGGGAAGCG-3' |
| Luc-RT-RP1 | 5'-TCTCACACACAGTTCGCCTC-3' |
| Anti T7bs2 (BG)[I] | 5'-CCACAGGTGCGGTTG-3' |

* Primer binding sites are marked and shown in *results chapter*(p61, Figure4.1).

[I] These primers are covalently coupled with BG-substrate through a thiol group with C6 spacer in the 5'-end. [137]

N= A, G, C or T.

### 2.2.2.2 Oligonucleotides for mutation studies

**Table 2.2** List of Oligonucleotides (mutation studies)

| Name | Sequence |
|---|---|
| T7C62mut1T | 5'-AGCAACCGCACCTGTGGTAATACGACTCACTATCGCGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut2T | 5'-AGCAACCGCACCTGTGGTAATACGACTCACATTCGCGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut3A | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAAACGCGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut4G | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATGGCGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut5C | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCCCGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut6G | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGGGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut7 | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCCGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut8C | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCGCAGTTGTGAGCGGATAACAATTCCCC-3' |

**Table 2.2** List of Oligonucleotides (mutation studies)

| Name | Sequence |
|---|---|
| T7C62mut9T | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCGGTGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut10C | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCGGACTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut6A | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGAGGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut6T | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGTGGAGTTGTGAGCGGATAACAATTCCCC- |
| T7C62mut7A | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCAGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut7T | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGCTGAGTTGTGAGCGGATAACAATTCCCC-3' |
| T7C62mut6G7C | 5'-AGCAACCGCACCTGTGGTAATACGACTCACAATCGGCGAGTTGTGAGCGGATAACAATTCCCC-3' |

### 2.2.3 Other nucleic acids

| | |
|---|---|
| Hering sperm DNA as carrier DNA | Roche |
| Poly A RNA as carrier RNA | Amersham |

## 2.3 Antibodies

| | |
|---|---|
| Anti- GFP rabbit IgG fraction, polyclonal | Invitrogen |
| Alexa Fluor 488 goat anti-rabbit IgG | Invitrogen |
| Polyclonal Anti-Rabbit Imunoglobulin/HRP | Dako |

## 2.4 Enzymes

### 2.4.1 Restriction endonuclease

All the restriction endonucleases with their respective buffer systems are purchased either from New England Biolabs or MBI(Fermantas)

### 2.4.2 Other Enzymes

| | |
|---|---|
| Dream Taq Polymerase | Fermentas |
| Faststart taq Polymerase | Roche |
| RNAse IF | New England Biolab |

## 2.5 Chemicals and reagents

| | | | |
|---|---|---|---|
| Acrylamide solution | AppliChem | HCL | J. T. Baker |
| Agar | AppliChem | Hexane | J.T.Baker |
| Agarose | USB | Isopropanol | Merck |
| Ammoniumpersulfate(APS) | Roth | Methanol | J.T.Baker |
| Ampicillin | Roth | PBS (10X) | Gibco |
| Bromophenol Blue | Gibco | SOC medium | Invitrogen |
| BSA | New England Biolab | Sodium acetate | J.T.Baker |
| DMSO | Applichem | Sodium chloride | J.T. Baker |
| dNTPs | Merck | β-mercapto-ethanol | Sigma |
| EDTA | Sigma | Sterile water | Braun |
| Ethanol | Riedel-deHaen | TEMED | Sigma |
| Ethidium bromide | Roth | Triton X 100 | AppliChem |
| Fat Free dry milk | Spinnard | Tween 20 | AppliChem |
| Sodium Hydroxide | AppliChem | Tris | Sigma |
| | | Yeast extract | AppliChem |

### 2.5.1 Special chemicals and reagents for Multi-copy bead display reagents

| | |
|---|---|
| Myone carboxylic acid Dyna Beads | Invitrogen |
| M270 carboxylic acid Dyna beads | Invitrogen |

| | |
|---|---|
| 2-(N-morpholino)ethanesulfonic acid (MES) | AppliChem |
| Abil EM 90 | Evonik |
| Span-80 | Sigma |
| Mineral oil | Sigma |
| TritonX 100 | Sigma |
| 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide | Thermo scientific |
| SNAP-Vista green | New England Biolabs |
| BG-$NH_2$ | New England Biolabs |
| BG-malemide | New England Biolabs |
| BG-agarose beads | New England Biolabs |

#### 2.5.1.1 Cell free expression systems

| | |
|---|---|
| Eco Pro T7 system | Novagen |
| Expressway Cell- free S 30 expression kit | Invitrogen |
| RTS 100 E.coli HY kit | Roche |
| TNT Quick for PCR DNA | Promega |

## 2.6 Kit reagents

| | |
|---|---|
| Ampliscribe High Yield T7 transcription kit | Epicenter |
| Bright Glo Luciferase substrate | Promega |
| Clone JET PCR Cloning Kit | Fermentas |
| ECL ChemiGlow | Alpha Innotech |
| Geneclean III | Q- Biogene |
| Jetquick DNA Clean-up Spin Kit | Genomed |
| Jetstar Plasmid Maxi Kit | Genomed |
| QIAMP Blood DNA kit | Qiagen |
| QiaQuick Gel Extraction Kit | Qiagen |
| Qiagen® Plasmid Purification Kit | Qiagen |
| Quantitect SYBR Green PCR kit | Qiagen |

| | |
|---|---|
| Quantitect SYBR Green RT-PCR kit | Qiagen |
| TaKaRa DNA Ligation Kit | TAKARA |
| TOPO-TA Cloning Kit | Invitrogen |
| Turbo DNA-*free*™ | Ambion |

## 2.7 Media, Buffers and solutions

## 2.7.1 Media for bacteria culture

| | |
|---|---|
| LB Medium | 1 % (w/v) Trypton, 0.5 % (w/v) yeast extract, 170 mM NaCl, pH 7.2 -7.5 |
| LB Agar | 1.5 % (w/v) Agar in LB-Medium |

Both the LB media and LB agar are autoclaved and antibiotics either Amplicillin or Kanamycin are added to them 50-100 μg/ml, 25-50 μg/ml, respectively before use.

| | |
|---|---|
| S.O.C medium(Invitrogen) | 2 % (w/v) Trypton, 0.5 % (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM Glucose |

## 2.7.2 Preparation of Buffers and solution for nucleic acid biochemistry

### 2. 7. 2. 1 Reagents for DNA/RNA precipitation/extraction

| | |
|---|---|
| 3 M Sodium Acetate | 3 M NaOAc.3$H_20$ in $H_20$, pH=5.2 with 3 M acetic acid. |
| 4 M Lithium Chloride | 4 M LiCL in $H_2O$ |

5 M Ammonium acetate (AppliChem)

| | |
|---|---|
| Phenol:Chloroform:Isoamyl alcohol (Sigma) | 25:24:1 saturated with 10mM Tris,1mM EDTA |
| Chloroform: Isoamyl alcohol (Sigma) | 24:1 |

### 2.7.2.1 Solutions for DNA gel electrophoresis

| | |
|---|---|
| TAE-Buffer (50x) (AppliChem) | 2.0 M Tris<br>0.05 M $Na_2EDTA$<br>1 M Acetic acid |
| TBE-Buffer (10x) (Applichem) | 0.89 M Tris<br>0.2 M $Na_2EDTA$<br>0.89 M Boric acid |

These buffers were used in 1X concentration during electrophoresis.

| | |
|---|---|
| DNA loading dye (6x) (Fermentas) | 0.03 % (w/v) Bromophenolblau<br>0.03 % (w/v) xylene cyanol FF<br>60 % (v/v) Glycerol<br>60mM EDTA |
| Ethidium Bromide Stock solution | 10mg/ml of ethidium bromide in $H_2O$ |

### 2.7.3 Buffers and solutions for Protein biochemistry

### 2.7.3.1 Solutions for SDS-PAGE and Western Blot

| | |
|---|---|
| 1x SDS-PAGE loading buffer | 50mM Tris-HCl (pH 6.8)<br>2% SDS<br>0.1% bromophenol blue<br>10% glycerol<br>100mM dithiothreitol |

| | |
|---|---|
| Running Buffer (10x) | 0.25 M Tris<br>1.9 M Glycin<br>1 % SDS |
| Coomaie staining solution | 0.1 % (w/v) Coomassie Bril. Blue R250<br>50 % (v/v) Methanol<br>10 % (v/v) Acetic acid<br>40% distill water |
| Destaining solution | 50 % (v/v) Methanol<br>10 % (v/v) Acetic acid<br>40% distill water |
| Western Blot transfer Buffer | 0.02 M Tris<br>0.15 M Glycin<br>20 % (v/v) Methanol |
| Western Blot washing buffer | 0.1 % (v/v) Tween 20 in PBS (1x)<br>(PBS-T) |
| Western Blot Blocking solution | 5 % (w/v) non fat milk in PBS-T |

## 2.7.4 Buffers and solution for Multicopy Bead Display

| | |
|---|---|
| SSC Buffer (20x) | 3M Sodium chloride<br>300mM Trisodium citrate<br>pH7.5 titrated with HCl |

For hybridization experiments, the working concentration used is 1x.

| | |
|---|---|
| MES Buffer | 1M 2-(N-morpholino)ethanesulfonic acid)<br>in deionized H2O, pH=6.0 |

| | |
|---|---|
| | titrated with 0.1N NaOH |

The working concentration of this buffer to be used for coupling of oligonucleotides is 25mM.

| | |
|---|---|
| TE buffer | 1 mM EDTA, pH 8.0<br>10 mM TrisCl, pH 8.0<br>in $H_2O$ |
| Bind and wash Buffer<br>(for first emulsion breaking) | 10.0 mM Tris-HCl (pH 7.5)<br>1.0 mM EDTA<br>2.0 M NaCl |
| Breaking buffer<br>(for second emulsion breaking) | 1 mM DTT<br>10pM BG-$NH_2$<br>1x PBS |
| FACS Buffer | 0.5 % (w/v) BSA, 1 mM $NaN_3$in PBS |
| FACS washing buffer | 0.05 % (v/v) Tween 20 in PBS(1x) |

## 2.8 Standards

## 2.8.1 DNA size Standards

DNA 1kb Ladder Fermentas

*(size in bp: 75, 200, 300, 400, 500, 700, 1000, 1500, 2000, 3000, 4000, 5000, 7000,10000, 20000)*

DNA 100bp Ladder Invitrogen

*(size in bp: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2072)*

## 2.8.2 Protein molecular weight marker

Prestained Precision Protein Standard Biorad

*(size in kDa: 10, 15, 20, 25, 37, 50, 75, 100, 150, 250)*

## 2.7 Device instruments

| | |
|---|---|
| Balance SPB 63 | SCALTEC |
| Benchtop Centrifuge- Small | 415R, Eppendorf |
| Benchtop Centrifuge- Large | Hettich-Rotina 420R |
| Biophotometer | Eppendorf |
| Cell counting chamber | Beckman |
| Deep Freezer -80 ºC | Sanyo |
| Electrophoresis and botting apparatus | Bio-Rad |
| FACS calibur | Bectin Dickenson |
| FACS Di Va cell sorter | Bectin Dickenson |
| Flourescence microscope Axiovert 100 | Zeiss |
| Gel chamber | Peqlab |
| Light microscope | Nikon/Leica |
| Light cycler | Roche Diagnostics |
| Luminometer | Hamamatsu Photonics |
| Magnetic stirrer IKA-combimag RET | Janke & Kunkel |
| Microplate Luminometer | Berthold detection Systems |
| Microwave R- 220A | Sharp |
| PCR Cycler PTC-100 | MJ Research Corporation |
| PCR Gradient Cycler PTC-200 | MJ Research Corporation |
| pH Meter | Hanna instruments |
| Pipette filler | Hirschmann Laborgereate |
| Pipettes | |
| - 2 µl - 20 µl | |
| - 20 µl - 200 µl | ABIMED |
| - 100 µl - 1000 µl | |
| Piston pipettes | |
| - 10 µl - 50 µl | Gilson |
| - 100 µl - 1000 µl | |
| Precision balance | SBA 31 SCALTEC |
| Refrigerators | AEG, Siemens, Sanyo, Liebherr |
| Shaker Rokomat | Technomara |
| Speedvac | Uniequip |
| Thermcycler | MJ Research |
| Thermomixer | Eppendorf |
| Vortex Genie 2 | Scientific Industries |
| Water Distillation Apparatus | Pro 90 CN Seral |
| Water bath | Memmert |
| Western-Blot | Bio-Rad |
| Vacuum pump | Uniequip |

### 2.7.1 Device for Emulsion Preparation

| | |
|---|---|
| Flat Bottom Cryovial | Greiner Bio-one |
| Digital Magnetic Stirrer | IKA |
| Spin Bar Magnets | |

# 3. Methods

## 3.1 General Molecular Biology

### 3.1.1 Purification of DNA by phenol / chloroform extraction

Equal volumes of phenol / chloroform / isoamylalcohol solution (25:24:1) (v:v:v) were added to the solution containing the DNA and mixed by vortexing until an emulsion was formed. After a short centrifugation, the upper aqueous phase containing the DNA was transferred into a clean 1.5 ml reaction tube. Then equal volume of chloroform / isoamylalcohol solution (24:1) (v:v) was added to the recovered aqueous phase and mixed by vortexing for 15 sec. The mixture was briefly centrifuged for separation of phases and the upper aqueous phase containing DNA was transferred to a clean 1.5 ml reaction tube. This process was repeated for one more time and the DNA was recovered by ethanol precipitation.

### 3.1.2 Purification of DNA by ethanol precipitation

Three molar sodium acetate (pH=5.2) was added to the DNA solution at 1:10 ratio and mixed by vortexing followed by addition of 3 volumes of ice-cold Ethanol (99% v/v) to the solution. The mixture was shortly vortexed and then centrifuged at 4°C for 30 min at 14000 rpm using a bench-top centrifuge. The pellet was washed with ice-cold 70% ethanol following by another round of centrifugation for 10 min. The supernatant was carefully discarded and the pellet was air-dried for 5 to 10 min, then resuspended into an appropriate volume of nuclease-free water and stored at –20°C.

### 3.1.3 Analytical isolation of plasmid DNA

Bacteria colonies (10 – 20) were picked from the agar plates and each was seeded into 5 ml of LB medium containing the appropriate antibiotic and incubated for 16 to 20 h under gentle shaking at 37 °C. 2 ml of the bacterial suspension was thereafter collected and cells were pelleted by centrifugation at 4000 rpm for 5 min. The plasmid DNA was then extracted using the Qiagen miniprep kit (Qiagen). The DNA was eluted in an appropriate volume of TE buffer or water and subjected to enzymatic restriction or kept at –20 °C for further experiments.

### 3.1.4 Preparative isolation of plasmid DNA

One mililiter of bacterial suspension from the minipreps (found positive after analytic isolation and restriction of plasmid DNA) was seeded into 100 to 250 ml of LB medium containing the appropriate antibiotic, and incubated for 16 to 20 h under gentle shaking at 37 °C. Bacterial cells were thereafter pelleted by centrifugation and the DNA was extracted using the Qiagen Plasmid Maxi Kit (Qiagen). The DNA was eluted with TE buffer or water, quantified in photometer and stored at –20 °C till further use.

### 3.1.5 Preparative isolation of plasmid DNA from agarose gel

Following electrophoresis, the gel was observed under a UV lamp with wavelength of 366 nm and the fragment of interest was cut from the gel using a scalpel blade. The DNA was then recovered from the gel using a commercial kit (QIAquick Gel Extraction Kit, Qiagen) according to the manufacturer's manual.

### 3.1.6 Determination of DNA concentration

The DNA concentration was determined as described by Sambrook and Russell, 2001. The optical density (O.D) of the DNA solution was measured at wavelength of 260 and 280 nm. The formula used for the determination of the DNA concentration was: 1 O.D. = 50 μg/ml of double stranded DNA. The ratio

between the optical density at 260 nm and 280 nm ($OD_{260}/OD_{280}$) of a pure DNA ranges between 1.8 and 2.0.

### 3.1.7 Digestion of plasmids with restriction endonucleases

Endonuclease digestion of DNA was carried out according to enzyme supplier's manual. 0.5 to 1 μg of DNA was digested in case of analytical isolations whereas up to 5 μg of DNA was digested overnight for a preparative DNA digestion.

### 3.1.8 Dephosphorylation of 5' ends of DNA

To avoid self ligation of plasmid vector, phosphate groups at the 5' ends of the vector were removed as follows: 2 μg of DNA was first digested with the appropriate restriction enzyme in a total volume of 20 μl. The DNA was thereafter heat-inactivated for 10 min at the appropriate temperature depending on the inactivation requirement of a particular enzyme. 5 μl of 10X phosphatase buffer, 1μl (1 unit) of calf intestinal alkaline phosphatase and 24 μl of $H_2O$ were added to the DNA solution. The mixture was incubated at 37°C for one hour, then heat inactivated at 70°C for 10 min.

### 3.1.9 Separation of DNA by agarose gel electrophoresis

Agarose was boiled in either TAE or TBE buffer at concentrations of 0.8 to 2% (w/v) according to the experiment. Then ethidium bromide was added to the gel solution at the final concentration of 1μg/ml. The gel was thereafter casted into an electrophoresis chamber containing an appropriate comb. The DNA was mixed with the DNA loading buffer (p.35, 2.7.3.1) and loaded on the gel together with a DNA molecular weight marker. DNA was then separated for 30 to 120 min at 100 to 150 Volts. The bands could be visualized under an ultra violet lamp.

### 3.1.10 DNA ligation

The DNA ligation was performed using Takara DNA ligation kit (Clonetech). After purification (p.42, 3.1.9), plasmid vectors and inserts were mixed at a molar ratio of approximately 1:4 and the ligation reaction was performed using a commercial kit (Takara DNA ligation kit) as recommended by the manufacturer.

### 3.1.11 Transformation of chemically competent bacteria

An aliquot of chemically competent bacteria (p. 28, 2.1) was thawed on ice. 50 to 200 ng of plasmid DNA, or 1 to 2 µl of ligation products were added to cells, and incubated on ice for 20 min. The cells were then heat-shock treated at 42 °C for 45 sec, and then snap chilled on ice for 2 min. 200 µl of S.O.C. medium (Invitrogen) were added to the bacteria which were thereafter incubated for one hour at 37 °C with gentle shaking. Depending on the experiment, the cells were plated diluted or undiluted on agar plates containing the appropriate antibiotic, and then incubated 12 to 16 h at 37 °C.

### 3.1.12 Polymerase chain reaction (PCR)

The general PCR reactions were set up as follows:

- 5 µl of 10 x PCR buffer
- 5 µl of 10 µM forward or reverse primers
- 1 µl of 10 µM dNTPs
- 0.25 µl of Taq polymerase (5U/ µl)
- 1 µl of DNA template (plasmid or genomic DNA at different concentrations)
- $H_2O$ up to a final volume of 50 µl

Samples were subjected to a first denaturation step at 94 °C, followed by 30 to 40 amplification cycles consisting of annealing (55 to 65 °C), extension (72 °C) and and a final extention step at 72 °C. The incubation time for each step depended on the size of the expected PCR product.

## 3.1.13 Analysis of proteins

### 3.1.13.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated according to their molecular weight in a SDS-PAGE. The gel is discontinous system containing stacking gel and resolving gel. The percentage acrylamide in a resolving gel ranges from 5 to 10 %. Whereas in stacking gels it is alwaays 5%. The composition of a 10 % SDS-PAGE gel is shown in the table 3. 1.

**Table 3.1** Composition of a 10 % SDS-PAGE

| | **Resolving gel** | **Stacking gel** |
|---|---|---|
| 2 M Tris-HCL, pH 8.8 | 6.0 ml | - |
| 0.5M Tris-HCl, pH 6.8 | - | 2.0 ml |
| 30% Acryl-bisacrylamide | 4.0 ml | 560 µl |
| $H_2O$ | 1.8 ml | 1.4 ml |
| 10% SDS | 120 µl | 60 µl |
| 10% APS | 120 µl | 60 µl |

The glass plates of the electrophoresis device were assembled according to the manufacturer's manual. The acrylamide resolving gel was poured into the 0.75 mm thick glass chamber leaving about 1.5 cm for the stacking gel. This solution was overlaid with 2-isopropanol until the complete polymerization of the gel. The 2-isopropanol solution was thereafter discarded and the gel washed with deionized water. The stacking gel was prepared immediately before use and poured onto the surface of the resolving gel. The combs were placed into the staking gel until polymerization of the gel. Protein samples were mixed with SDS-PAGE loading buffer (p.36, 2.7.3.1) and denatured at 80 °C for 5 min, and were loaded into the wells of the gel together with a protein size marker. This was thereafter separated at a voltage of 100 V for 1 to 2 h until the bromophenol blue reached the bottom of

the resolving gel. The gel was then either stained with coomasie stain for protein visualization or directly used for western blotting.

To visualize the protein directly on the gel, the gel was first stained with coomasie for 15- 30 min and then detained with destaining solution. As the protein bands starts apearing, destaining is stopped. The gel is then dried and documented.

#### 3.1.13.2 Western Blot

For western blotting, the proteins were first seperated by SDS-PAGE as described in 3.1.13.1. The seperated proteins were then transfered to nitrocellulose membrane for western blotting. To transfer the proteins, the nitrocellulose membrane was first equilibrated for 15 sec in pure methanol, then 2 min in deionized water and finally equilibrated into the transfer buffer for 15 min. The blot containing blotting paper, nitrocellulose membrane, gel, sponges were assembled according to the manufacturer's instructions. This was thereafter placed into a western blot transfer device and containing transfer buffer (p.36, 2.7.3.1). The transfer of the protein from the gel to the nitrocellulose membrane was performed for 1- 2 h at 100 V. The membrane was thereafter incubated at 4°C overnight in WB blocking buffer (p.36, 2.7.3.1) with gentle shaking. It was then incubated for two hours with primary antibodies and then washed three times with PBST to eliminate any trace of unbound antibodies. It was then incubated for one hour with horseradish peroxidase-conjugated secondary antibodies. The membrane was again washed four times with PBST and probed with chemiluminescent substrate from ECL Chemiglow kit (Biozym) according to the manufacturer's manual. The chemiluminescence was then observed in a luminometer (Hamamatsu) and documented.

## 3.2 Construction and characterization of model templates

The expression cassette GFP SNAP interlinked with a 4 GS linker was cloned in the multiple cloning site of pET 21 vector between Nco I and Cla I restriction sites and was present in the lab before. The construct y contained the T7 promoter

followed by a ribosome binding site (RBS), the GFP-SNAP fusion protein and a T7 terminator. This expression cassette was then amplified by PCR with 15bs2T7 and 15bs2T7 Term A (sequences shown in table) primers to obtain linear templates required for characterizations.

Another control fusion protein MS2-SNAP having the similar expression modules as GFP-SNAP was also constructed. The primer binding sites 15bs2T7 and 15s2T7termA were introduced in this construct through cloning to be able to amplify MS2-SNAP expression with the same set of primers as for GFP-SNAP expression cassette.

### 3.2.1 Western Blot analysis of GFP-SNAP fusion protein

GFP-SNAP fusion protein was expressed in a cell free expression system using the *in vitro* coupled transcription-translation kit TNT T7 Quick for PCR DNA (Promega) according to the manufacturer's guidelines. 800 ng of purified PCR product were used for cell free expression of proteins. 5 µl of translated reaction mixture was ran on 10% SDS PAGE and transferred to a nitrocellulose membrane as described above (p.45, 3.1.13.2.). The blot was blocked overnight at 4 °C blocking buffer, then incubated for 2 h with a 1:2000 diluted anti-GFP monoclonal antibody. After washing, the blot was incubated for 1 h with 1:3000 diluted anti-mouse IgG secondary antibody conjugated to horseradish peroxidase. After a serial of three washings, bands were revealed by adding ECL (Biozym) substrate.

### 3.2.2 BG-Binding assay of GFP-SNAP

GFP-SNAP fusion protein was expressed in a cell free expression system using the *in vitro* coupled transcription-translation kit TNT T7 Quick for PCR DNA (Promega) according to the manufacturer's guidelines. 800 ng of purified PCR product were used for cell free expression of proteins. 10 µl of the translation reaction were incubated with 10 µl of BG-agarose beads (Covalys) for 25 min at

room temperature in the dark. All beads were then analyzed under a fluorescence microscope.

#### 3.2.2.1 Competition Assay

GFP-SNAP fusion protein was expressed in a cell free expression system using the *in vitro* coupled transcription-translation kit TNT T7 Quick for PCR DNA (Promega) according to the manufacturer's guidelines. 800 ng of purified PCR product were used for cell free expression of proteins. 10 µl of the translation reaction were first incubated with 10 µl BG-$NH_2$ (2 mg/ml) (New England Biolabs) for 25 min at room temperature in the dark. This reaction mixture was then incubated with 10 µl of BG-agarose beads for 25 min at room temperature in the dark. All beads were then analyzed under a fluorescence microscope.

## 3. 3 Covalent coupling of oligonucleotides to microbeads

The 5'-amino-modified forward primers 15bs2T7-s2+TG spacer were coupled to magnetic beads (Dynabeads Myone carboxylic acid, Invitrogen) according to the manufacturer's guidelines and a previously published report [110] with minor modifications. 7-12×$10^8$ beads (100 µl of the bead solution) were transferred into 1.5 ml eppendorf tubes and washed three times with 100 µl of 0.01 N NaOH and three times with 100 µl deionized water. Beads were then resuspended in 50 µl 25 mM MES (2-[N-morpholino] ethane sulfonic acid, pH 6.0) buffer. 20 µl of a 100 µM 5'-amino modified primer solution was added to this solution and incubated on an end-over-end rotator for 30 min. After 30 min 3 mg of EDC (3-dimethylaminopropyl carbodiimide hydrochloride) dissolved in 30 µl of 25 mM MES buffer (pH 6.0) were added and incubated overnight on a end-over-end rotator at 4 °C. Beads were then washed 5 times with 100 µl TE (pH 8.0) and stored in 100 µl of TE buffer at 4 °C.

### 3.3.1 Verification of primer coupling on microbeads through probe hybridization.

10 µl of 15bs2T7+ TG spacer coupled beads were washed three times with deionized water. The beads were then suspended in 80 µl of nuclease free water. 10 µl of 1 µM 6-FAM labeled anti-15bs2T7 was added, followed by 10 µl of 10x PCR buffer to increase the final volume to 100 µl. The reaction mixture was then incubated at 50 ºC for 15 min in thermal cycler. The beads were then washed three times in 1x SSC buffer and resuspended in 500 µl of 1x PBS and analyzed in flow cytometry.

## 3. 4 Covalent coupling of oligonucleotide to BG

The primers 15bs2T7 Term A or anti15bs2 (BG) were chemically coupled to BG substrate. BG-malemide containing a thiol reactive group was purchased from New England Biolabs (Germany) and coupling was perfomed using a protocol previously published [117] through a custom service of the Sigma Aldrich company (Germany) who also sysnthesized the oligonucleotides to be coupled to BG substrate.

## 3. 5 Water in oil emulsion PCR

The aqueous phase is composed of 100 µl PCR mix containing the following reagents: $3\times10^9$ copies of linear template DNA, $7\text{-}12\times10^7$ beads coupled with forward primer (T7-s2+TGSpacer), 0.01 µM of soluble forward primer (T7-s2), 3 µM of BG coupled reverse primer, 0.5 mM of dNTPs, 15 U of DreamTaq DNA polymerase (Fermentas) and 10 µl of BSA (10 mg/ml). The mineral oil mixture containing 2 % Abil-90 EM and 0.05 % Triton X-100 was prepared as published previously [138]. In a 2 ml cryovial with flat bottom the aqueous phase was gradually added drop wise to 500 µl of the oil phase over a period of 5 min while stirring constantly on ice at 2000 rpm. Stirring was continued for another 3 min to obtain a homogenous emulsion. The emulsion was then dispensed in 50 µl aliquots

into ten 200 µl thin-walled PCR tubes. PCR was then performed in a conventional thermocycler with the following temperature profile: 95 °C for 3 min, 45 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 90 s, respectively followed by a final extension at 72 °C for 10 min. After PCR, all samples were collected in a 1.5 ml eppendorf tube and centrifuged at 13,000 g for 5 min. The oil was removed from the top and intact emulsion droplets remain concentrated at the bottom. The emulsion was disrupted by adding 100 µl of Bind and Wash buffer (p.37,2.7.4), 1 ml of hexane and vortexing for 10 seconds. The disrupted oil phase (top) in hexane was removed and discarded and beads settled down in the aqueous layer. To completely remove the oil from the bead suspension, hexane extraction was repeated three more times. Residual hexane was removed by centrifugation under vacuum at room temperature for 5 min. The beads were then washed 5 times with TE pH 8.0 and resuspended in 9 µl of nuclease free water.

### 3.5.1 Determination of size distribution of picoliter reactors

The emulsions were prepared as descibed in [110] . They were then observed under the microscope at 40 X magnification and an image was taken.This image was then processed for size analysis. The size analysis was performed using Image J, a public domain software package developed at National Institutes of Health (NIH), USA for image processing. Dune Sciences, Inc (USA) developed the protocols to process the images and further calculation of the particle sizes and size distribution. Both the images used for processing are shown in the figure 3.1 and the final graph depicting the size distribution is shown in the *results chapter* (p64, figure 4.3). This analysis was kindly done by Dr. DurgamMadhab Mishra of Physics Department, Ruhr University Bochum, Germany.

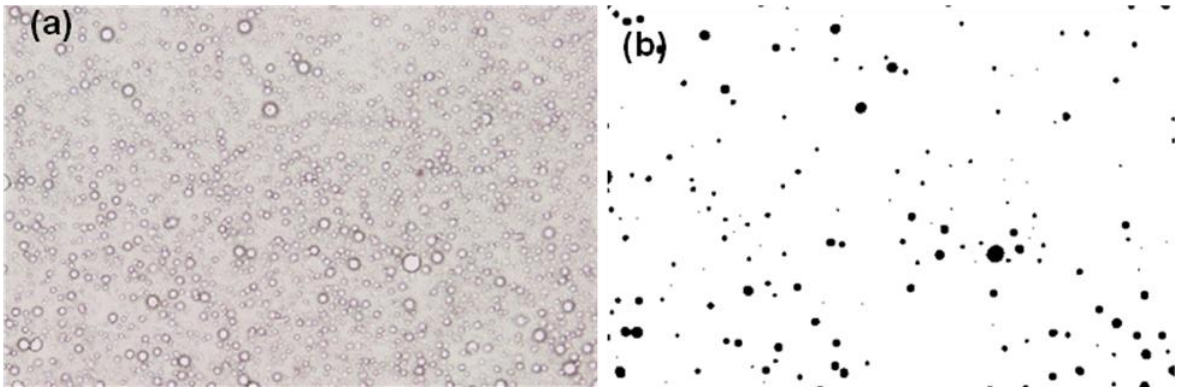



**Figure 3.1 Processing of micrscopic image for size analysis of droplets.** (a) The microscope image of the oil emulsion, which was used for the image processing. The background subtraction yields the image showing the particles only in black and the rest in white as shown in Figure (b). The dark and bright contrast determines the threshold to distinguish particles from the background. The black circular particles can be seen in the white background. These black particles were fitted with elliptical areas, which yield an average area A. The mean diameter can be calculated from the equation $d = A/\pi$.

## 3.5.2 SYBR green staining of PCR products on beads

The microbeads containing PCR products were stained with SYBR Green. Beads were resuspended in 90 µl of nuclease free water. Then 10 µl of 1:10000 diluted SYBR Green solution was added. The reaction mixture was then incubated at room temperature for 30 min in dark. The beads were then washed three times with 1x SSC buffer, resuspended in 500 µl of 1x PBS and analyzed in Flow cytometry. The process of SYBR Green staining of beads is shown in the figure 3.2.

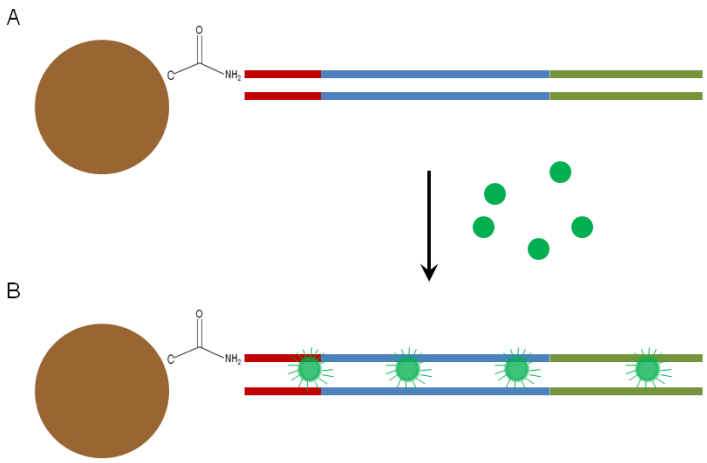


**Figure 3.2 *(contnued)* Outline of SYBR Green Staining of PCR products on beads**. (A) PCR products on beads are incubated with SYBR Green. (B) SYBR Green intercalates into double stranded DNA.

## 3.5.3 Probe hybridization to PCR products on microbeads.

The microbeads containing PCR products were hybridized with a fluorescently labeled oligonucleotide probe. Beads from emulsion PCR were first re suspended in 0.1 M NaOH and incubated for 2 min. The beads are put in a magnetic separator and supernatant is removed. These beads were then washed three times with 1x SSC buffer. The probe hybridization reaction is set up with 6 FAM labeled anti-pAGTSM oligonucleotide (which is complementary to a portion of SNAP) as described above (p.48, 3.3.1). The temperature profile used for hybridization was ramped from 72 °C to 50 °C in 20 min, 50 °C for 10 min. The beads were then washed three times in 1x SSC buffer, resuspended in 500 µl of 1x PBS and analyzed by flow cytometry.This process is also illustrated in the figure 3.3.

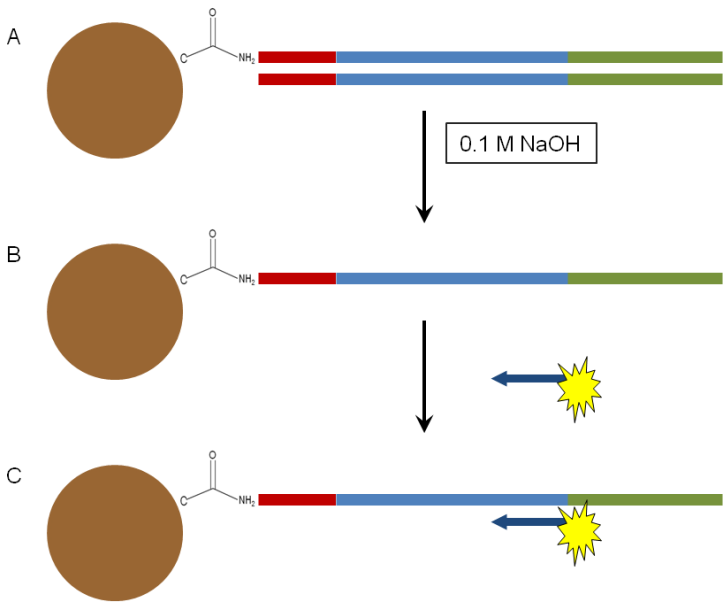


**Figure 3.3 *(continued)* Outline of oligohybridization of PCR products on beads.** (A) PCR products on beads treated with 0.1 M NaOH to melt the double stranded DNA. (B) Melted strand is removed and beads with single strand then hybridize with oligo coupled with a fluorophore which is complementary to the DNA strand on the bead. (C) The final hybridized product on the beads can be analyzed through flow cytometry**.**

## 3.6 Normalization of BG binding sites on beads with BG-oligo

BG moeities on microbeads were increased through simply hybridizing a BG-labeled oligonucleotide (anti- 15bs2-BG) to the capture oligonucleotide (15bs2 T7+ TG spacer) of the bead after the emulsion PCR. This was done as described in 3.3.1. This process is also illustrated in the figure 3.4.

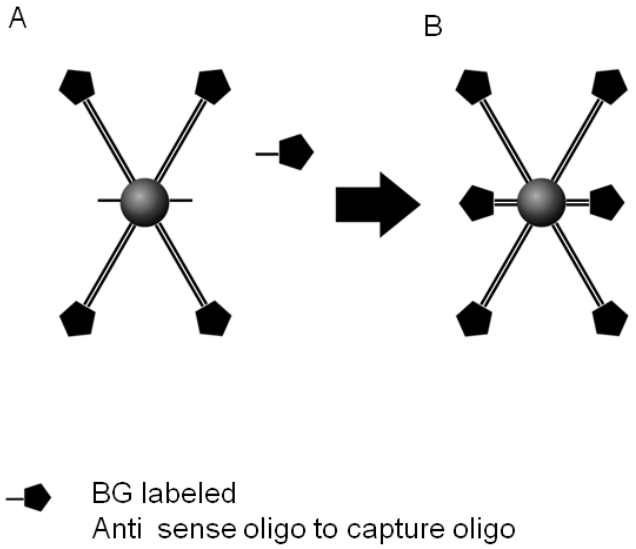



**Figure 3.4 Normalization of Bg binding sites on Beads. After emulsion PCR,** (A) the beads were extracted and they contain some of the unused capture oligo on them. (B) A BG- labeled oligonucleotide anti sense to the capture oligo is allowed hybridized with the capture oligonucleotides on the beads to increase BG moeities on the beads.

## 3.7 Cell free expression of proteins in water in oil emulsion

The TNT T7 Quick for PCR DNA system (Promega) was used to perform *in vitro* transcription–translation reaction in emulsions. 50 µl of reaction mix were prepared on ice by combining 40 µl of TNT T7 Quick for PCR DNA mix supplemented with 1 µl of methionine (1 mM) and 9 µl of beads (extracted from emulsion after emPCR) suspended in nuclease free water. The reaction mixture was added to the 500 µl oil mixture as described above for the emulsion PCR to form the emulsion

IVTT. The emulsion was then incubated at 30°C for 5 -120 min for the expression of proteins. The emulsion was put on ice for 5 min and breaking of emulsion was done as previously described [139]: 100 µl of breaking buffer (p.37, 2.7.4)) was also added during the recovery of beads. The breaking buffer contains DTT and BG-$NH_2$.The use of DTT is for protein stability and BG-$NH_2$ was added to bind unbound proteins expressed in picoliter reactors and so avoid binding of these proteins to beads from other reactors. The beads were then washed twice with breaking buffer and resuspended in 100 µl of FACS buffer.

## 3.8 Antibody staining of microbeads

100 µl of a 1:100 dilution of anti-GFP rabbit serum (Invitrogen) in FACS buffer were added to the beads, mixed well and incubated at room temperature for 1 h with intermittent mixing every 20 min. The beads were then washed twice with 100 µl of FACS buffer and stained with 100 µl of 1:100 diluted anti rabbit IgG coupled to Alexa 488 (Invitrogen) for 1 h in the dark at room temperature with intermittent mixing every 20 min. The beads were then washed two times with 100 µl of FACS buffer, resuspended in 500 µl FACS buffer and stored in 4 ºC till analysis.

## 3.9 Analysis of beads through Flow Cytometry

The microbeads displaying either DNA or GFP-SNAP fusion protein and stained with appropriately with flourscent probe or a flourecently labelled antibody as described above in FACS calibur (Bectin Dickenson). The flourescence intensity of beads was analyzed using a standard protocol as described in the manufacturer's manual. A population of single beads was gated using forward and side scatter and analyzed. The acquired data were then analyzed through FCS express Flow Cytometry Data Analysis (De Novo software).

## 3.10 Generation of beads containing GFP-SNAP and MS2-SNAP through competition

GFP-SNAP fusion cassette was amplified and captured on microbeads by emulsion PCR as described above (p.48, 3.5). After breaking of the emulsion, these beads are then resuspended in 9 µl of nuclease free water containing different concentrations ($1\times10^{10}$ - $1\times10^{12}$) of linear MS2-SNAP templates. These beads solution then underwentcell free expression of proteins in a second water in oil emulsion as described above (p.52, 3.7). Beads were then stained against GFP as described above (p.53, 3.8) and analyzed by flow cytometry as described above (p.53 3.9). Competition assay is also illustrated in the figure 3.5.

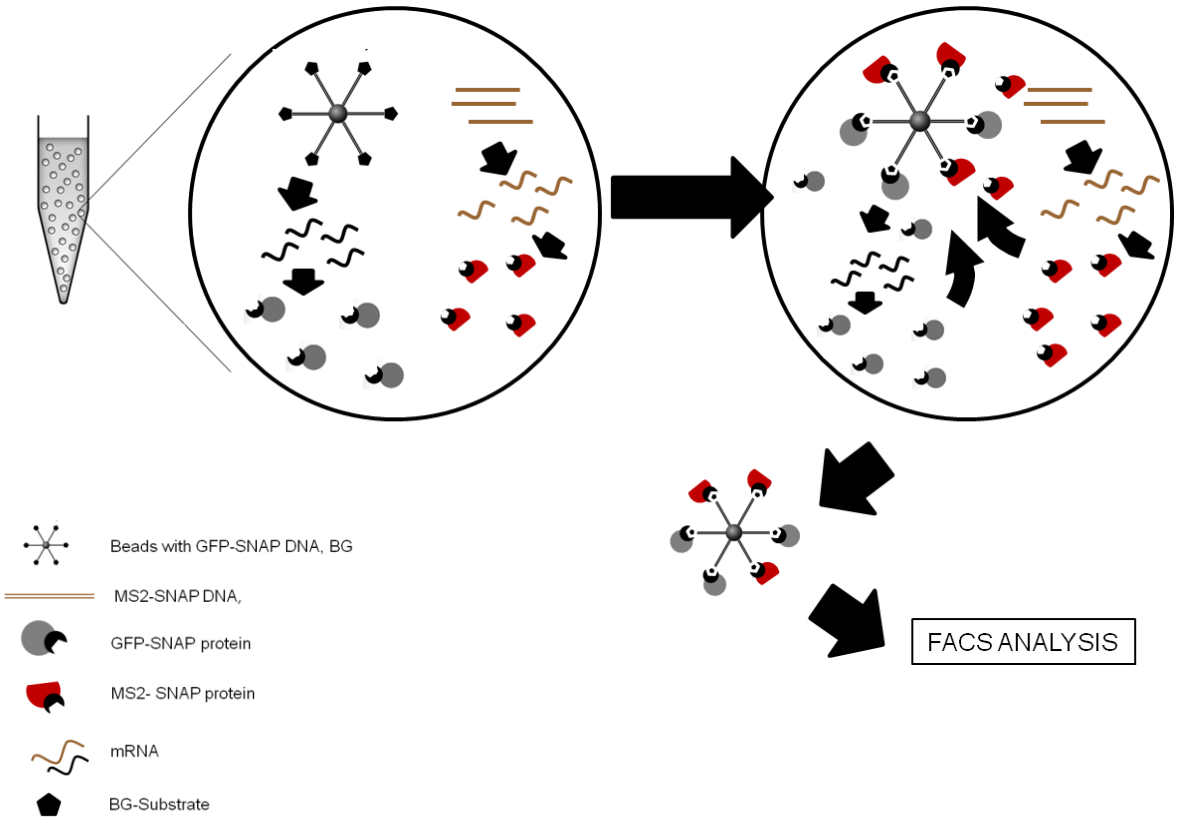



**Figure 3.5 Schematic representation of competition assay.** Microbeads containing GFP-SNAP expression cassette with BG and soluble MS2-SNAP expression cassette without BG are mixed and added to IVTT in emulsion. Both the DNA expresses encoded fusion proteins and these proteins compete for BG- binding sites on the beads. These beads are extracted from emulsion, probed with anti-GFP antibodies and analyzed by flow cytometry.

## 3.11 Construction of a T7 promoter library

A T7 promoter library was constructed by PCR amplification of a cloned open reading frame encoding a GFP-SNAP fusion protein with the partially randomized primer T7-wobble-s and the primer 15T7Term-a (sequences of all primers are shown in table 2.1 in *materials chapter*). T7-wobble-s starts with the sequence of the 15bT7-s2 primer followed by the T7 promoter sequence in which 10 nucleotides flanking the transcription initiation site completely randomized (Figure 4.1 B) followed by a spacer and primer binding site on the ORF of GFP-SNAP fusion protein. Thus, the PCR product is a library of randomized T7 promoter variants followed by the GFP and SNAP coding regions. The wild type T7 promoter construct was also produced using primers T7-consensus and T7Term-a, thereby maintaining the same reading frame as the library construct. PCR products were run on an agarose gel and eluted to completely remove the template plasmid DNA. The quality of randomization of the 10 nucleotides was confirmed by sequencing the PCR product.

## 3.12 Screening of beads displaying the T7 promoter library by flow cytometry

The microbeads displaying GFP-SNAP fusion protein under control of the T7 promoter library were first stained as described above (p.53, 3.8) and then sorted for by FACS DiVa cell sorter (Bectin Dickenson) using standard sorting protocol as described in manufacturer's manual with FACS DiVa software for analysis. Single bead population was gated and sorted usingforward and side scatter. Each round of sorting protocol involved two sorts. Beads were first yield sorted with a speed of 5000-10000 beads/s. After this yield sort, sorted beads underwent purity sort which was performed at a speed below 2000 beads/s. The purity sort beads were directly collected in 50 µl nuclease free water for subsequent PCR amplification.

## 3.13 Re-amplification of DNA from beads

A shorter fragment containing the T7 promoter variant separately as shown in the figure3.6 was amplified directly from beads using the primers 15bT7-s2 and T7pro_RP_seq2 with Faststart Taq polymerase (Roche) with a limited number of PCR cycles (25 cycles) with the following temperature profile: 94 ºC for 4 min, followed by 25 cycles of 94 ºC for 15 s, 62 ºC for 30 s, 72 ºC for 30 s and analyzed in a 1.5 % agarose gel. This DNA is either utilized for sequencing or building the entire expression cassette for next rounds of display.

## 3.14 Generation of expression cassettes by overlapping PCR

The longer fragment containing GFP and SNAP-tag as shown in figure 3.6 was amplified by PCR using the primers ovrlprim_NNB_T7 and 15bT7Term-a from a pEGFP-SNAP plasmid with Faststart Taq polymerase (Roche). The temperature profile for this PCR was 94 ºC for 3 min followed by 35 cycles of 94 ºC for 30 s, 60 ºC for 30 s, 72 ºC for 90 s and final extension at 72 ºC for 7 min. The entire cassette of GFP-SNAP for expression was rebuilt by using both - the smaller fragment as described above (p.55, 3.12) and the longer fragment in equimolar concentrations through overlap extension PCR using the primers15bT7-s2 and T7Term-a with Dreamtaq polymerase (Fermentas). This is also shown in the figure 3. 5. The temperature profile for this PCR was 94 ºC for 3 min followed by 35 cycles of 94 ºC for 30 s, 60 ºC for 30 s, 72 ºC for 90 s and a final extension at 72 ºC for 10 min. The resulting fragment was gel purified and used for another round of bead display as described above (p.55, 3.11) and then sorted as described in the *results chapter* ( p.85, 4.8.2). After the final round single beads were sorted directly in wells of a 96 well PCR plate, each containing 10µl of nuclease free water.

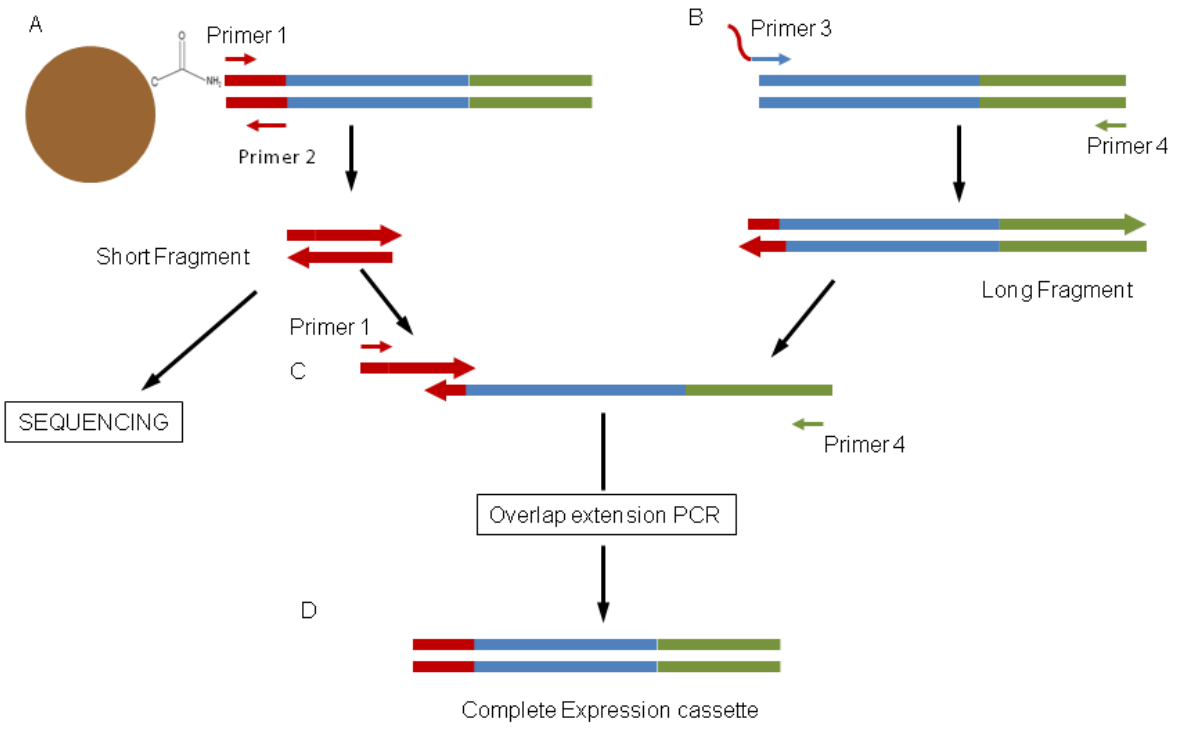



**Figure 3.6** ***(continued)*** **Generation of a complete expression cassette for Bead display**.(A) The shorter fragment containing the T7 promoter variant is amplified directly either from a population of beads or a single bead (a single bead contains multiple copies of PCR product in contrast to one as shown in the figure) with primer 1 (15T7-s2) and primer 2 (T7pro_RP_seq2).(B) The longer fragment containing the expression cassette GFP-SNAP without T7 promoter is amplified from the plasmid pGFP-SNAP with primer 3 (ovrlprim_NNB_T7) and primer 4 (15b T7Term-a). (C) Both short and long fragment are mixed in equimolar concentration for overlap extension PCR. (D) Generation of complete expression cassette with T7 promoter variant upstream of the sequence for GFP-SNAP fusion protein.

## 3.15 Sequencing of T7 promoter variants

15 µl of PCR product as described above (p.55, 3.12) were purified by ethanol precipitation as described in 3.1.1 and 10 µl were sequenced using the T7pro_RP_seq2 primer. The sequencing was done using the sequencing facility at Chemistry department, Ruhr University Bochum, Germany.

## 3.16 Cloning of T7 promoter variants

The positive clones of T7 promoter variants were cloned in a vector to be utilized for future purpose using the Clone JET PCR cloning kit (Fermantas) as described in the manufacturer's protocol.

## 3.17 IntroducingT7 promoter variants upstream of luciferase gene

The mutant T7 promoter sequences were PCR amplified using the primers 15bT7-s2 and Luc_ov_Rev to confer each of these variants with a segment of the luciferase gene by subsequent overlap extension PCR. The PCR was performed using Faststart Taq polymerase (Roche) with the following temperature profile: 94 ºC for 4 min followed by 25 cycles of 94 ºC for 15 s, 60 ºC for 30 s, 72 ºC for 30 s and a final extension at 72 ºC for 7 min. The luciferase gene without any promoter was amplified from pCR-Luc plasmid with the primers Luc_ov_Fwd and Luciferase_RP using Faststart Taq polymerase (Roche) with the following temperature profile: 94 ºC for 4 min followed by 25 cycles of 94 ºC for 15 s, 60 ºC for 30 s, 72 ºC for 90 s and a final extension at 72 ºC for 7 min. Both of these fragments were then mixed in equimolar concentration for overlap extension PCR (similar to process described in figure 3.5) to produce the entire expression cassettes for the luciferase gene under control of different T7 promoter variants. This PCR was performed using Dreamtaq Polymerase (Fermentas) with the following temperature profile: 94 ºC for 3 min followed by 35 cycles of 94 ºC for 30 s, 62 ºC for 30 s, 72 ºC for 90 s, respectively and a final extension at 72 ºC for 10 min. Products were gel purified and the concentration was determined photometrically.

## 3.18 Expression of T7 promoter variants in luciferase assay system

The luciferase gene under control of different promoter variants was expressed using the TNT T7 Quick for PCR DNA kit (Promega). One fifth of one reaction was used for analysis. 200 ng of DNA were mixed with 8 µl of master mix containing 1 mM of methionine in 0.5 ml tubes. Expression of proteins was done at 30 °C for 15-120 min. After expression, the reactions were snap chilled on ice for 5 min prior to depleting the reaction of RNA by adding 0.5 µl of RNase IF (New England Biolabs) and incubating the reaction at 37 °C for 30 min.

Expression levels of luciferase were analyzed in a luminometer using Bright Glo luciferase substrate (Promega) as described by the manufacturer.

## 3.19 Production of RNA for transcription assay

DNA containing the T7 promoter upstream of the luciferase gene was transcribed into RNA by the TNT T7 Quick for PCR DNA kit (Promega) at 30 °C for 15-120 min. The reaction was stopped by immediately adding lysis buffer from the QIAMP Blood DNA mini kit (Qiagen) and incubating at 56 °C for 10 min. This lysis buffer contains Proteinase K which should completely destroy RNA polymerase and thereby terminate the transcription of DNA. RNA was purified using the DNA Blood kit (Qiagen), which is known to co-purify RNA efficiently. The purified RNA was further treated with TURBO DNA-*free*™ (Ambion) to completely remove residual DNA.

### 3.19.1 Real time PCR analysis of RNA transcripts

The amount of RNA transcripts produced as described in 3.18 were then determined by a quantitative RT-PCR using primers Luc_FP1 and Luc_RP1. RT-PCR was performed using the Quantitect SYBRGreen RT-PCR Kit (Qiagen) in a total volume of 20 µl with 5µl (diluted 1:1000) purified RNA and each primer in a final concentration of 0.5 µM. All quantitative PCRs were performed on a Rotor-Gene 3000 (Corbett Research). The temperature profile was: 50 ºC for 20 min for reverse transcription, 94 ºC for 15 min, 40 cycles of 94 ºC for 15 s, 60 ºC for 30 s, 72 ºC for 30 s and acquiring SYBR Green flourescence at 81 ºC followed by standard melting curve analysis of products. Appropriate no RT and no template controls were used for each experiment. A standard curve derived from *in vitro* transcripts of known amount was used to calculate the copy numbers.

## 3.20 Production of RNA for translation assay

To determine the translational activity of the different T7 promoter sequences, DNA was transcribed to RNA using the Ampliscribe High Yeild T7 transcription

kit (Epicentre) as recommended by the manufacturer. The purified RNA was further treated with TURBO DNA-*free*™ (Ambion) to completely remove residual DNA. The RNA transcripts were quantified photometrically.

### 3.20.1 Translation assay through luciferase assay

2 ng of purified RNA transcripts as described in 3.19 were then added to the TNT T7 Quick for PCR DNA kit (Promega). Expression of luciferase was quantified as described in 3.17.

# 4. Results

## 4.1 Templates used for the development of Multi-copy Beads Display

In order to establish the multi-copy bead display approach, the template DNA was constructed with T7 promoter upstream of GFP-SNAP fusion protein in pET vector. A control construct containing a T7 promoter followed by bacteriophage coat protein MS2 and SNAP fusion protein was also generated (figure 4.1 A). The primers applicable for the amplification of GFP-SNAP were also introduced in the MS2-SNAP construct to be used in bead display later. Linear DNA templates from both of these constructs were constructed with 15bs2 and 15T7-Terma primers to be used for establishing this approach.

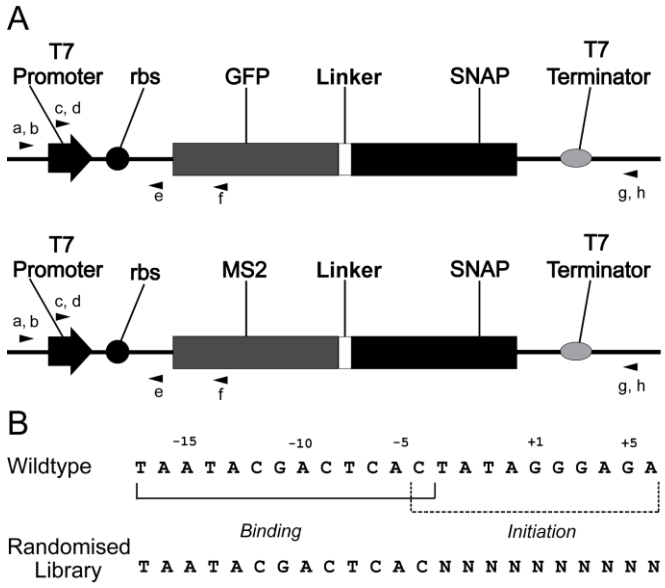



**Figure 4.1 Templates for multi-copy bead display.** (A) Maps of the different expression cassettes. The open reading frames for GFP or the coat protein of the bacteriophage MS2 are fused in frame via a flexible linker region to the open reading frame of SNAP. The cassettes are flanked by the T7 promoter, a ribosome binding site (rbs) and a T7 terminator sequence. The binding sites of primers used (Table 1) are also indicated (a-h). (B) Sequence of the wild type and the randomized T7 promoter. The polymerase binding site and the transcription initiation site are indicated.

## 4.2 Creation of stable emulsions

To optimize the number of pico liter reactors in an emulsion based system, four different oil mixtures were tested as listed in table 1. Water in oil emulsions were then prepared with 100 µl of aqueous phase in 900 µl of oil mixture as described in the methods section. Each of these emulsions were then analyzed for stability, and homogenicity of emulsion formation.

**Table 4.1** Oil Compositions

| | | |
|---|---|---|
| 1. | Mineral Oil, 4.5% Span-80, 0.5% Tween-80 | Tawfik and Griffith[24] |
| 2. | Mineral Oil, 4.5% Span-80, 0.5 % Tween-80, 0.1 % TritonX-100 | Ghadessy *et al* [120] |
| 3. | Mineral Oil, 4.5 % Span-80, 0.5 % TritonX-100 | Griffith *et al* [113] |
| 4. | Mineral oil, 2% Abil EM-90, 0.05% TritonX-100 | Williams *et al* [138] |
| 5. | Mineral oil, 7% Abil EM-90 | Diehl *et al* [115] |

### 4.2.1 Analysis of picoliter reactors

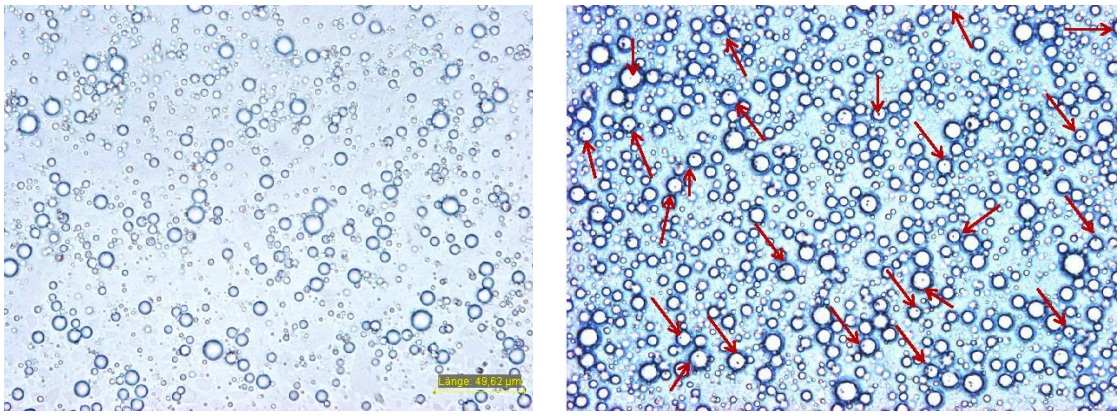
The picoliter reactors created through different oil mixtures were analyzed under the light microscope. 5 µl of emulsion were streaked on a glass slide and observed at 40x magnification as shown in figure 4.2 A. Initially all emulsions from different oil mixtures showed stable and homogenous emulsion formation which remained stable at room temperature upto one week. Stability was tested under temperature cycling conditions for mock emulsion PCRs with different oil mixtures. Here it was observed that the emulsion prepared from Span 80 was disrupted after 37 cycles, whereas the emulsion prepared using ABIL EM 90 remained stable upto 50 cycles (Figure 4.2 B). So, the latter oil mixture was used for all subsequent emulsion preparations.

The number of reaction centers in emulsion was calculated with haemocytometer under the microscope. The number of picoliter reactors was estimated from three

seperate oil emulsions to be approximately $10^{10}$-$10^{11}$ per mililiter of emulsion and the size of reaction centers ranged from 2-5 μm (figure 4.3).

A.

I. II.



B.

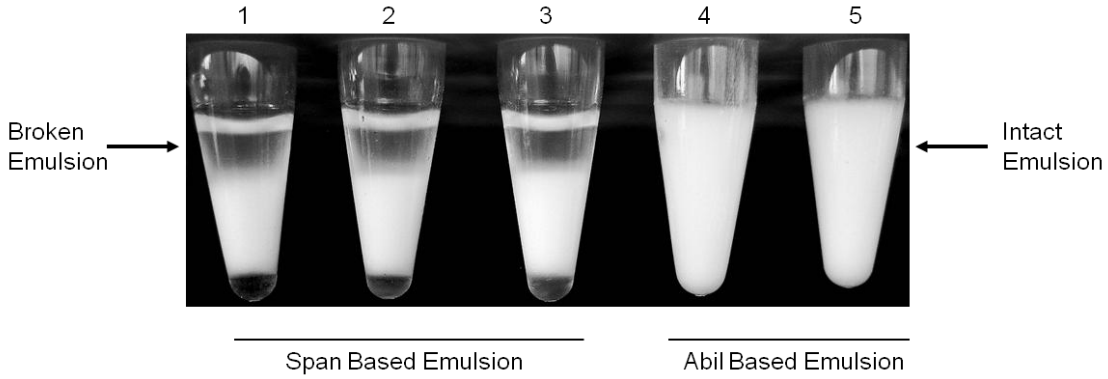



**Figure 4.2 Physical Characterization of Emulsion droplets.** A. Emulsions droplets prepared under standard conditions as described in and observed under the microscope (I) Emulsions droplets without any beads, (II) emulsion droplets containing beads in them. droplets containing single beads are marked with an arrow. B. Emulsion prepared with Span based and Abil based. They underwent PCR cycling for 40 cycles and then Images of the PCR tubes were taken. The numbers on the top of each tube corresponds to the oil composition as shown in the table 4.1.

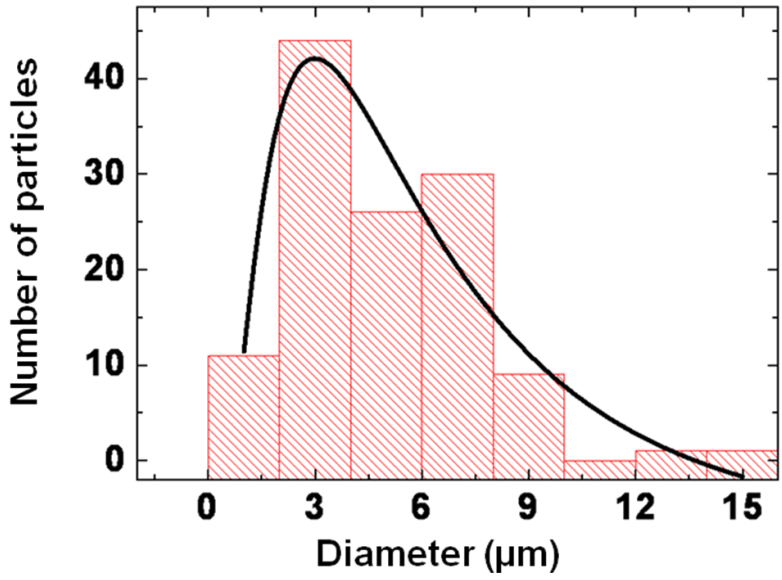



**Figure 4.3 Determination of average diameter of emulsion droplets.** A graph was plotted with the number of particles versus diameter as obtained after processing the microscopic image of emulsion droplets as described in methods chapter previously (p 49, 3.5.1). It shows a peak at 3 μm and a log-normal distribution. The black solid line is a log-normal fit to the curve. The fit yields a mean diameter of 5.64 μm and a broad size distribution.

## 4.3 PCR in water in oil emulsion

Emulsion PCR is a type of PCR, which is performed in millions of picoliter reactors created through water in oil emulsion. Each reactor is a reaction centre with its own reagents containing a single template and there is no exchange of macromolecules between different reaction centers. EmPCR is being applied successfully in next generation DNA sequencing techniques like 454 sequencing [116] and microfluidic applications [140–142]. It has been successfully applied in digital PCR for identifying mutations in cancer alleles through BEAMing [114] and it was also successfully shown to prevent generation of chimeric products through recombination while amplifying complex gene mixtures [138]. EmPCR involves primer coupled beads which would then clonally amplify and capture DNA on the beads. The rationale to apply emPCR in this approach was not only to capture multiple copies of single gene variant beads but also the beads would provide monoclonality to the system till the selection of phenotype when the emulsions are disrupted. Also the presence of multiple copies of the same genotype- phenotype on a same bead, this would increase sensitivity and stability of the system.

### 4.3.1 Amplification of gene products through emPCR

In order to standardize PCR condition in emPCR and to obtain similar amounts of PCR products both in solution and emPCR, the protocol described by Williams *et al.* [138] was followed. After the PCR, both the products were ran on an agarose gel as shown in figure 4.4 A. As it can be seen clearly from the figure, similar amount of PCR product was obtained both in emPCR and PCR in solution. It is also clear that product amplified through emPCR was also devoid of any byproducts.

After obtaining comparable amounts of PCR products both in emulsion and solution, the capability of emPCR to prevent any size competition was also tested. For this purpose two different templates of different sizes containing the same primer binding sites were used. The templates used were pGFP-SNAP and pMS2-SNAP (Figure 4.1). The contruction details of these templates are discussed in *methods chapter* (p.45, 3.2) and the salient features of these constructs will be discussed in the later sections of the results. Using the same set of primers, pGFP-SNAP gives an amplicon of 1.6 kbp in size whereas pMS2-SNAP gives an amplicon of 1.2 kbp in length. As seen from the figure both templates were amplified seperately in similar amounts both in emulsion and in solution with the same set of primers (Figure 4.4 B lane 1 and lane 2, lane 5 and lane 6).

Then to ascertain if there is any interchange of primers or other reagents between different reaction centers, emulsion PCR were assembled with pGFP-SNAP and pMS2-SNAP in different tubes. However in the assembly of the shorter template pMS2-SNAP, the primers were excluded. Both the emulsions were then mixed and PCR was set up. As it can be seen from the figure 4.4 lane 4 only the long template was amplified (pGFP-SNAP) and not the shorter template which implied that there was no transfer reagents between the different emulsions and the picoliter reactors remained intact till the 45 cycles of PCR. But when two templates were mixed and amplified both emPCR and PCR in solution with all the reagents present in both PCRs, only the smaller fragment is amplified. This experiment shows that emPCR

is capable of preventing any size competition and amplify specific PCR fragments. The experimental set up in this experiment was similar to as described in Williams *et al* [138].

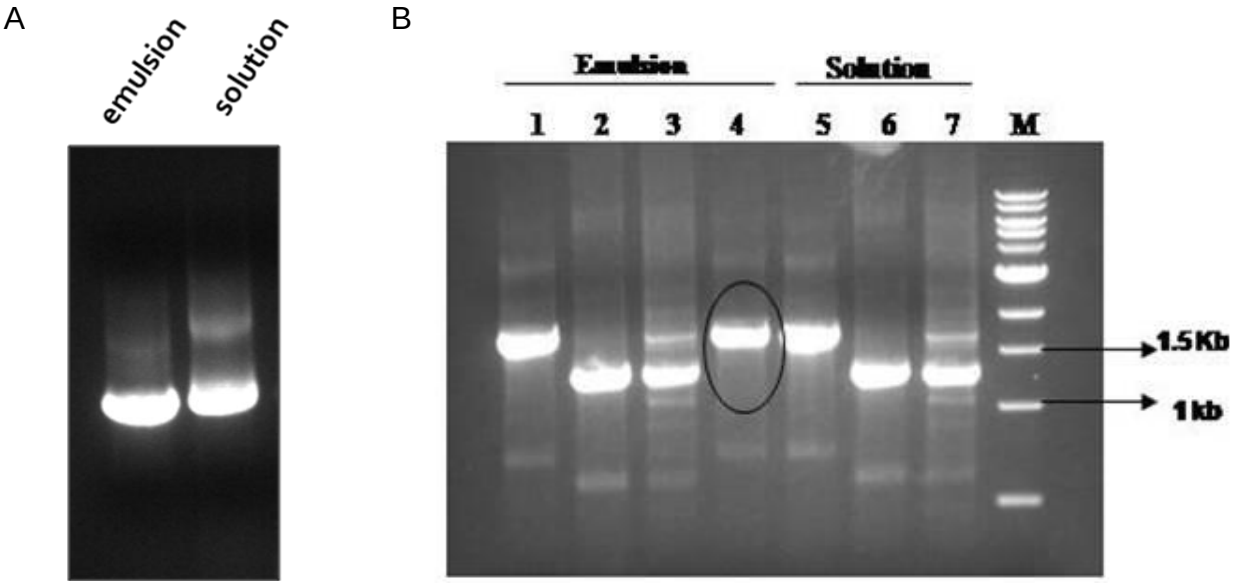



**Figure 4.4 Amplification of gene products through emPCR**. (A) Agarose gel analyzis depicting the comparision of PCR products generated through emPCR and PCR in solution. (B) Agarose gel analysis of PCR products, *lane 1*- pGFP-SNAP (emulsified and amplified through emPCR), *lane 2*- pMS2-SNAP(emulsified and amplified through emPCR), *lane- 3* pGFP-SNAP and MS2-SNAP mixed (emulsified together and amplified through emPCR), lane- 4 pGFP-SNAP and MS2-SNAP (emulsified separately, then mixed and amplified through emPCR), *lane- 5* pGFP-SNAP (amplified through PCR in solution), *lane- 6* pMS2-SNAP (amplified through PCR in solution), *lane-7* pGFP-SNAP and MS2-SNAP mixed (amplified through PCR in solution).

## 4.3.2 PCR on microbeads in emulsion

As outlined in the principle of multi copy bead display approach (Figure 1.9), the first step is to capture multiple copies of DNA on microbeads through emPCR. The use of microbeads would serve as a solid support to the captured DNA and protein thereafter. These microbeads are 1 micrometer in diameter and paramagnetic in nature. These beads also have activated carboxyl (-COO) groups on their surface. The use of magnetic beads provides the ease of handling compared non magnetic beads especially during washing steps as the entire solution

can be removed from them by magnetic separation and also the loss of beads are comparative less as compared to non-magnetic beads.

#### 4.3.2.1 Analysis of forward primer on the beads

In order to capture PCR products on the beads, the forward primer is coupled to their surface. The protocol of coupling is described in details in the *methods chapter.*(p.47, 3.3) The oligos used for coupling are amino modified at the 5' end. They react to activated carboxyl group on the beads, forming an amide bond between them. A number of reports have discussed the problem of steric hindrance during PCR amplification on beads [143,144] and it is recommended to add a minimum spacer of 15 bases before the primer binding sequence to lessen the effect of steric hindrance[145]. Keeping that in mind the forward primer (15bT7s2+TGspacer) used for coupling to the beads included a spacer of 30 nucleotides composed of T and G upstream of the primer sequence.

The quality of coupling was analyzed by hybridizing a fluorescently labeled oligo complementary to the coupled primer on the beads. Beads were then analyzed in flow cytometry. The hybridization protocol is described in *methods chapter (*p.51, 3.5.3). As seen from the figure 4.5, a single sharp peak for most of the beads was achieved which indicated that most of them had captured similar amounts of oligos. It can also be seen from the figure that hybridizing a non specific fluroroscently labeled oligo to the coupled primer results in very low signals confirming that the hybridization was highly specific.

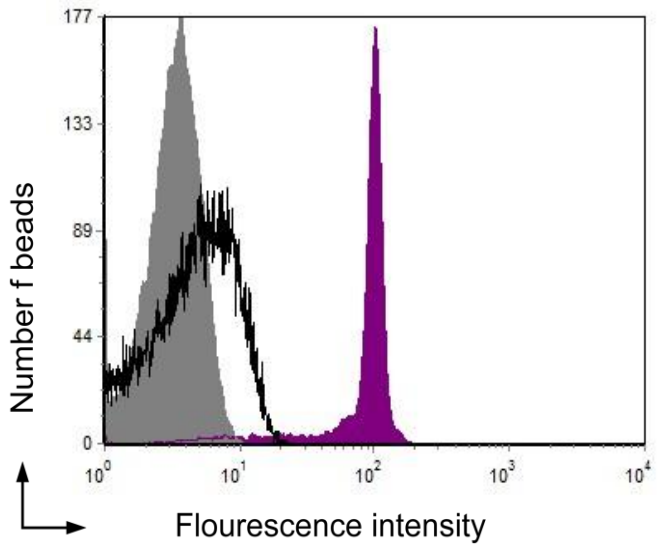


**Figure 4.5 *(continued)* Analysis of forward primer coupled to the microbeads.** Beads coupled to forward primer were hybridized with fluorescently labeled oligo complementary to the coupled primer and analyzed in flow cytometry. The levels of fluorescence (FL1H) are plotted as histograms. Overlays of beads coupled to primer and hybridized with complementary oligo (purple filled) or beads coupled to primer and hybridized to a non specific oligo (black line) or beads without any coupled primer and hybridized with the fluorescently labeled oligo (filled grey) are shown.

To determine the right amount of oligo required for coupling to the beads so that all the active sites on the beads are saturated, a titration with different amounts of coupling oligo was performed. They were then analyzed in flow cytometry in the same manner as described above. By comparing the mean fluorescence intensity of all of them, 15 μM and 20 μM of primers were able to saturate binding sites on each bead (figure 4.6). Thereafter 20 μM of primers were used for all coupling on the beads (for $10^9$ beads).

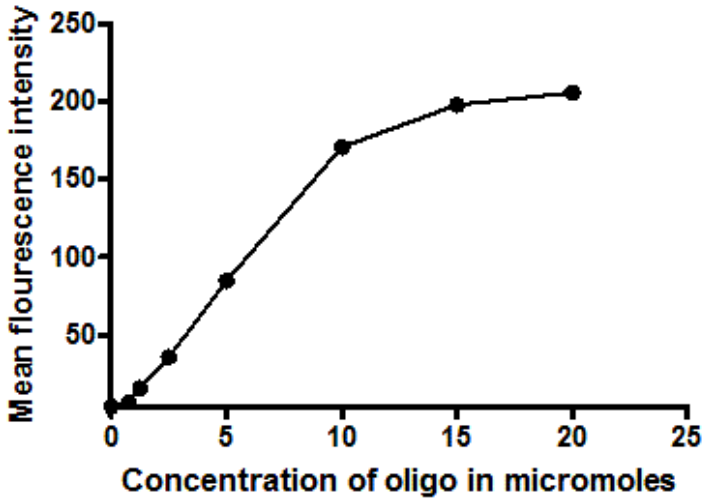



**Figure 4.6 Optimization of amount of coupling primer to beads.** Different sets of beads were coupled with different concentrations (0.5, 1, 2.5, 5, 10, 15, 20 μM) of forward primer. These beads were then hybridized with fluorescently labeled oligo complementary to the coupled primer and analyzed in flow cytometry. A graph was plotted with concentration of coupled oligos versus mean fluorescence intensity.

### 4.3.1.2 Analysis of reverse primer coupled to BG

As outlined in the principle of multi-copy bead display approach (Figure 1.9), genotype-phenotype linkage is created through BG-SNAP interaction. The proteins

translated from the gene bound on to the beads are captured back to its coding DNA through the SNAP domain of the fusion proteins which interacts with BG present on the encoding gene. The BG is introduced to the PCR product on the beads through the reverse primer which is coupled to BG in emPCR.. The details of the coupling are described in the *methods chapter* (p48, 3.4). Functionality of BG on the primer was analyzed by its ability to bind to SNAP. An aliquot of GFP-SNAP fusion protein was incubated with BG coupled reverse primer and then analyzed in western blot probed with anti-GFP antibody. As seen in figure 4.7, the primer incubated with GFP-SNAP induced a clear shift as compared to GFP-SNAP without any incubation with primer. This proved that the primer coupled to BG is functionally active with its interaction with SNAP protein.

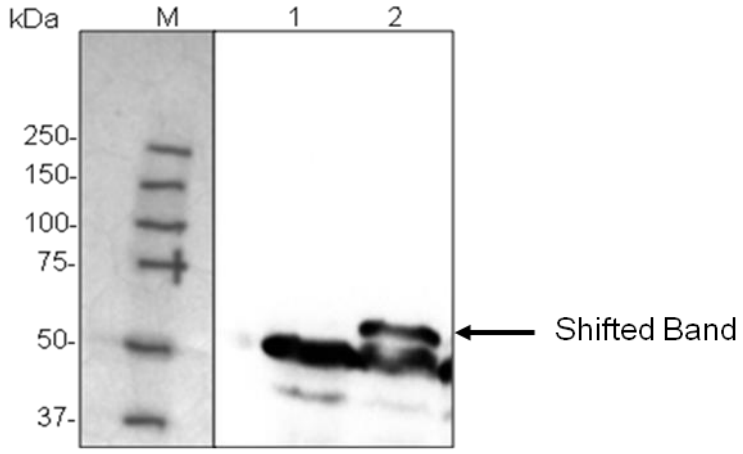



**Figure 4.7 Analysis of functionality of BG-coupled primer.** BG-coupled reverse primer was incubated with GFP-SNAP and analyzed through western blotting with anti-GFP antibody. *Lane 1*: GFP-SNAP protein without incubation with BG coupled oligo, and *lane 2* shows GFP-SNAP protein incubated with BG-coupled oligonucleotide**.**

## 4.3.2 Standardization of PCR on beads in water in oil emulsion

There is an inherent difference between normal emulsion PCR and emulsion PCR with beads as solid support. The initial protocol described by Williams *et al.* [92] for emulsion PCR failed to work in emulsion PCR with beads as solid support. Another approach described by Diehl *et al.* [115] - molecular BEAMing, which

includes emulsion PCR on beads was tested for multi copy bead display. However, this strategy is shown to work efficiently for amplifying amplicons with a maximum size of 400 bp, which is not sufficient for our approach, where around 1.6 kbp need to be amplified. Also, this approach failed to amplify and capture adequate amounts of PCR products on the beads. Analyzing the amount of PCR product on beads through real time PCR showed an average of only 300 copies. To screen phenotype effectively in flow cytometry, a minimum of around 2000 fluorescence molecules are required. Although, 300 copies of DNA would translate to a lot of proteins but eventually only 300 protein molecules will be captured back on to the beads and it would not provide enough signal for screening in flow cytometry. Therefore, it is important to have an optimized emPCR which would capture more than 300 DNA on beads and subsequently more proteins.

#### 4.3.2.1 Supplementing with soluble forward primer for emPCR

The need for a certain amount of soluble primer in emPCR with microbeads has been discussed before [110]. This is important because each picoliter reactor contains only one single DNA molecule to be used by the primer present on the beads. However, probably due to steric hindrance on the beads it might become difficult to amplify long templates. The rationale to use soluble primer is to amplify few DNA copies during the initial rounds of PCR and thereby produce more copies of the same template, which would be available to be amplified and captured on the beads thereafter. However, it is also very important not to use saturating amounts of soluble primer, because it would compete out the primer on the beads and lead to accumulation of PCR products in solution. The soluble primer used for this purpose (15s2T7) has the same sequence as the forward primer coupled on the beads (15bs2T7+TG spacer) without the TG spacer). A titration of different amounts of soluble primer was done for emulsion PCR. After emPCR, PCR products on the beads were incubated with SYBR Green, which is known to intercalate into double stranded DNA. The beads were then analyzed in flow cytometry. As seen from figure 4.8, the PCR reaction with 0.01 μM of soluble

primer showed the best signal. It can also be seen that 1 µM soluble primer induces a reduction of the signal, which indicates that soluble primer competes out the primer on the beads and most of the PCR products are in solution – an effect, whose tendency can be seen already at 0.1 pM soluble primer. Addition of 1 µM primer also showed some signal at $10^4$ which probably appeared from non-covalent binding of PCR products from solution on to the beads. It was therefore decided to add 0.01 µM soluble primer for subsequent emulsion PCRs.

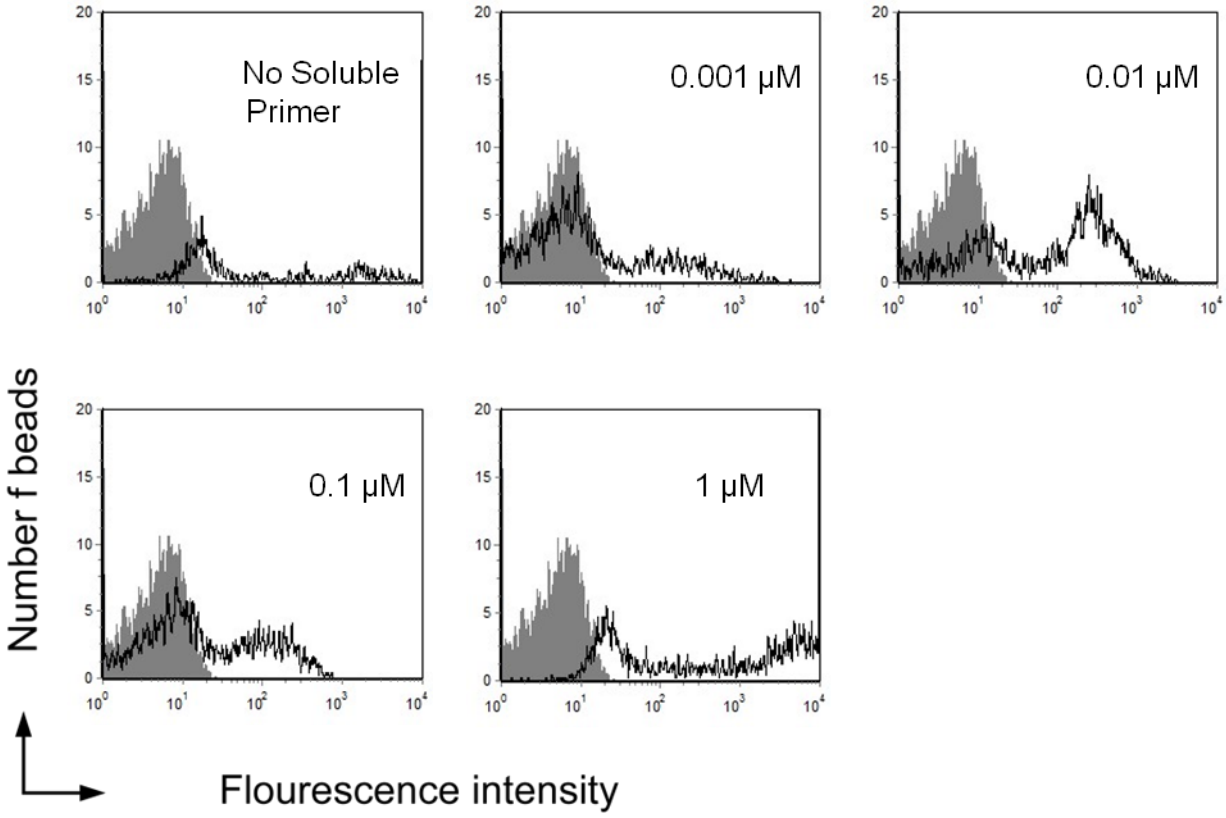



**Figure 4.8 SYBR Green staining of DNA on the beads.** Emulsion PCR on beads was performed with different concentrations (0.001, 0.01, 0.1, 1 µM) of soluble forward primer. The PCR products captured on beads were then incubated with 0.1 pM of SYBR Green solution for 15 min and analyzed in flow cytometry. The levels of fluorescence (FL1H) were plotted as histograms. Overlays of beads with PCR products (black line) stained with SYBR green and beads which underwent without any templates (Grey filled) stained with SYBR Green are shown.

### 4.3.2.2 Optimization of the reverse primer concentration

20 µM of sense primer was used to couple to around $10^9$ bead of which $10^8$ beads were used for emPCR. Considering the fact that huge amount of forward primers

were present on the beads and also some amount soluble primer was added seperately to the PCR reaction, reverse primer was also required to be increased to further optimize the amount of PCR products on beads. Emulsion PCR with different amounts of reverse primer was performed to further optimize the capture of PCR products on beads. The final concentrations of the primers used are 0.3 μM, 1 μM, and 3 μM respectively. After emulsion PCR, the beads were treated with 0.1 N NaOH to melt the double stranded DNA on the beads. The strand in solution is washed away and the DNA strand on the bead is then hybridized with a fluorophore containing oligo which is complementary to a segment of the DNA on the beads, which are then analyzed in flow cytometry. This is described in details in the *method chapter* (p.51, 3.5.3). As seen from the figure 4.9, increasing reverse primer led to more fluorescence signal which indicates that PCR products captured on beads increased substantially. The use of 3 μM of reverse primer showed the best signal however, it was still less than the positive signal (beads containing single stranded forward primer coupled to beads and hybridized with a complementary oligo). As seen from the figure, there is an appearance of a small populaton of beads, which had the same mean fluorescence intensity as the positive control beads. This might be due to the fact that in the corresponding picoreactors, either the reactors were bigger so they could accommodate more than one DNA. This amount of reverse primer was used for all subsequent emulsion PCR.

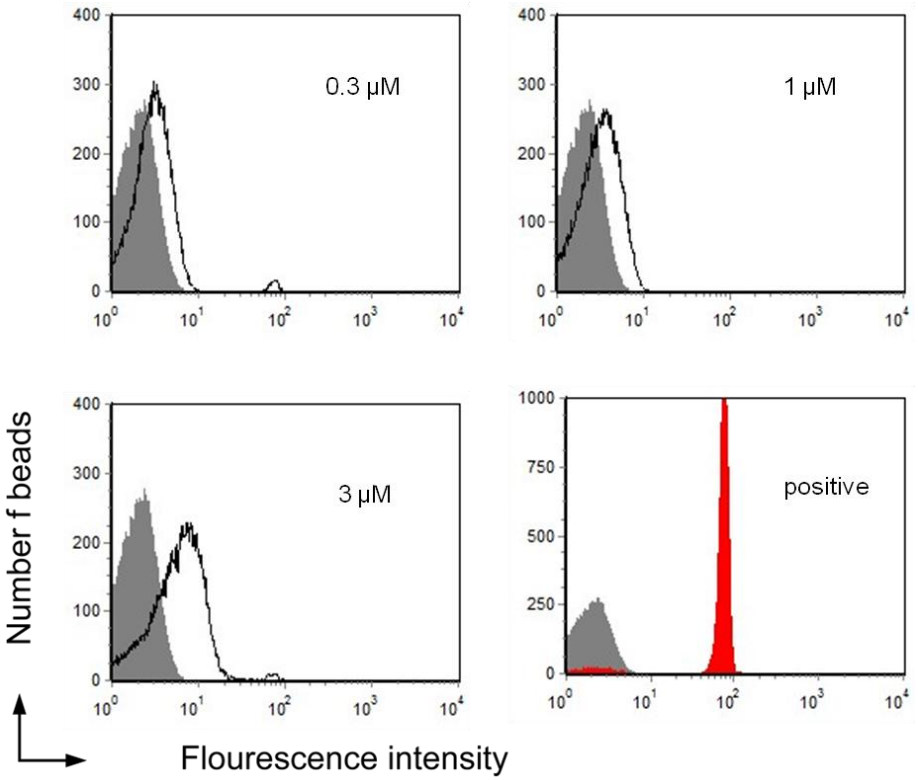



**Figure 4.9 Reverse primer concentration optimization.** Emulsion PCR was performed on beads with different amounts of reverse primer (0.3, 1, 3 µM). After PCR, the products on beads were melted and the remaining single strand DNA on the bead was then hybridized with a fluorophore coupled oligo complementary to a part of it. The level of fluorescence is plotted as histograms. Overlays of beads with DNA ( black line) and beads without DNA are shown. As a positive control (red), beads coupled to an oligo and hybridized with a complementary oligo to it is shown, too.

The utilization of higher amounts of reagents in emPCR is important because of the fact that there is a chance of exhaustion of reagents in some reactors during PCR. Apart from above mentioned modifications, higher amounts of dNTPs and taq polymerase were also introduced in all future emPCRs. The optimized protocol for emPCR on beads is described in the methods section. The amount of PCR products were analyzed again through real time PCR and showed a substantial increase in the copy numbers compared to non-optimized conditions: 800-1000 copies per bead.

## 4.4 Expression of protein in water in oil emulsion

### 4.4.1 Construction and expression of Proteins

As outlined in the principle of multi-copy bead display approach (Figure 1.9), expression of protein encoded by DNA on beads is performed in a second water in oil emulsion followed by subsequent capture of the proteins on the encoded DNA before passing to the screening process. The proteins were seperately analyzed for expression in *in vitro* transcription translation (IVTT). The templates used for establishing of IVTT, encode for GFP-SNAP and MS2-SNAP fusion proteins under the control of a T7 promoter (figure 4.1 A).

### 4.4.2 Expression of fusion protein in IVTT in emulsion

The GFP-SNAP fusion protein was expressed in an IVTT system both in solution and in emulsion, respectively and was analyzed for their activity through Western blot probed with anti-GFP antibody. A number of different IVTT systems were tested for their ability to express proteins in emulsion as listed in *materials* (p.33, 2.5.1.1) However, most IVTT kits worked well when plasmid was used as a template for expression both in solution and emulsion, but most of them failed to have a good expression when PCR products were used as templates (figure 4.10 A) except for rat reticulocyte expression system specific for PCR products (figure 4.10 B). This is very important because in subsequent experiments with libraries the encoding DNA on beads were linear in nature. It is also important to note that, rat reticulocyte expression systems work only for oil mixtures containing ABIL EM 90. Other oil mixtures lead to the rapid discoloration and translation shutdown in rat reticulocyte system [146]. The oil mixture with ABIL EM90 which was used for emPCR was also employed for the IVTT in emulsion step.

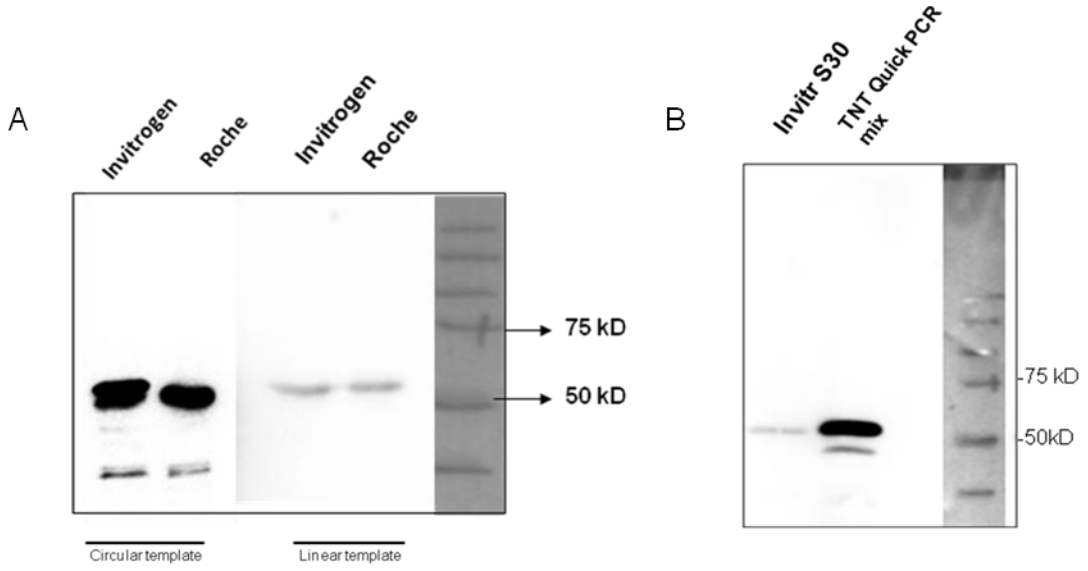



**Figure 4.10 Expression of GFP-SNAP in IVTT**. (A) Different IVTT systems were tested of which only two of them are shown here. First two lanes show the expression of the fusion protein with plasmid as template for IVTT and second two lanes shows expression of fusion protein with PCR product as template for IVTT. (B) TNT rat reticulocyte system for PCR product in IVTT is shown in comparison to S30 bacterial expression system using a PCR product as template for both IVTTs.

The SNAP partner of GFP-SNAP fusion protein was also analyzed for its functionality to interact with BG. For this purpose, BG-agarose beads were incubated with an aliquot of GFP-SNAP fusion protein expressed through IVTT in emulsion. Then the beads were analyzed under a fluorescence microscope. Only Beads incubated with GFP-SNAP showed fluorescence under the microscope as compared to beads with no proteins incubated (figure 4.11 A and B). To ascertain, that the interaction between BG and SNAP was specific, a competition assay was also performed, where the expressed GFP-SNAP fusion protein was first incubated with excess amount of soluble BG and then loaded on to BG-agarose beads. Fluorescence microscopic analysis of these beads showed only minimal binding of the protein (Figure 4.11 C), which indicated complete competition of SNAP by soluble BG and proved that the IVTT expressed GFP-SNAP fusion protein in emulsion was functionally active.

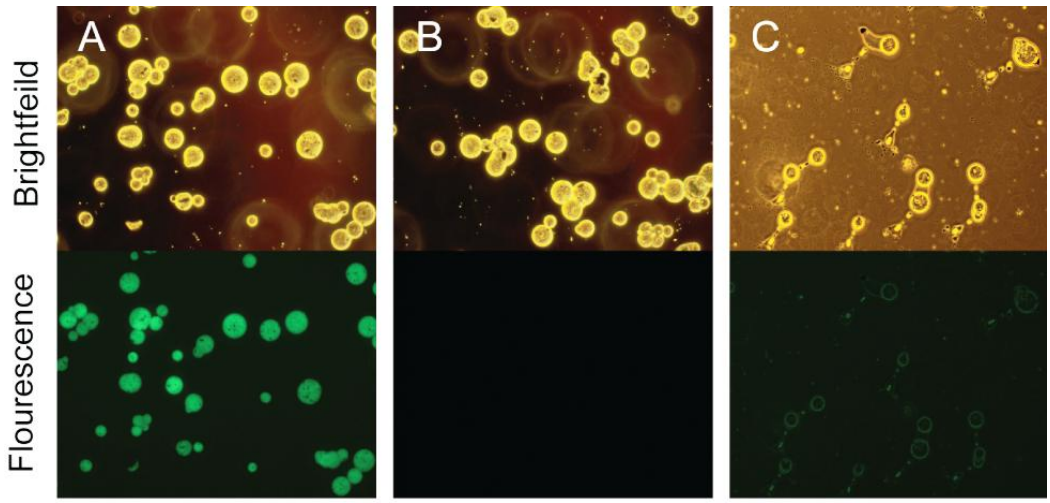



**Figure 4.11 Expression and functionality of the GFP-SNAP fusion protein.** PCR products spanning the GFP-SNAP expression cassette were expressed in a cell free IVTT reaction. BG-coupled agarose beads were incubated with the IVTT reaction for 30 min at room temperature and analyzed by fluorescence microscopy. The upper panel shows bright field images of the beads, the lower panel the images obtained by fluorescence microscopy. (A) BG-coupled beads after incubation with the GFP-SNAP-primed IVTT reaction. (B) BG-coupled beads without the GFP-SNAP-primed IVTT reaction. (C) GFP-SNAP-primed IVTT reactions were first incubated with an excess of BG prior to adding the IVTT reaction to the BG-coupled beads.

## 4.5 Coupling of DNA and encoded proteins to beads by emulsion PCR and emulsion IVTT

After initial standardization of emPCR on beads and IVTT in emulsion separately, the entire process of multi-copy bead display was performed with GFP-SNAP template. Approximately $10^8$ beads coupled with forward primer were used for multi-copy bead display. $3\times10^9$ linear templates encoding GFP-SNAP fusion were added to emPCR which amounts to 0.3 times the minimal amount of the number of the reactors (so that one out of three reactors gets one copy of template considering that there are $10^{10}$ reactors in total). EmPCR on beads was performed with the optimized protocol as discussed above. After PCR, beads were isolated from the emulsion and DNA captured on beads was allowed to express encoded GFP-SNAP fusion protein through IVTT in a second emulsion.

After IVTT in emulsion, beads were extracted and analyzed for GFP-SNAP expression through flow cytometry. In contrast to the experiment with BG-agarose beads the amount of protein on the beads was not enough to be directly detectable by flow cytometry or microscopy. This may be due to the fact that the size of the beads used for display is 1.0 micrometer whereas the size of BG-agarose beads is 50- 70 micrometer, therefore the protein loading on them was higher and can be directly analyzed under the microscope. Also, the amounts of BG molecules on the magnetic beads were limited. To enhance the signal on the microbeads, beads were stained with an anti-GFP primary and a fluorescently-labeled secondary antibody. However, the signal on those beads were still weak and not stable which might be due to the fact that the number of BG binding sites on the beads were still not normalized.

### 4.5.1 Normalization of BG binding sites with BG oligo hybridization

It might be a factor that not enough number of BG molecules was available on the beads for the sufficient SNAP to bind. This might have arised due to some of coupled primers were not extended correctly and thereby led to generation of truncated products on the beads without any BG on them. This would lead to less stable and weak labeling of translated proteins thereafter. To overcome this, an intermediate step was introduced in the multi-copy bead display. In this step, BG molecules on the beads were normalized by hybridizing a BG-labeled oligo complimentary to the capture oligo of the bead. This would therefore ensure that averagely the same number of BG residues is present on each bead at subsequent IVTT and is not totally dependent of completely efficient emPCR. The protocol of normalizing BG molecules on beads is described in details in the *method chapter* (p.52, 3.6 and figure 3.4). This step was thereby introduced subsequently in all the latter experiments.

## 4.5.2 Flow cytometric analysis of Protein on beads

After the realization of the final standardized protocol of multi-copy bead display system, the displayed GFP-SNAP proteins on the beads were analyzed in presence of different staining controls. Also to avoid the possibility of covalent linkage of an excess of expressed SNAP fusion proteins of one picoliter reactor to free BG moieties on beads from other reactors during breaking-up of the emulsion, excess amounts of free BG were added at the recovery step of the beads from the IVTT emulsion reaction. Using these conditions, the displayed proteins on DNA-coupled beads were recovered from the emulsion IVTT and stained with the anti-GFP antibody and then fluorescently labeled secondary antibody. A population of GFP-positive beads can clearly be detected by flow cytometry (figure 4.12).

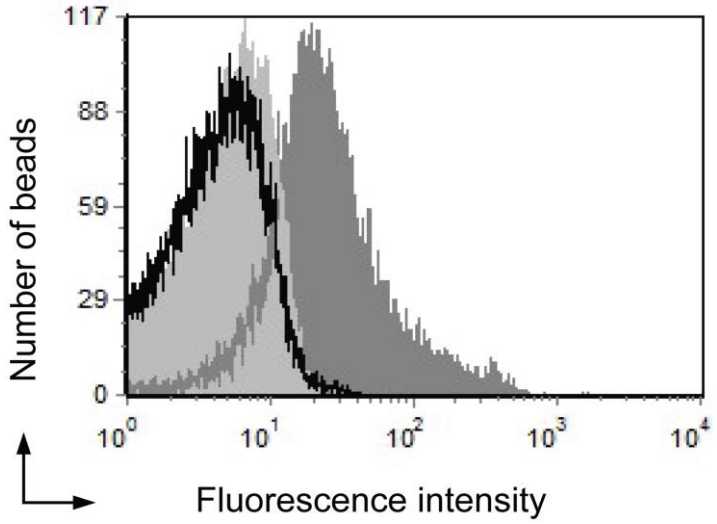



**Figure 4.12 Detection of the GFP-SNAP fusion protein after multi-copy bead display by flow cytometry**. Filled dark grey: Beads were loaded with DNA encoding GFP-SNAP and added to a compartmentalized IVTT reaction. After recovery of the beads from the IVTT reaction, the beads were stained with an anti-GFP antibody and a fluorescently labeled secondary antibody prior to flow-cytometric analysis. After gating on the single beads, the fluorescence intensity of single beads is plotted as a histogram. Filled light grey: beads displaying GFP-SNAP proteins were incubated with fluorescently labeled secondary antibody in absence of the anti-GFP primary antibody. Black: beads were processed through emulsion PCR and IVTT without a GFP-SNAP DNA template and then stained for GFP.

## 4.6 Clonality and sensitivity of the multi-copy bead display approach

After having established the principal methods for the multi-copy bead display, mixing experiments were performed to confirm clonal amplification of the template and to determine the sensitivity for the detection of GFP-SNAP in the presence of an excess of a control SNAP fusion protein. The template encoding the coat protein of the bacteriophage MS2 fused to SNAP (figure 4.1) was mixed at different ratios with the template encoding GFP-SNAP. As mentioned before, both these templates have the same primer binding site for amplification of the expression cassettes. The ratios used for mixing of MS2-SNAP to GFP-SNAP are 1:1, 10:1, 100:1, and 1000:1, respectively. After multi-copy bead display of the mixed DNA templates, the beads were stained with anti-GFP antibody and fluorescently labeled secondary antibody. At a 1:1 ratio, there are clearly two populations of beads, one being GFP positive, the other GFP negative (Figure 4.13 A). This indicates that at the copy, numbers of template added to the emulsion PCR, each bead is loaded with only DNA encoding for one of the SNAP fusion proteins. It also confirms, that cross-contamination of SNAP fusion protein produced in one picoliter reactor to the bead of another picoliter reactor does not occur at detectable levels. Reducing the amount of GFP-SNAP template gradually decreased the GFP-positive bead population, but not the mean fluorescence intensity of the positive beads. However, even at a 1000-fold excess of the MS2-SNAP template, a GFP-positive bead population remains detectable (Figure 4.13 D), which indicates that the multi-copy bead display approach is sensitive.

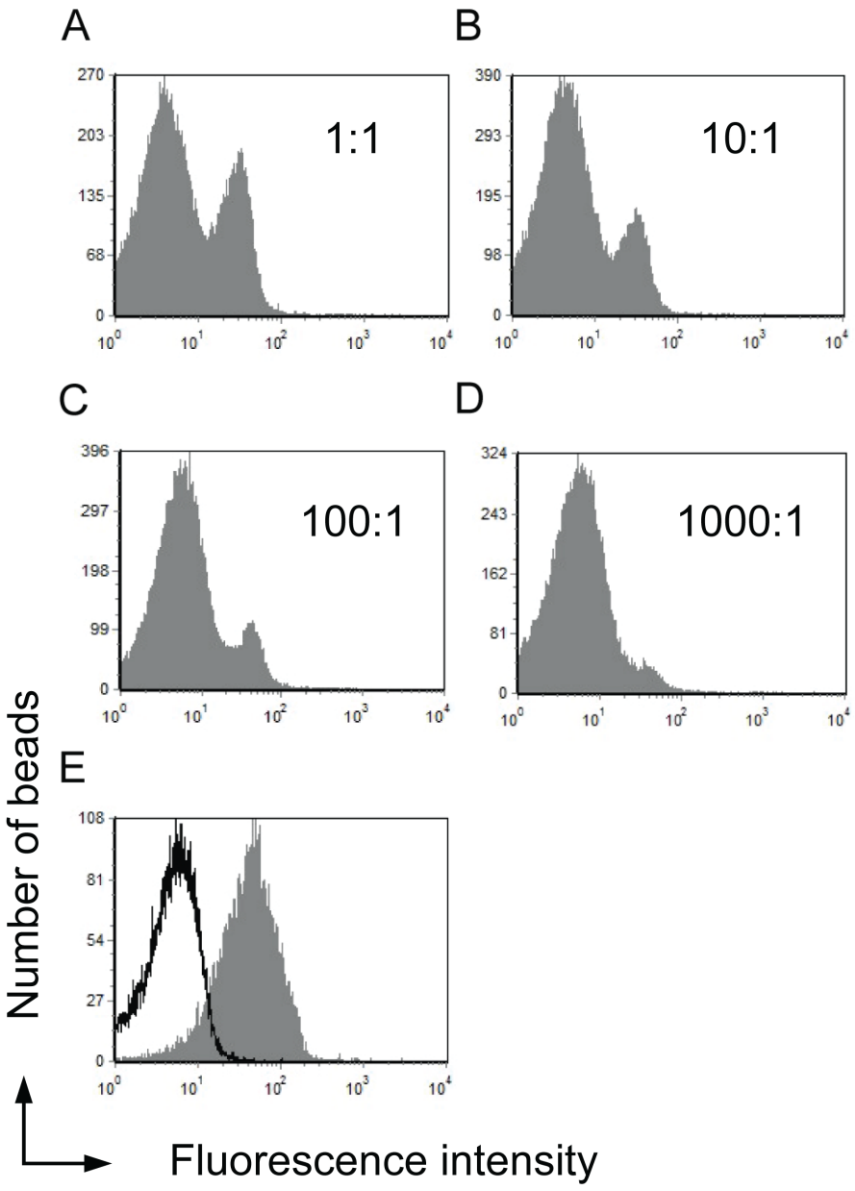



**Figure 4.13 Detection of GFP-SNAP carrying beads in the presence of an excess of a competing SNAP fusion protein**. (A-D) Bead-containing emulsion PCRs were performed with the indicated ratios of MS2-SNAP to GFP-SNAP template DNA. Beads were then transferred to compartmentalized IVTT reactions and stained with anti-GFP antibody and fluorescently labeled secondary antibody prior to flow-cytometric analysis. (E) Overlay of histograms of beads loaded with either GFP-SNAP (filled grey area) or MS2-SNAP (black line) only.

## 4.7 Estimation of number of DNA molecules per bead in multi-copy bead display

Although it was estimated in initial experiments, the average number of DNA molecules captured on the beads after emPCR with real time PCR analysis to be

between 800-1000 copies per bead, another estimation of the number of copies of DNA molecules post multi-copy bead display was done, which would provide a realistic estimation of the DNA molecules on beads after the entire process. To perform this, $3\times10^9$ copies of GFP-SNAP templates were added to emPCR. After PCR, the beads carrying GFP-SNAP DNA were then incubated with IVTT mixture in emulsion supplemented with increasing concentration of soluble MS2-SNAP encoding DNA. This is done with the intent that with increasing concentration of MS2-SNAP DNA, which would eventually translate equally as GFP-SNAP present on beads and compete GFP-SNAP protein for BG-binding sites on beads. Starting concentration of MS2-SNAP was chosen 100-fold higher than the amount of GFP-SNAP template added as template for emulsion PCR. The concentration of MS2-SNAP DNA was increased up to 1000-fold ($1\times10^{12}$ copies). The experiment is described in detail in the *methods chapter* (p.54, 3.10 and figure 3.5). The beads recovered from IVTT emulsion were stained with anti-GFP antibody and fluorescently labeled secondary antibody. As seen from the figure 4.14 addition of $1\times10^{12}$ copies of MS2-SNAP DNA completely competed out GFP-SNAP protein for BG-Binding on the beads indicated by a reduction of the fluorescence signal bringing it close to signal of GFP negative beads (beads which underwent bead display with without any template). However, with $1\times10^{11}$ copies of MS2-SNAP, competition was less than 50%. Fine tuning of this competition assay i.e. to achieve approximately 50% competition, showed that $4\times10^{11}$ copies of MS2-SNA was able to reduce the mean fluorescence intensity to half maximal, bringing it in between those from GFP negative and GFP positive (GFP-SNAP without any competition) beads. Approximately, $10^8$ coupled beads were used for multi-copy bead display. Assuming that the beads remained intact till the end of the display without any loss during the process, the average number of DNA copies present on each bead would be around 500.

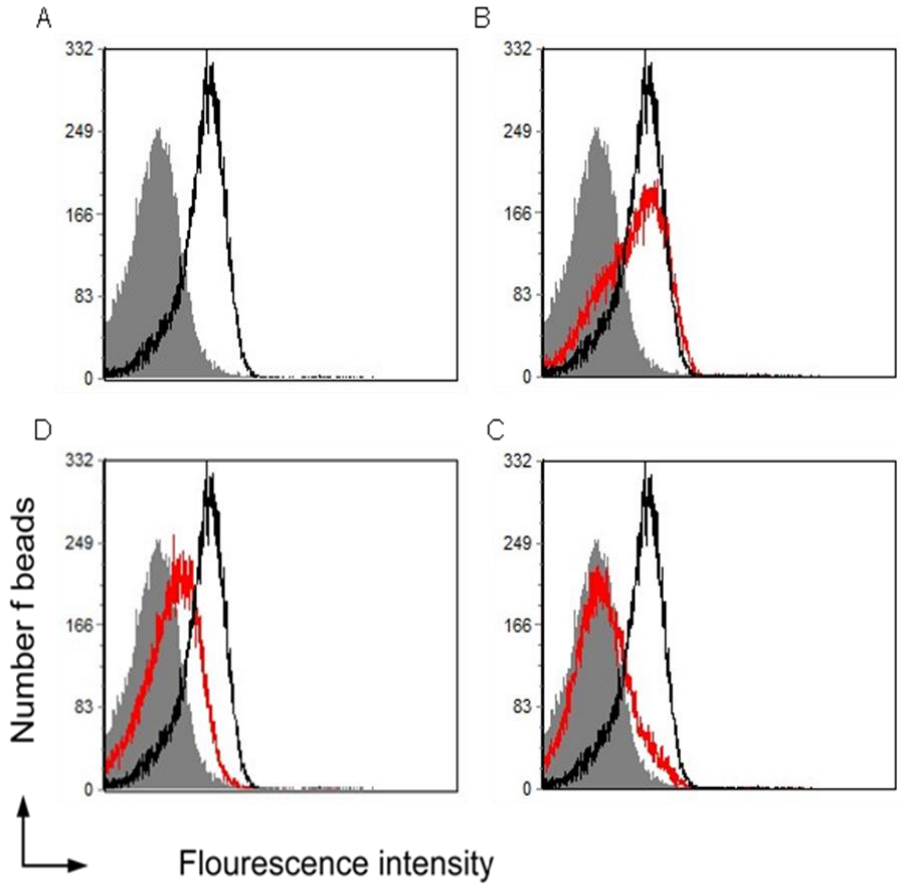



**Figure 4.14 Generation of mixed populations on beads.** Different sets of beads with gene product (GFP-SNAP) after emulsion PCR was expressed in IVTT in a second emulsion in presence of increasing amounts of soluble MS2-SNAP DNA. Both DNAs expressed their respective fusion proteins, which competed for BG binding on the beads. The beads were then labeled with anti-GFP antibody and fluorescently conjugated secondary antibody and analyzed in flow cytometry. The level of fluorescence (FL1H) was plotted as histograms. Overlay of histograms of beads with no GFP-SNAP template DNA processed through emulsion PCR and emulsion IVTT (filled grey area), loaded with only GFP-SNAP (black line) or beads with GFP-SNAP expressed in presence of MS2-SNAP (red line) are shown. Overlays of histograms (A) in absence of MS2-SNAP, in presence of $1\times10^{11}$ (B)„ $4\times10^{11}$ (C) and $1\times10^{12}$ (D) copies of MS2-SNAP.

## 4.8 Application of Multi-copy Bead Display Approach

### 4.8.1 T7 promoter library and its display

After establishment of multi-copy bead display, the first application chosen was the development of T7 promoter library and selection of a T7 promoter mutant with enhanced *in vitro* activity. Phage promoter has evolved to have efficient expression in prokaryotes. There can be other T7 promoters, which might be more efficient than wildtype in its expression in different IVTT. Also understanding of promoter

activities of different promoters in synthetic circuits is one of the aims of researchers in synthetic biology [147–149]. With this idea, the T7 promoter library was chosen as the first application of multi-copy bead display library.

The T7 bacteriophage promoter is 22 nucleotides in length, which comprises a polymerase binding site and a transcription initiation site (Figure 4.1 B). For generation of the T7 promoter library, 10 nucleotides of the transcription initiation site were chosen and randomised completely. (Figure 4.1 B). This resulted in a theoretical library complexity of $10^6$. The library generation was done in a PCR based setting through oligonucleotides containing randomised T7 promoter sequence. They were introduced upstream of GFP- SNAP fusion template. This is descibed in the *methods chapter* (p.55, 3.11)

#### 4.8.1.1 Optimal expression condition for T7 promoter variants

The quality of promoter activity of any promoter depends on its ability to express proteins that is under its control effectively. In order to evaluate the promoter on the basis of the protein expressed a balance between time of expression and the amount of protein expressed is essential. However, with limited binding sites present on the beads, even promoters with weak activity which could produce proteins at extended expression time and bind to these sites leading to overloading and which in turn would lead to false selection during screening steps.

To circumvent this, a time dependant expression of GFP-SNAP was performed with the wildtype T7 promoter, and protein loading on beads was measured. The protein expression on beads was measured by staining with anti-GFP antibody and then with fluorescently labeled secondary antibody. 30 min of expression time showed around 50 % of maximum protein loading on beads (indicated by half maximal mean fluorescence intensity) (Figure 4.15 A and B). Following this limited expression time would then allow to differentiate and detect between promoters with a higher and lower activity. Therefore, for the selection T7 promoter variants

expressing GFP-SNAP, expression of proteins was done for 30 minutes in subsequent display experiment.

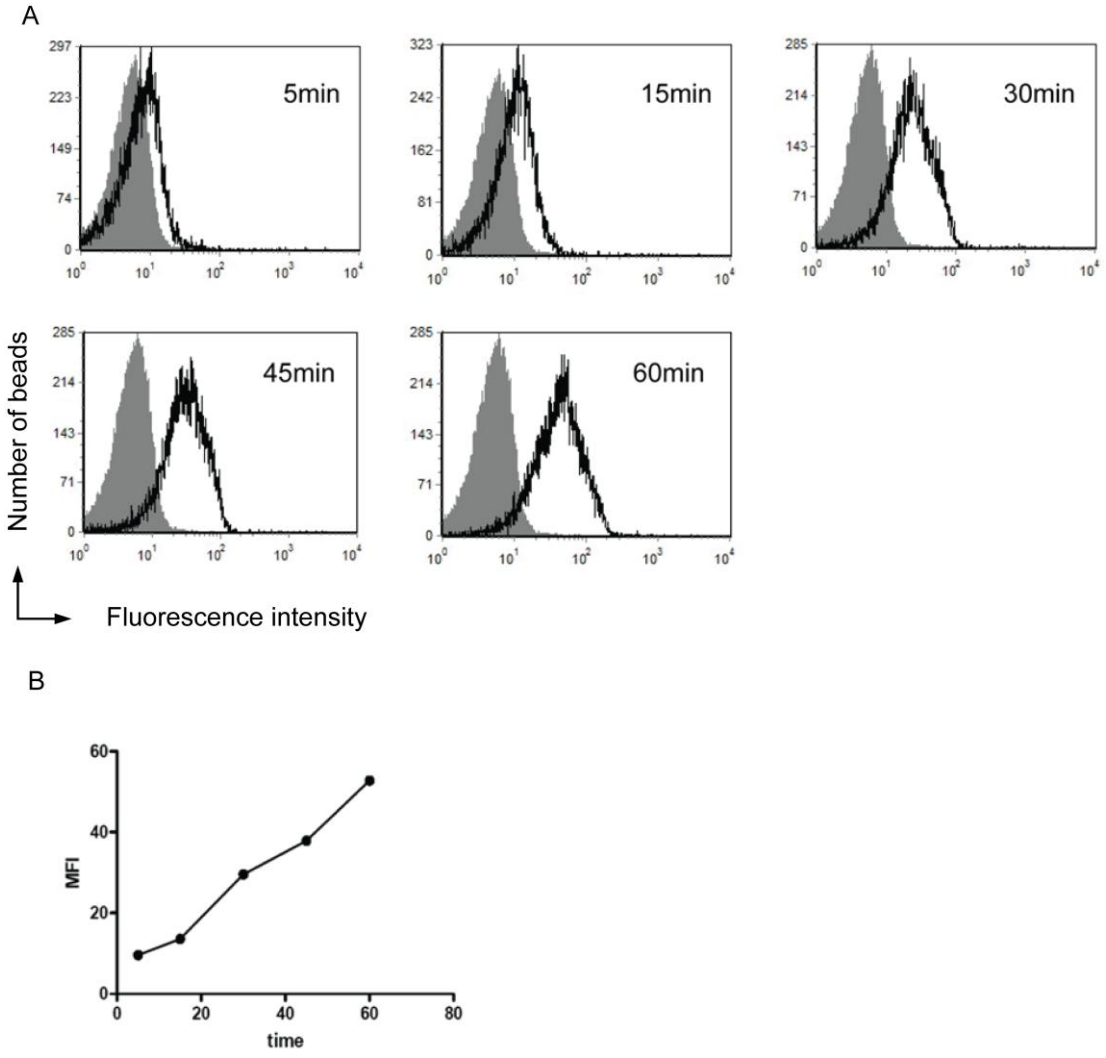



**Figure 4.15 Time dependancy assay for protein loading on beads**. (A) Same sets of beads with gene product (GFP-SNAP) controlled by wildtype T7 promoter after emulsion PCR were expressed in a cell-free expression system in water in oil emulsion. Reactions were stopped at different time points (5, 15, 30, 45 and 60 min). Beads were labeled with anti-GFP antibody and fluorescently conjugated secondary antibody and analyzed in flow cytometry. The level of fluorescence (FL1H) was plotted as histograms. Overlay of histograms of beads with no GFP-SNAP template DNA processed through emulsion PCR and emulsion IVTT (filled grey area) or loaded with GFP-SNAP (black line) are shown. (B) A graph was also plotted with time versus mean fluorescence intensities (MFI) of beads at different time points.

## 4.8.2 Bead display and selection of T7 promoter variants

The multi-copy bead display of the T7 promoter library was performed for three rounds. The entire display and sorting procedure is illustrated in Figure 4.16 for better understanding.

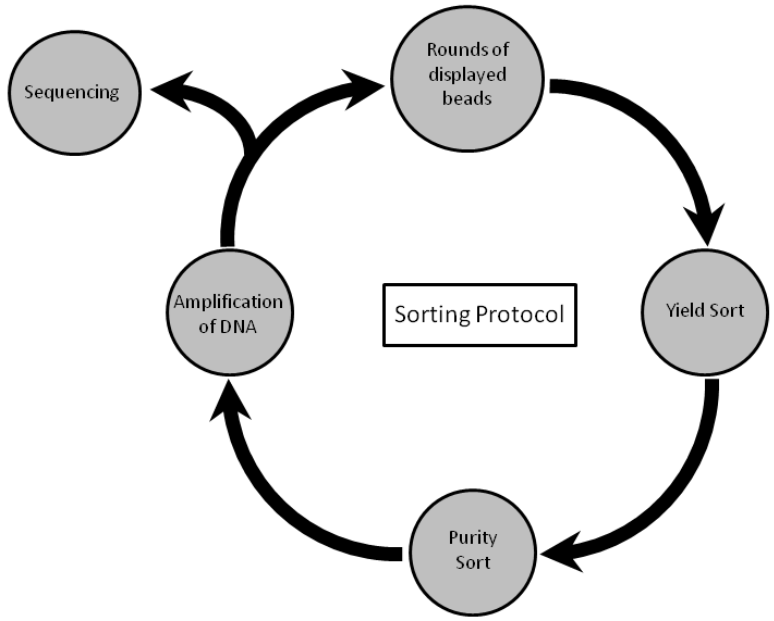



**Figure 4.16 Schematic Diagram depicting sorting and reamplication procedure of desired genotype-phenotype from beads.**

$3\times10^9$ copies of templates containing T7 promoter variants to control GFP-SNAP expression and approximately $10^8$ beads were added for multi-copy bead display. After display, beads carrying GFP-SNAP expression cassette under control of different promoter variants were stained with anti-GFP antibody and fluorescently labeled secondary antibody. Beads were then sorted through flow cytometry. The sorting protocol used for different rounds is shown in figure 4.16. A total of three rounds of multi-copy bead display were performed for T7 promoter variants. To select beads with the highest GFP fluorescence and to avoid selection of false positive beads a two-step flow-cytometric sorting strategy was used for each round of display. During a first yield sort, around 1% of beads with the highest fluorescence intensities were selected. These pre-sorted beads were submitted in a second, slower and more accurate round of sorting (purity sort) selecting again

around 1% of the remaining beads with the highest fluorescence intensity. Fluorescence intensity distribution of the beads during yield and purity sort clearly reveals an increase in mean fluorescence intensity from yield to purity sort (Figure 4.17.). The efficacy of the purity sort was not directly analyzed, since the number of recovered beads was low ($<10^4$). The beads recovered from purity sort were used to amplify the randomized T7 promoter region by PCR. The amplified T7 promoter region was fused by overlap extension PCR to the remaining GFP-SNAP expression cassette regenerating the entire GFP-expression cassette as a new library composed of the selected T7 promoter variants (Figure 3.6, *methods chapter*). This PCR product was used for a second round of multi-copy bead display.

During the subsequent yield sort in the second round a small population of GFP-highly-positive beads is detectable (Figure 4.17). Analyzing the 1% of beads selected in the yield sort during the purity sort demonstrates a further enrichment of highly GFP-positive beads. There is also an increase of mean fluorescence between the two rounds of selection, however, they cannot be compared directly as they were performed on two different days.

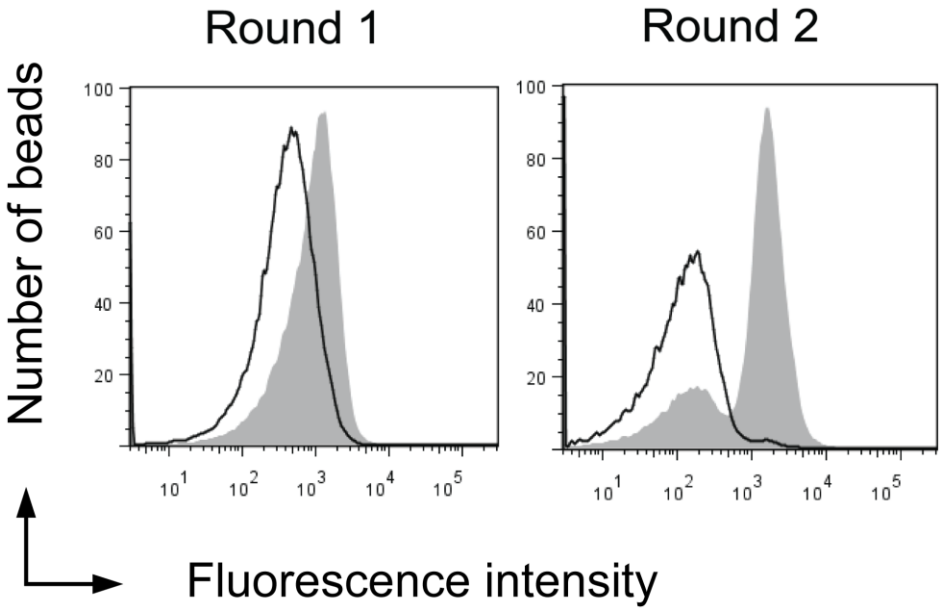


**Figure 4.17** ***(continued)*** **Selection of variants from the T7 promoter library.** Beads were loaded with amplicons from the randomized T7 promoter library upstream of the GFP-SNAP coding region by emulsion PCR, transferred to compartmentalized IVTT reactions and stained for GFP. Overlays of histograms of flow-cytometric analyses during yield (black line) and purity sort (grey area) of the first and second round of selection are shown.

Again, a total of around 5000 beads were sorted for their highest promoter activity. The T7 promoter region of the beads recovered from the purity sort during the second round was amplified as described above and used to reconstitute the entire expression cassette for a third round of multi-copy bead display.

In the third round of display, single high GFP-positive beads obtained were directly sorted into microtiter PCR plates. The T7 promoter region could be successfully amplified from 31 of the 40 single beads selected. The presence of multiple copies of DNA per bead in this system is an advantage when amplifying the DNA directly from a single bead. These 31 T7 promoter variants were subsequently analyzed for their promoter activity in a different assay system.

## 4.9 Characterization of the selected T7 promoter variants

### 4.9.1 Optimization of a luciferase assay system for analyzing promoter variants

Characterization of the selected T7 promoter variants was performed in quantitative assay system. For this, luciferase assay system was employed. To standardize the assay, the luciferase gene under control of the wildtype T7 promoter was expressed in IVTT in solution upto 2 hours. The reaction was stopped at different time points by putting it on ice and subsequently treating it with RNAse to completely remove all mRNA from the system. This stringent step is important for correct quantification of the amount of protein expressed by different promoter variants, so that no residual RNA is left to produce any more proteins after the desired stopping of the reaction. As seen from the figure 4.18 time points between 15 min and 30 min showed the log phase of luciferase

production. Post 30 min, the production of proteins were saturated. For the final characterization of different promoter variants, 15 min of luciferase expression was chosen.

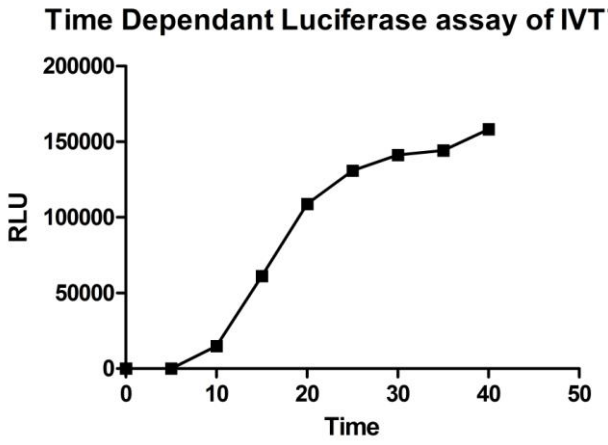



**Figure 4.18 Optimization of Luciferase assay. Linear DNA templates encoding the luciferase gene under control of wildtype T7 promoter were expressed in IVTT in solution.** The reaction was stopped at different time points (0, 5, 10, 15, 20, 25, 30, 35, 40 min) and treated with RNAase and analyzed for luciferase expression.

## 4.9.2 Quantitative characterization of T7 promoter variants through luciferase assay

After standardization of luciferase assay with wildtype T7 promoter, the activities of the selected T7 promoter variants were determined by fusing them upstream of luciferase gene through overlap extension PCR. Protein expression was done as mentioned in the above protocol. They were then analyzed for luciferase activity. More than 50% of the selected promoter variants showed high or higher luciferase activity as compared to wildtype T7 promoter (Figure 4.19). However, there were also some of the selected promoters who had less luciferase activity as compared to the wildtype. This is due to some leak during the sorting protocol of beads. The top six luciferase expressing promoters were then sequenced.

To ensure, that the results do not reflect a random situation of the original library, also 40 clones were randomly picked from T7 promoters of the original library

None of the clones had an activity close to the one of the wild type promoter. Around 20% of the wild type promoter´s activity was detected in 4 of the random clones (Figure 4.20).

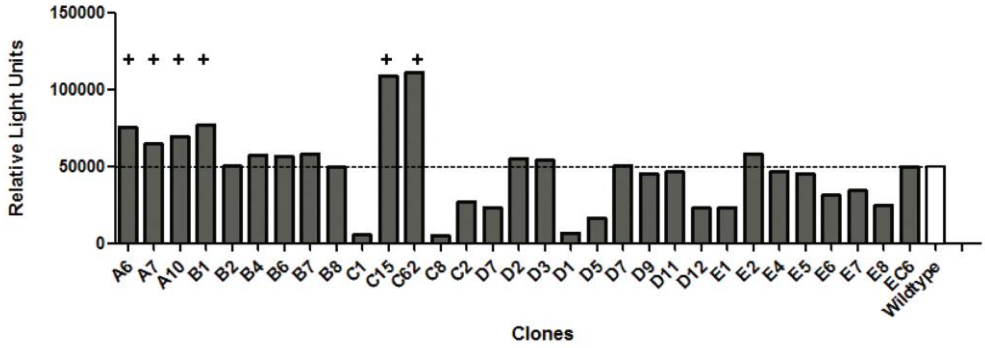



**Figure 4.19 Protein expression activity of promoter variants selected from the T7 promoter library by multi-copy bead display.** The luciferase reporter gene activity of 31 T7 promoter variants obtained from selected single beads is shown in comparison to the wild type T7 promoter.

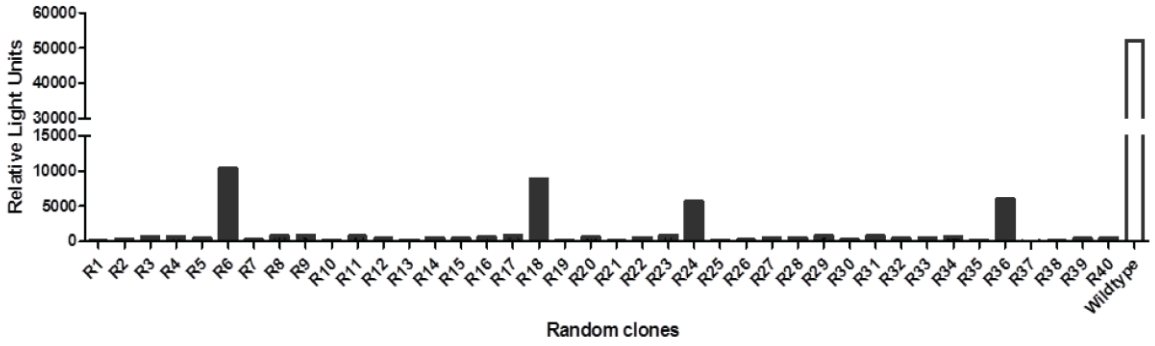



**Figure 4.20 Protein expression activity of promoter variants of randomly picked clones from original library.** The luciferase reporter gene activity of 40 T7 promoter variants obtained from original library without any selection is shown in comparison to the wild type T7 promoter.

## 4.9.3 Sequence analysis of promoter variants

The six T7 promoters with the highest activity were sequenced directly from amplicons. Since two sequences were represented by two independently selected beads, a total of four different T7 promoter sequences were obtained with a higher activity than the wild type promoter sequence (Table 4.2). The only nucleotide of

the randomized region conserved between the wild type T7 promoter and all four selected sequences is the guanine at position +1 (Table 4.2 ).

As described before, in addition to sequencing these 6 selected promoter variants, six random promoter sequences were also sequenced from the library which did not undergo multi-copy bead display. They are completely random without any similarity to wild type T7 promoter. They are also shown in the table 4.3.

**Table 4.2** Sequences of slected T7 promoter variants

| Variants | Sequence | % activity relative to wild type T7 promoter* in terms of protein expression |
|---|---|---|
| | -15 -10 -5 +1 +5 | |
| Wildtype | TAATACGACTCAC**TATAGGGAGA** | 100 |
| A6, B1 | TAATACGACTCAC**TCCGGCAATC** | 151 |
| A7 | TAATACGACTCAC**TTCGGCAACC** | 118 |
| A10 | TAATACGACTCAC**ACGAGCGGCA** | 138 |
| C15, C62 | TAATACGACTCAC**AATCGCGGAG** | 193 |

Nucleotides from randomized library are in bold.
Conserved residue is inverted in color.
* Mean values from two independent experiments

**Table 4.3** Sequences of random variants from unselected T7 promoter library

| Variants | Sequence | % activity relative to wild type T7 promoter* in terms of protein expression |
|---|---|---|
| | -15 -10 -5 +1 +5 | |
| Wildtype | TAATACGACTCAC**TATAGGGAGA** | 100 |
| R6 | TAATACGACTCAC**TTTGAGGGTG** | 20 |
| R18 | TAATACGACTCAC**TAGAAACTTA** | 17 |
| R19 | TAATACGACTCAC**CTTACTACCA** | 0.2 |
| R36 | TAATACGACTCAC**GGTTTAGACT** | 0.25 |
| R24 | TAATACGACTCAC**TGCATTATGT** | 11 |

Nucleotides from randomized library are in bold.
*Mean values from two independent experiments

## 4.10 Characterization of the C62 T7 promoter variant

### 4.10.1 Comparison of protein expression kinetics of C62 and wildtype T7 promoter

To further characterize the T7 promoter variant, which had shown the highest expression efficiency (C62), luciferase under control of C62 was expressed for different time periods (15 min, 30 min, 45 min, 60 min, 90 min, and 120 min) and compared to expression under control of wildtype T7 promoter. C62 showed an overall faster luciferase expression where expression levels at each time point were at least two fold higher than those for wild type T7 promoter (Figure 4.20). Luciferase levels equal to a 1 hour wild type T7 promoter activity were obtained after 15 min of expression for the C62 variant and total luciferase amounts after 2 hours were two fold higher than those for the wild type T7 promoter. Taken together, the variant promoter showed two fold higher efficiency in terms of speed of expression as well as total amount of expressed protein after 2 hours. This experiment was performed twice at two different days.

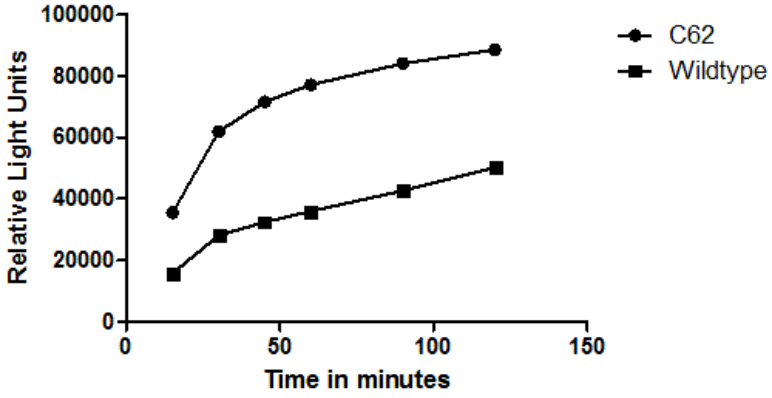



**Figure 4.21 Promoter activity on protein expression**. Time dependant (15 min, 30 min, 45 min, 60 min, 90 min, and 120 min) luciferase activity in IVTT reactions of the luciferase gene driven by the C62 or wild type T7 promoter.

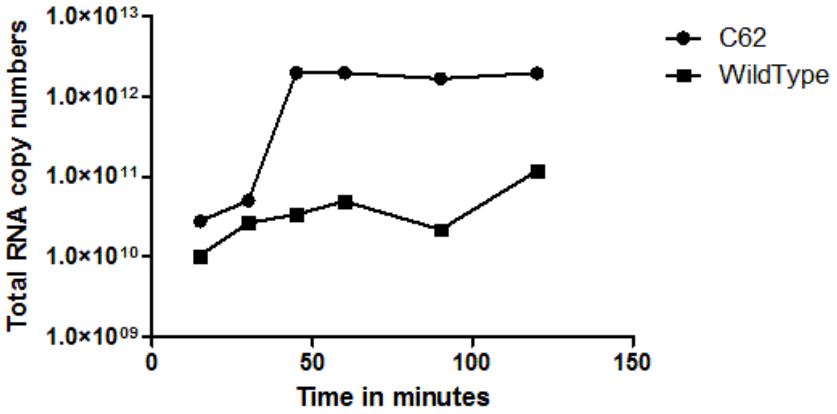
### 4.10.2 Effect of promoter variation on transcription and translation

Although only the transcription initiation site of the T7 promoter was randomized it was unlikely but not impossible, that modifications affected not only the transcription activity but also the translation efficiency. Therefore both activities were analyzed separately. For the transcription efficiency analysis, we decided to stop transcription at different time points (up to 2 hours), isolate the RNA and quantify RNA copy numbers by quantitative RT-PCR. This experiment was performed twice separately at different days. Maximum numbers of transcripts were more than tenfold higher for C62 than for wild type T7 promoter and were obtained within 45 minutes for C62, whereas wild type T7 promoter seems not to have reached the maximum copy numbers even after 2 hours (Figure 4.21 A). Taken together, wild type T7 promoter appeared to be much slower in RNA production and RNA copy numbers were tenfold less after 2 hours of reaction time compared to the C62 promoter variant.

To analyze, if the promoter also affects translation efficiency, a defined copy numbers of *in vitro* transcribed expression cassettes (RNA) was introduced to the cell free expression system without any DNA. This is done to simulate, that same

amount of RNA were transcribed ready for translation. At different time points, translation was stopped by adding RNase and luciferase expression levels were analyzed as described above. No differences in amounts of protein expression between wild type and C62 T7 promoter were observed (Figure 4.21 B) and both constructs showed saturation levels of protein production at 45 min. This experiment was also repeated twice at two different days.

A




B

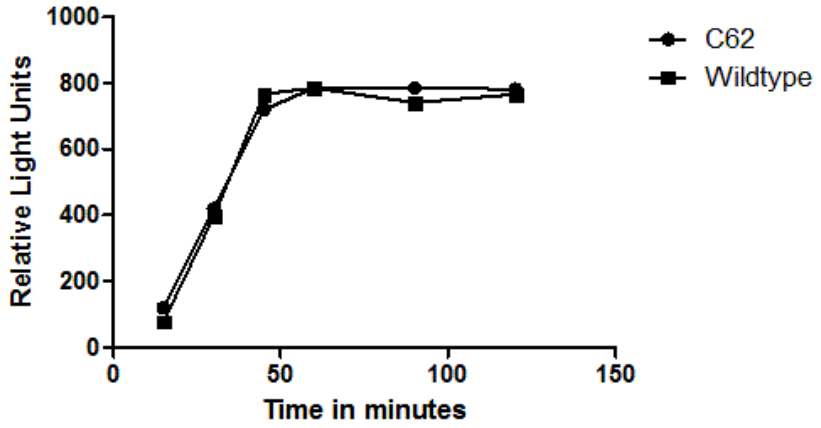



**Figure 4.22 Effect of promoters on transcription and translation**. (A) Time course of transcribed RNA levels in *in vitro* transcription reactions of the luciferase gene driven by the C62 or wild type T7 promoter. (B) Time course of luciferase activity of *in vitro* translation reactions of equal amounts of luciferase mRNA containing either the first transcribed nucleotides of the wild type or the C62 T7 promoter. Results of one representative experiment out of 2 independent experiments are shown.

### 4.10.3 Mutation studies on the new T7 promoter variant C62

C62 T7 promoter variant had a completely novel sequence in the transcription initiation site compared to wild type T7 promoter as seen from table 4.2. To understand, if any change of residue can affect promoter activity, it was decided to mutate one residue at a time for all 10 nucleotides in the transcription initiation site of C62. A was mutated to T, T was mutated to A, G was mutated to C and C was mutated to G. This led to 10 different promoter variants which required to be analyzed. The variants were fused upstream of the luciferase gene through overlap extension PCR as described above. Transcription efficiencies were analyzed through RNA production up to 40 min in IVTT and the amount of transcripts was quantified through real time PCR and calculated as percent activity relative to the activity of C62. The result of this analysis is documented in table 4.4.

As seen from the table, the mutations in all positions led to drastic reduction of promoter activity of C62 except for position +2 and +3. These positions had a G and C respectively in C62 which were replaced with a C and G and showed 44% for +2 (mut 6C) and 80% for +3 (mut 7G) activity relative to C 62. Compared to other positions which showed the importance of conserving residue respective for that position, the residues at +2 and +3 appeared less important i.e replacing residues in these positions has drastic effect on promoter activity of C62. These residues are the part of transcription start site residues along with +1 position of the T7 promoter.

It was then decided to further analyze position +2 and +3 by replacing with other residues. These new variants (mut 6A, mut 6T, mut 7A and mut 7T) were analyzed for their promoter activities and compared with earlier mutations from the experiment described above (Table 4.4). It was expected that position +2 and +3 are less important for promoter activity of C62 and any replacement will still show considerable amount of promoter activity. However, these new mutants had completely destroyed the promoter activity of C62 (shown in table 4.4).

**Table 4.4** Mutation studies

| | Sequence | | | | | | | | | | | | | | | | | | | | | | % Activity * |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | -1 | | | | | -10 | | | | | | -5 | | | | +1 | | | | | | +5 | |
| C62 | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **G** | **G** | **A** | **G** | **100** |
| mut1T | T | A | A | T | A | C | G | A | C | T | C | A | C | **T** | **A** | **T** | **C** | **G** | **C** | **G** | **G** | **A** | **G** | **1.52** |
| mut2T | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **T** | **T** | **C** | **G** | **C** | **G** | **G** | **A** | **G** | **0.58** |
| mut3A | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **A** | **C** | **G** | **C** | **G** | **G** | **A** | **G** | **0.48** |
| mut4G | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **G** | **G** | **C** | **G** | **G** | **A** | **G** | **0.27** |
| mut5C | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **C** | **C** | **G** | **G** | **A** | **G** | **0.69** |
| mut6G | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **G** | **G** | **G** | **A** | **G** | **44.8** |
| mut6A | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **A** | **G** | **G** | **A** | **G** | **8.08** |
| mut6T | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **T** | **G** | **G** | **A** | **G** | **1.65** |
| mut7C | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **C** | **G** | **A** | **G** | **87.7** |
| mut7A | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **A** | **G** | **A** | **G** | **1.13** |
| mut7T | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **T** | **G** | **A** | **G** | **0.83** |
| mut8C | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **G** | **C** | **A** | **G** | **0.63** |
| mut9T | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **G** | **G** | **T** | **G** | **0.47** |
| mut10C | T | A | A | T | A | C | G | A | C | T | C | A | C | **A** | **A** | **T** | **C** | **G** | **C** | **G** | **G** | **A** | **C** | **0.29** |

Nucleotides from transcription initiation site are in bold.

Mutated residues are inverted in color.

* % Activity relative to C 62 (mean values from two independent experiments) in terms of RNA synthesis

## 4.10.4 Protein expression kinetics of C62 mutants

All 14 different C62 mutants from the mutation study were also analyzed for their ability to express protein. Luciferase gene under the control of these C62 variants was expressed in IVTT for 15 min as described before (p.87, 4.9.1) and analyzed for luciferase expression. As shown in Figure 4.21, most of the mutants showed almost no protein expression. However, mut 6G and mut 7C showed around 50% of protein expression capability as compared to C62

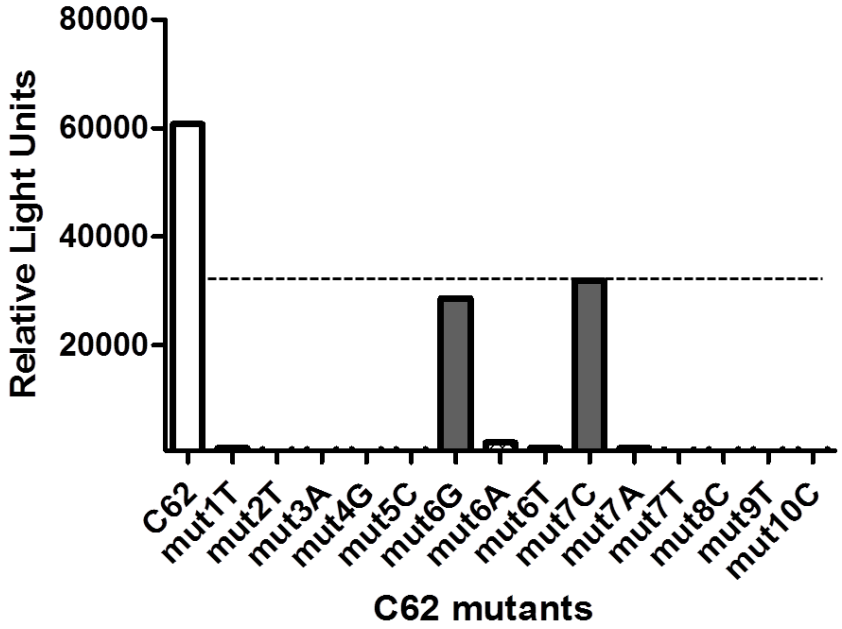



**Figure 4.23 Protein expression kinetics of C62 mutants.** The luciferase reporter gene activity of 14 C62 mutants is shown in comparison to the C62.

# 5. Discussion

Emulsion based compartmentalization introduced new scopes for display of proteins and evolution in a much faster and economical way. In this context, different novel strategies have been developed for making the process general in mode of selection other than enzymatic or DNA based substrate. One such attempt was done by Griffith *et al* [113], where they modified the binding strategy by enzyme substrate interaction on a microbead and selection through flow cytometry. Flow cytometric based screening is highly sensitive and could be applied to screening of wide range of proteins. Using this strategy they could evolve a faster phosphodiesterase enzyme. However, still the mode of selection remained dependent on some catalytic property of the enzyme and so possible applications are again strongly limited. There is a need for a system, which is more general in its purpose, which would evolve other proteins beyond enzymes and in which screening.is independent from the *in vitro* transcription translation reaction. All DNA display even those which do not involve IVC systems described [24, 101, 103, 104, 108, 113–115, 118, 120] all lack a generic suitability for library expression and selection.

The aim of this work was to develop a novel bead display approach for display and evolution of proteins in a more generic way in *in vitro* compartments by (i) introducing a universal system for covalent genotype-phenotype linking (SNAP), (ii) covalent linking of genotype to a solid phase such as. paramagnetic microbeads for maintaining monoclonality of the system and (iii) the beads serve as the unit of selection which contains multiple copies of the DNA stably linked to multiple molecules of the encoded protein and thereby would provide great stability and robustness to the system. This would permit both, quantitative and qualitative

selection criteria making this system universal in its applicability. This chapter discusses three main aspects of this thesis-(i) how the multi-copy bead display approach was realized, (ii) screening and selection of T7 promoter library, and (iii) the future scope of multi-copy bead display approach.

## 5.1 Multi-copy bead display system

The multi-copy beads display approach (Figure 1.9), is an emulsion based compartmentalized system in which $10^{10}$-$10^{11}$ picoliter reactors were generated where multiple copies of DNA and proteins are simultaneously linked to a solid surface, which would then be used for screening of libraries with very high complexity. Each of these compartments acts as a reaction center containing its own reagents and approximately one copy of DNA for display and evolution. Assuming there are $10^{10}$ picoliter reactors , $3x10^{9}$ template DNA molecules were added which results in a probability of having two template DNA molecules in the same picoliter reactor of 0,09 ($=0.3^2$). This ensures that the majority of reactors do not contain more than one DNA copy. As pertinent with earlier reports and the initial experiments performed, it showed that these reactors were stable enough to withstand longer PCR cycling especially the emulsions which were prepared with Abil EM 90 (Figure 4.2 B). The emulsions were prepared at different rotation speed. The best emulsion preparation was achieved at a speed of 2000 rpm, in which most of the reactors were in the average size of 2-5 micrometer (Figure 4.3) and most of them contained one single bead (Figure 4.2 A II).

The standardization of multicopy bead display system was divided in to two main parts. In the first part, emPCR on beads was optimized and in the second part, the subsequent expression and capture of proteins in emulsion was optimized for flow cytometry based screening and selection. Although, emPCR without beads provided optimal results both in terms of amount of amplicons generated in comparision to a normal PCR in solution and also its robustness at longer PCR cycles to amplify specific gene product from a mixture of gene products preventing

any kind of size competition (Figure 4.4 A and B), this protocol however, failed to amplify and capture a sufficient amount of gene products in emPCR on a solid phase i.e. beads. Even though, these beads coupled with sense primers demonstrated saturated coupling efficiency (figure 4.5) but through the initial emPCR protocol followed to amplify and capture of DNA on them, these beads they were only able to capture around 300 copies of PCR products. The kinetics of PCR in emulsion with beads tends to be slower as compared to normal emulsion PCR, mainly due to the steric hindrance of primers coupled to the beads. So, these emPCR required different optimizations. Moreover, because of only one single template molecule present in a picoliter reactor makes it more difficult to amplify efficiently on beads. A number of standardization steps were inducted in the emPCR on beads. A final concentration of 0.01 μM of soluble sense primer was introduced to boost initial cycles of PCR, so that even at low numbers of template molecules the PCR is efficient thereby counteracting to a certain extent any possible stress due to steric hindrance. However, a balance is required for adding soluble primer, because at higher concentration, soluble primers tend to compete with the primers bound to beads and result in most of the PCR product in solution and not on beads (Figure 4.8). The reverse primer is chemically coupled to BG. This is involved in the creation of genotype-phenotype linkage through BG-SNAP covalent interaction. This primer introduces BG to the PCR products during emPCR. The amplicons labeled with BG is translated to its encoded proteins, which are then captured to their encoding DNA through this covalent linkage (Figure 1.9). This BG coupled antisense primer requires being functionally active for the success of the multi-copy bead display approach (Figure 4.7). The amount of this reverse primer in emPCR was also empirically optimized. A final concentration of 3 μM of this primer was included in the final protocol of emPCR (figure 4.9). The amount of liquid phase in these reactors is in the range of femtoliters. There is a high chance that during extended PCR cycling process, most of the reagents in these compartments are exhausted. Along with the changes in the

primer concentrations as discussed above, dNTPs, bovine serum albumin (BSA) and Taq polymerase were substantially increased for optimized solid phase emPCR. The number of amplicons captured on the beads could be substantially increased to 800-1000 copies per bead. It is also an important development that this optimized solid phase emPCR could efficiently amplify PCR product of the size of around 1.6 kbp in contrast to other published reports on solid phase PCR which could only amplify amplicons of a maximum size of 400bp [114,115]

In the second stage of standardization of multi-copy bead display approach, the functionally active proteins were expressed and subsequently captured on beads for screening. The model expression cassette GFP-SNAP under the control of the T7 promoter was first expressed in a number of cell free expression system. However, only a rat reticulocyte expression system specifically for PCR templates could expresses both the fusion partners of the proteins in its functionality both in solution and in emulsion from PCR products as templates in contrast to other cell free expression systems which only work with circular templates (Figure 4.10). Circular plasmids tend to be more stable than linear templates and thereby special expression systems might be required for templates with PCR products. The IVTT expressed GFP-SNAP fusion protein was also able to bind to BG through its SNAP partnerwhen incubated with BG agarose beads (Figure 4.11) and they could also fluoresce through its GFP partner when observed in fluorescence microscope. The BG-SNAP interaction was also highly specific (Figure 4.11). Flow cytometric analysis of the beads post coupling of DNA and encoded proteins to it through emPCR and emIVTT initially failed to give to any signal for GFP fluorescence in contrast to the experiments with BG-labeled agarose beads. One of the possible reasons might be due to the fact that BG-labeled agarose beads are 70 times larger and containBG residues directly coupled to the bead as compared to the magnetic beads used for multi-copy beads display. Also the number of BG moieties of these beads is too low (around 1000 BG moiities from 1000 amplicons). This suggests that emPCR on beads was still not efficient enough and might be some products

are not extended fully. Thereforem the number of BG residues on the beads had to be increased for efficient capture of proteins thereafter. Introducing a simple step of oligohybridisation through BG oligonucleotide which is antisense to the coupled sense primer on the beads after emPCR in the multi-copy bead display approach was able to additionally increase the BG residues on the beads. The system was further made independent of GFP fluorescence through antibody staining (Figure 4.12). This increases the potentiality of the multi-copy bead display approach to be multiparametric in flow cytometric based screening and selection. It means selection criteria can be combined through Boolean operations. For example a peptide sequence can be selected for its binding to antibody A AND NOT antibody B. thereby more precise selection of libraries can be accomplished. The proof of the principle experiment clearly demonstrates that multi-copy bead display approach is monoclonal and sensitive. Mixing of templates encoding either a binding (GFP-SNAP) or a non-binding (MS2-SNAP) partner of the ligand clearly revealed two different populations of the beads (Figure 4.13). This supports the fact that flow cytometric screen can be efficiently applied to identify protein variants with increased binding affinity to a fluorescently labeled ligand. The ability to detect GFP at a very low amount shows, that this system is highly sensitive and robust. Moreover, this system is also able to clearly discriminate a mixed population from a homogenous population of only binding or non-binding ligands (Figure 4.14). This could be highly interesting especially when screening complex molecules like an antibody from a library. This system would be able to differentiate populations of high, medium and low binders and thereby effectively can be applied to study affinity maturation of antibodies where the low and the mdium binders can be further manipulated by error prone PCR or chain shuffling to further mature in subsequent selection steps. The proof of principle experiment also suggests that in addition to qualitative selection criteria also quantitative selection criteria can be applied through this approach. Calculating the number of amplicons post multi-copy bead display (from the competition assay, figure 4.14)

and the real time PCR analysis post emPCR showed that both calculations are consistent with each other. This system also makes it feasible to transfer the protein and DNA loaded beads recovered from the IVTT emulsion to cell-like microcapsules for independent enzymatic screening reactions and subsequent microfluidic selection methods [150] extending the potential applicability of the multi-copy bead display to all kinds of enzymatic optimizations in the field of synthetic biology. Thus, the combination of emPCR and IVTT in emulsion in multicopy bead display approach increases the scope of DNA display systems in diverse applications beyond enzymes and DNA binding proteins.

## 5.2 *In vitro* evolution of a novel T7 promoter variant

All the display strategies till date are employed to evolve protein through qualitative selection criteria based on the property of the protein. This is also probable with multi-copy bead display approach. Apart from this, quantitative selection criteria are also possible with multi-copy bead display and therefore it can be applied to evolve certain functional nucleic acids which have a role in biological organization such as a promoter which controls the expression of a gene. As the first application of multi-copy bead display approach it was chosen to evolve a T7 promoter variant with higher promoter activity in IVTT compared to the wild type T7 promoter. This would prove that multi-copy bead display approach is more generic in its application and could even be applied to the evolution of nucleic acids in contrast to other IVC display strategies. The 10 nucleotides at the transcription initiation site of wild type T7was completely randomized and this gave a library of T7 promoter with a complexity of $10^6$. These promoter variants introduced upstream of GFP-SNAP fusion expression were evaluated for their promoter activity through the expression of the gene under their control. This promoter variant library underwent multi-copy bead display approach. Under the experimental conditions described, the $10^8$ beads used in the multi-copy bead display approach limit the maximum number of variants that can be screened. But, it was enough for a library of this complexity. The number of emulsions generally ranges from $10^{10}$-

$10^{11}$; a ten- to hundred-fold increase of the number of beads and simple upscaling of the total volumes would allow screening of libraries with larger diversity. The screening and selection of these promoter variants with highest promoter activities were taken relative to the expression of GFP-SNAP fusion protein in flow cytometry. Flow cytometric based screening and selection proved to be successful for re-iterative rounds of multicopy bead display approach.

Although a small population of GFP positive beads was collected in each round of display, there was a definitive enrichment of the GFP positive population between different rounds of screening and selection (Figure 4.17). After the final round of multi-copy bead display, single beads were sorted to amplify the DNA sequence from them. This is another advantage of multicopy bead display approach. This is possible in multi-copy bead display approach because of the presence of multiple copies of DNA on each bead. This DNA can be then directly utilized for further characterization. In contrast to other DNA display systems where a population of beads is required to be collected in order to amplify the DNA from them because of the presence of single DNA molecule each of them. Again to isolate single sequence one has to then clone these sequences in bacteria. However, it can lead the problem of bottleneck which arises through inefficient transformation in bacteria (as reviewed in [151]) Absence of any bacterial cloning step makes this system robust and versatile.

The promoter variants (a total of 40 in number)) obtained through multi-copy bead display approach, were evaluated for their promoter activity. Generally, to evaluate a promoter through its ability to express proteins under its control, a quantitative assay system is preferred. Luciferase assay is a perfect quantitaive assay system to determine the promoter activity of different selected promoter variants. More than half of the selected clones were able to express more or equivalent luciferase relative to luciferase expressed under the control of wild type T7 promoter (Figure 4.19). Only 4 out of randomly picked 40 clones from the original library that did not underwent multi-copy bead display showed only 20 % luciferase activity

relative to wild type T7 promoter which proves the fact that high activity T7 promoters variants selected through multi-copy bead display suggests that it is not a random effect but due to the stringent selection procedure.

DNA sequence analysis of the promoters with highest activity had completely novel sequences at the transcription initiation sites compared to wild type T7 promoter (Table 4.2). Repetition of sequences in the selected clones (A6, B1 and C15, C62) indicated that there is already an accumulation and enrichment of highly positive sequences through three rounds of multi-copy bead display and subsequent round of multi-copy bead display was not necessary. Interestingly guanine (G) at the transcription start site +1, was conserved in all the evolved promoter variants like the wild type T7 promoter confirming the importance of G at this position which is generally involved in the stabilization of the transcriptional unit along with other residues at +2 and +3 during transcription[152]. The sequences obtained through multi-copy bead display approach were unambiguous in nature which once again proved that clonality of the system is maintained at sufficient level at all stages of the display procedure.

Further analysis of C62, one of the T7 promoter variants, screened and selected through multi-copy bead display showed almost two fold protein production capacity compared to wild type T7 promoter (Figure 4.20). Also, the protein expression is faster with C62 as wild type failed to catch up even with extended time of expression. C62 effects the level of transcription by production of 10 fold higher RNA than wild type T7 promoter in IVTT (Figure 4.21 A) which had led to this high protein production. However, it did not show any effect on translation level when RNA transcribed under the control C62 and T7 promoter respectively and similar amounts of them were added in IVTT for the expression of proteins (Figure 4.21 B). Therefore, C62 promoter variant has a definitive role in transcription which in turn affects protein production validating its role as better promoter than wild type T7 promoter.

*In vitro* evolution allows a researcher to evolve novel molecules with improved physical or chemical properties for a defined function without the necessity to answer the question, why these novel molecules are better compared to already standard biomolecules for the same function. There might be more variants of C62, which have an equal or better promoter activity compared to C62. To assess this, all the 10 residues of C62 from +1 to +10 were further mutated which resulted in 10 new variants of C62. Most of the 10 variants showed no or minimal RNA expression activities. Only two mutants (mut 6G and mut 7C) tested, showed remarkable transcriptional activities of around 50 % and 80 % compared to C62. (Table 4.4). These mutants had an exchange at position +2 and +3 in which G was mutated to C and C mutated to G, respectively. Further mutating these positions with A or T resulted again in a complete drop of promoter activity. These results suggest that G or C is preferred at these positions along with a G at +1 for promoter activities. These findings resonates with an earlier published report which shows the importance of these residues in wild type T7 promoter [152] for T7 polymerase to transcribe might also hold true for C62 promoter variant. These mutation studies throw a light on the fact that certain specific arrangement of residues at different positions is important for the promoter activity. This is confirmed by the fact, that the C62 promoter variant found here has no similarities with wild type T7 promoter, nevertheless is highly active.

In parallel to RNA expression activities under these different C62 mutants, luciferase protein expression studies under the contol of these mutants too showed, that most of the mutants of C62 had failed to produce any luciferase activity (Figure 4.21) except for mut 6G and mut 7C which produce only 50% of total amount of protein relative to C62.. It is interesting to observe that mut 7C, which had around 80 % of RNA expression activity relative to C62 showed around 50% protein expression relative to C62. This proves once again that selection of C62 in bead display approach was not a random process; the selection criteria was set on protein expression only and not on RNA expression (figure 1.9).

C62 promoter variant was evolved through a coupled *in vitro* transcription translation system which is a eukaryotic system and works with T7 promoter. This promoter variant was able to show high level of *in vitro* activity both in terms of RNA production and subsequent protein production in a rat reticulocyte system. However, this promoter variant was not evaluated side by side analysis in different cell free expression systems such as wheat germ, *E.coli* S 30 systems etc. Nevertheless this novel C62 evolved through multi-copy bead display is functional to carry out its promoter activity and it also turned out to be a better promoter than wild type T7 promoter. It can further be used to optimize *in vitro* RNA synthesis protocols through a strong reduction of costs in a variety of biotechnological and therapeutic applications.

To finally summarize, multi-bead display approach combines a number of advantages and would provide broad applicability and robustness as demonstrated with the evolution of a novel T7 promoter variant. It was also demonstrated that multiple copies of DNA on a solid phase (microbead) increases the sensitivity of the system and also easy amplification of DNA without the requirement of any bacterial cloning step. Moreover, the use of microbeads can make the system monoclonal till the end of display and thereafter. Also, the coupling of the DNA and the encoded protein on beads allows easy removal of IVTT reagents by a simple magnetic separation step. Thus, the selection conditions can be chosen independent from the IVTT reaction providing greater flexibility and avoidance of high background signals. In contrast to the DNA display approach, encoded proteins not coupled to DNA can also be removed avoiding competition and interference during the selection step. Since multiple molecules of the same protein are concentrated on a single bead, the signal used for the selection step is also strong enough for flow cytometric screening. This is advantageous since it is possible to precisely define what percentage of the library is selected during each round. These advantages would definitely make multi-copy bead display an alternative display platform for *in vitro* evolution of proteins and nucleic acids.

## 5.3 Scope of Multicopy Bead Display Approach

Being generic in nature, a wide number of applications could benefit from multi-copy bead display approach. The evolution of novel promoter variant as demonstrated in this work was the first application of this display approach. Another possible application of it, currently being addressed in the lab is screening and selection of a novel DNA library that encodes for a combinatorial mimotope library. A mimotope is a protein molecule which mimics the structure of an epitope. This mimotope library will be then screened and selected with a broadly neutralizing anti- HIV antibody through flow cytometry. With this application, the versatility of multi-copy bead display also capable of in vitro evolution of proteins would be established. These mimotopes will lie in scaffolds of other protein. The use of scaffold would provide stability to these mimotopes, helping them to fold naturally as in cells. This novel mimotopes would be then characterized for their interaction with neutralizing antibody and used for development of possible neutralization antibody to HIV.

Generally yeast two hybrid system has been exploited for the isolation of novel libraries with scaffold protein. Despite its advent, it has not been widely exploited because of very low success rate mainly due to a large number of false positives (almost 70%) owing to failure in designing a good characterization strategy [153]. And also utilizing a living system like yeast, which is dependent on cellular expression pathways and the replication capacity of the respective units of selection, can make this system less reliable and more laborious as discussed in the *introduction chapter*. Multi-copy bead display has the characteristics to again address these concerns and be aptly applied to evolution of novel peptide in scaffold.

Emulsion PCR is an efficient system to amplify random DNA libraries for aptamer libraries[154] . Multi-copy bead display can further take screening and selection of different aptamers to the next level due to its high level of precision and robustness. With a slight modification of multi-copy bead display approach it can also be directly applied to direct characterization of different DNA or RNA

aptamers for target binding in emulsion. This can lead to selection of multiple copies of single sequence of aptamers without the need to clone them in E coli making the process less time consuming.

It would be very interesting to apply this approach to display and evolve an antibody library. Until today, most of the exclusive *in vitro* display systems such as ribosome or mRNA were limited to mostly single chain antibody (ScFv) display, one study however with DNA display system could successfully display Fab (fragment antigen-binding) heterodimeric antibody fragments[155]. However, the recovery rate was less than 1 %, which would influence the effective library size as pointed out by the authors. This could be overcome by multi-copy bead display, which includes emulsion PCR and thereby would be able to handle larger DNA libraries. Multi-copy bead display approach could be effectively applied to display of both ScFvs and Fabs. Moreover, single beads population can effectively be screened by flow cytometry without the requirement of biopanning and the sequence can directly amplified from a single bead without the need to clone them.

With all the advantages discussed in favor of multi-copy bead display, it can also be further applied to evolution of novel functional enzymes. However, different enzymatic assays required to be developed to further expand the potential of multi-copy bead display approach apart from flow cyometry based screening.

# Bibliography


1 Joyce, G. F. (2007) Forty years of in vitro evolution. *Angewandte Chemie (International ed. in English)*. 46, 6420–36

2 HARUNA, I. *et al.* (1963) AN RNA "REPLICASE" INDUCED BY AND SELECTIVE FOR A VIRAL RNA: ISOLATION AND PROPERTIES. *Proceedings of the National Academy of Sciences of the United States of America*. 50, 905–11

3 Haruna, I. and Spiegelman, S. (1965) Specific template requirments of RNA replicases. *Proceedings of the National Academy of Sciences of the United States of America*. 54, 579–87

4 Ellington, A. D. and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*. 346, 818–22

5 Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*. 249, 505–10

6 McGinness, K. E. and Joyce, G. F. (2003) In search of an RNA replicase ribozyme. *Chemistry & biology*. 10, 5–14

7 Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Critical reviews in biochemistry and molecular biology*. 39, 99–123

8 Yarus, M. (1999) Boundaries for an RNA world. *Current opinion in chemical biology*. 3, 260–7

9 Aharoni, A. *et al.* (2005) High-throughput screens and selections of enzyme-encoding genes. *Current opinion in chemical biology*. 9, 210–6

10 Waldo, G. S. (2003) Genetic screens and directed evolution for protein solubility. *Current opinion in chemical biology*. 7, 33–8

11 Matsuura, T. *et al.* (2004) Combinatorial approaches to novel proteins. *Chembiochem : a European journal of chemical biology*. 5, 177–82

12 Matsuura, T. and Yomo, T. (2006) In vitro evolution of proteins. *Journal of bioscience and bioengineering*. 101, 449–56

13 Ni, J. *et al.* (2002) Conversion of a typical catalase from Bacillus sp. TE124 to a catalase-peroxidase by directed evolution. *Journal of bioscience and bioengineering*. 93, 31–6

14 Ju, J. *et al.* (2005) Directed evolution of bacterial alanine racemases with higher expression level. *Journal of bioscience and bioengineering*. 100, 246–54

15 Nakamura, A. *et al.* (2005) In vivo directed evolution for thermostabilization of Escherichia coli hygromycin B phosphotransferase and the use of the gene as a selection marker in the host-vector system of Thermus thermophilus. *Journal of bioscience and bioengineering*. 100, 158–63

16 Sidhu, S. S. (2000) Phage display in pharmaceutical biotechnology. *Current opinion in biotechnology*. 11, 610–6

17 Ogino, H. and Ishikawa, H. (2001) Enzymes which are stable in the presence of organic solvents. *Journal of bioscience and bioengineering*. 91, 109–16

18 Voigt, C. A. *et al.* (2000) Rational evolutionary design: the theory of in vitro protein evolution. *Advances in protein chemistry*. 55, 79–160

19 Smith, G. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*. 228, 1315–1317

20 Georgiou, G. *et al.* (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature biotechnology*. 15, 29–34

21 Boder, E. T. and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology*. 15, 553–7

22 Hanes, J. and Plückthun, a (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*. 94, 4937–42

23 Takahashi, T. T. *et al.* (2003) mRNA display: ligand discovery, interaction analysis and beyond. *Trends in biochemical sciences*. 28, 159–65

24 Tawfik, D. S. and Griffiths, A. D. (1998) Man-made cell-like compartments for molecular evolution. *Nature biotechnology*. 16, 652–6

25 Gommans, W. M. *et al.* (2005) Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command. *Journal of molecular biology*. 354, 507–19

26 Hufton, S. E. *et al.* (1999) Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *Journal of immunological methods*. 231, 39–51

27 Bratkovic, T. *et al.* (2005) Affinity selection to papain yields potent peptide inhibitors of cathepsins L, B, H, and K. *Biochemical and biophysical research communications*. 332, 897–903

28 Lunder, M. *et al.* (2005) Peptide inhibitor of pancreatic lipase selected by phage display using different elution strategies. *Journal of lipid research*. 46, 1512–6

29 Yang, W. P. *et al.* (1995) CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *Journal of molecular biology*. 254, 392–403

30 Schier, R. *et al.* (1996) Isolation of picomolar affinity anti-c-erbB-2 single-chain Fv by molecular evolution of the complementarity determining regions in the center of the antibody binding site. *Journal of molecular biology*. 263, 551–67

31 Sidhu, S. S. *et al.* (2000) High copy display of large proteins on phage for functional selections. *Journal of molecular biology*. 296, 487–95

32 Messing, J. (1993) M13 cloning vehicles. Their contribution to DNA sequencing. *Methods in molecular biology (Clifton, N.J.)*. 23, 9–22

33 Fuh, G. and Sidhu, S. S. (2000) Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS letters*. 480, 231–4

34 Smith, G. P. and Petrenko, V. A. (1997) Phage Display. *Chemical Reviews*. 97, 391–410

35 Kehoe, J. W. and Kay, B. K. (2005) Filamentous phage display in the new millennium. *Chemical reviews*. 105, 4056–72

36 Malys, N. *et al.* (2002) A bipartite bacteriophage T4 SOC and HOC randomized peptide display library: detection and analysis of phage T4 terminase (gp17) and late sigma factor (gp55) interaction. *Journal of molecular biology*. 319, 289–304

37 Kalniņa, Z. *et al.* (2008) Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients. *Journal of immunological methods*. 334, 37–50

38 Lowman, H. B. (1997) Bacteriophage display and discovery of peptide leads for drug development. *Annual review of biophysics and biomolecular structure*. 26, 401–24

39 Bass, S. *et al.* (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins*. 8, 309–14

40 Hoogenboom, H. R. (2002) Overview of antibody phage-display technology and its applications. *Methods in molecular biology (Clifton, N.J.)*. 178, 1–37

41 McCafferty, J. *et al.* (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*. 348, 552–4

42 Clackson, T. *et al.* (1991) Making antibody fragments using phage display libraries. *Nature*. 352, 624–8

43 Marks, J. D. *et al.* (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *Journal of molecular biology*. 222, 581–97

44 Barbas, C. F. *et al.* (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proceedings of the National Academy of Sciences of the United States of America*. 88, 7978–82

45 Jespers, L. S. *et al.* (1994) Guiding the Selection of Human Antibodies from Phage Display Repertoires to a Single Epitope of an Antigen. *Bio/Technology*. 12, 899–903

46 Schorr, J. *et al.* (1991) Surface expression of malarial antigens in Salmonella typhimurium: induction of serum antibody response upon oral vaccination of mice. *Vaccine*. 9, 675–81

47 Freudl, R. *et al.* (1986) Cell surface exposure of the outer membrane protein OmpA of Escherichia coli K-12. *Journal of molecular biology*. 188, 491–4

48 Freudl, R. (1989) Insertion of peptides into cell-surface-exposed areas of the Escherichia coli OmpA protein does not interfere with export and membrane assembly. *Gene*. 82, 229–36

49 Charbit, A. *et al.* (1986) Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. *The EMBO journal*. 5, 3029–37

50 Steidler, L. *et al.* (1993) LamB as a carrier molecule for the functional exposition of IgG-binding domains of the Staphylococcus aureus protein A at the surface of Escherichia coli K12. *Molecular & general genetics : MGG*. 236, 187–92

51 Agterberg, M. *et al.* (1988) Expression of Escherichia coli PhoE protein in avirulent Salmonella typhimurium aroA and galE strains. *FEMS Microbiology Letters*. 50, 295–299

52 Nakajima, H. *et al.* (2000) Expression of random peptide fused to invasin on bacterial cell surface for selection of cell-targeting peptides. *Gene*. 260, 121–31

53 Samuelson, P. *et al.* (2002) Display of proteins on bacteria. *Journal of biotechnology*. 96, 129–54

54 Richins, R. D. *et al.* (1997) Biodegradation of organophosphorus pesticides by surface-expressed organophosphorus hydrolase. *Nature biotechnology*. 15, 984–7

55 Francisco, J. A. *et al.* (1992) Transport and anchoring of beta-lactamase to the external surface of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. 89, 2713–7

56 Fischetti, V. A. *et al.* (1993) Expression of foreign proteins on gram-positive commensal bacteria for mucosal vaccine delivery. *Current opinion in biotechnology*. 4, 603–10

57 Samuelson, P. *et al.* (1995) Cell surface display of recombinant proteins on Staphylococcus carnosus. *Journal of bacteriology*. 177, 1470–6

58 Acheson, D. W. K., Sonenshein, A.L., Leong, J.M., Keusch, and G.T (1997) Heat-stable spore-based vaccines: surface expression of invasin-cell wall fusion proteins in Bacillus subtilis. In: Brown, F., Burton, D., Doherty, P., Mekalanos, J., Norrby, E. (Eds.), . *Vaccines 97. Cold Spring Harbor Laboratory Press, New York, NY*.

59 Rathsam, C., Giffard, P.M., Jacques, N.A., (1993) The cell- bound fructosyltransferase of Streptococcus saliarius: the carboxyl terminus specifies attachment in a Streptococcus gordonii model system. *J. Bacteriol.* 175, 4520–4527

60 Mesnage, S., Tosi-Couture, E., Fouet, A (1999) Production and cell surface anchoring of functional fusions between the SLH motifs of the Bacillus anthracis S-layer proteins and the Bacillus subtilis levansucrase. . *Mol. Microbiol.* . 31, 927–936.

61 Lu, Z., Murray, K.S., Cleave, V.V., LaVallie, E.R., Sta˚ hl, and M.L., M., J.M., (1995) Expression of thioredoxin ran- dom peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. . *Bio/Technology* . 13, 366–372

62 Christmann, A., Walter, K., Wentzel, A., Kratzner, R., Kolmar, and H (1999) The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides. *Protein Eng.* . 12, 797–806

63 Gotz, F. and Strauss, A. (1996) n vivo immobilization of enzymatically active polypeptides on the cell surface of Staphy- lococcus carnosus. *Mol. Microbiol.* . 21, 491–500

64 Kondo, a and Ueda, M. (2004) Yeast cell-surface display--applications of molecular display. *Applied microbiology and biotechnology*. 64, 28–40

65 Cappellaro, C. *et al.* (1991) Saccharomyces cerevisiae a- and alpha-agglutinin: characterization of their molecular interaction. *The EMBO journal*. 10, 4081–8

66 Miki, B. L. *et al.* (1982) Possible mechanism for flocculation interactions governed by gene FLO1 in Saccharomyces cerevisiae. *J. Bacteriol.* 150, 878–889

67 Goossens, K. V. Y. *et al.* (2011) The N-terminal domain of the Flo1 flocculation protein from Saccharomyces cerevisiae binds specifically to mannose carbohydrates. *Eukaryotic cell.* 10, 110–7

68 Sato, N. *et al.* (2002) Long anchor using Flo1 protein enhances reactivity of cell surface-displayed glucoamylase to polymer substrates. *Applied microbiology and biotechnology*. 60, 469–74

69 Matsumoto, T. *et al.* (2004) Enantioselective transesterification using lipase-displaying yeast whole-cell biocatalyst. *Applied microbiology and biotechnology*. 64, 481–5

70 Kondo, A. *et al.* (2002) High-level ethanol production from starch by a flocculent Saccharomyces cerevisiae strain displaying cell-surface glucoamylase. *Applied microbiology and biotechnology*. 58, 291–6

71 Fujita, Y. *et al.* (2002) Direct and efficient production of ethanol from cellulosic material with a yeast strain displaying cellulolytic enzymes. *Applied and environmental microbiology*. 68, 5136–41

72 Boder, E. T. *et al.* (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences of the United States of America*. 97, 10701–5

73 Zou, W. *et al.* (2002) Screening of a molecule endowing Saccharomyces cerevisiae with n-nonane-tolerance from a combinatorial random protein library. *Applied microbiology and biotechnology*. 58, 806–12

74 Olsen, M. J. *et al.* (2000) Function-based isolation of novel enzymes from a large library. *Nature biotechnology*. 18, 1071–4

75 Boublik, Y. *et al.* (1995) Eukaryotic virus display: engineering the major surface glycoprotein of the Autographa californica nuclear polyhedrosis virus (AcNPV) for the presentation of foreign proteins on the virus surface. *Bio/technology (Nature Publishing Company)*. 13, 1079–84

76 Smith, A. D. *et al.* (1998) Human rhinovirus type 14:human immunodeficiency virus type 1 (HIV-1) V3 loop chimeras from a combinatorial library induce potent neutralizing antibody responses against HIV-1. *Journal of virology*. 72, 651–9

77 Taube, R. *et al.* (2008) Lentivirus display: stable expression of human antibodies on the surface of human cells and virus particles. *PloS one*. 3, e3181

78 Zhou, C. *et al.* Development of a novel mammalian cell surface antibody display platform. *mAbs*. 2, 508–18

79 Alonso-Camino, V. *et al.* (2009) Lymphocyte display: a novel antibody selection platform based on T cell activation. *PloS one*. 4, e7174

80 Lu, W.-C. and Ellington, A. D. (2012) In vitro selection of proteins via emulsion compartments. *Methods (San Diego, Calif.)*. DOI: 10.1016/j.ymeth.2012.03.008

81 Kitagawa, M. *et al.* (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA research : an international journal for rapid publication of reports on genes and genomes*. 12, 291–9

82 Bowers, P. M. *et al.* (2011) Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proceedings of the National Academy of Sciences of the United States of America*. 108, 20455–60

83 Persic, L. *et al.* (1997) Targeting vectors for intracellular immunisation. *Gene*. 187, 1–8

84 Fuchs, P. *et al.* (1991) Targeting recombinant antibodies to the surface of Escherichia coli: fusion to a peptidoglycan associated lipoprotein. *Bio/technology (Nature Publishing Company)*. 9, 1369–72

85 Cull, M. G. (1992) Screening for Receptor Ligands Using Large Libraries of Peptides Linked to the C Terminus of the lac Repressor. *Proceedings of the National Academy of Sciences*. 89, 1865–1869

86 Roberts, R. W. and Szostak, J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 94, 12297–302

87 Mattheakis, L. C. *et al.* (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences of the United States of America*. 91, 9022–6

88 He, M. (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. *Nucleic Acids Research*. 25, 5132–5134

89 He, M. *et al.* (1999) Selection of a human anti-progesterone antibody fragment from a transgenic mouse library by ARM ribosome display. *Journal of immunological methods*. 231, 105–17

90 He, M. and Taussig, M. J. (2002) Technique review Ribosome display : Cell-free protein display technology. 1, 204–212

91 Gersuk, G. M. *et al.* (1997) High-affinity peptide ligands to prostate-specific antigen identified by polysome selection. *Biochemical and biophysical research communications*. 232, 578–82

92 Hanes, J. *et al.* (1998) Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proceedings of the National Academy of Sciences of the United States of America*. 95, 14130–5

93 Hanes, J. *et al.* (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature biotechnology*. 18, 1287–92

94 Ellman, J. *et al.* (1991) Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods in enzymology*. 202, 301–36

95 Jermutus, L. *et al.* (2001) Tailoring in vitro evolution for protein affinity or stability. *Proceedings of the National Academy of Sciences*. 98, 75–80

96 He, M. and Taussig, M. J. (2001) Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic acids research*. 29, E73–3

97 Keefe, a D. and Szostak, J. W. (2001) Functional proteins from a random-sequence library. *Nature*. 410, 715–8

98 Fukuda, I. *et al.* (2006) In vitro evolution of single-chain antibodies using mRNA display. *Nucleic acids research*. 34, e127

99 Liu, R. *et al.* (2000) *RNA-Ligand Interactions Part B*, 318Elsevier.

100 Gold, L. (2001) mRNA display: diversity matters during in vitro selection. *Proceedings of the National Academy of Sciences of the United States of America*. 98, 4825–6

101 Wilson, D. S. *et al.* (2001) The use of mRNA display to select high-affinity protein-binding peptides. *Proceedings of the National Academy of Sciences of the United States of America*. 98, 3750–5

102 Xu, L. *et al.* (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & biology*. 9, 933–42

103 Weng, S. *et al.* (2002) Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. *Proteomics*. 2, 48–57

104 Cload, S. T. *et al.* proteins containing unnatural amino acids.

105 Li, S. *et al.* (2002) In vitro selection of mRNA display libraries containing an unnatural amino acid. *Journal of the American Chemical Society*. 124, 9972–3

106 Odegrip, R. *et al.* (2004) CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*. 101, 2806–10

107 Stein, V. and Hollfelder, F. (2009) An efficient method to assemble linear DNA templates for in vitro screening and selection systems. *Nucleic acids research*. 37, e122

108 Reiersen, H. *et al.* (2005) Covalent antibody display--an in vitro antibody-DNA library selection system. *Nucleic acids research*. 33, e10

109 Yonezawa, M. (2004) DNA Display of Biologically Active Proteins for In Vitro Protein Selection. *Journal of Biochemistry*. 135, 285–288

110 Kojima, T. *et al.* (2005) PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets. *Nucleic acids research*. 33, e150

111 Doi, N. and Yanagawa, H. (1999) STABLE : protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *FEBS Letters*. 457, 227–230

112 Bertschinger, J. and Neri, D. (2004) Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein engineering, design & selection : PEDS*. 17, 699–707

113 Griffiths, A. D. and Tawfik, D. S. (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *The EMBO journal*. 22, 24–35

114 Diehl, F. *et al.* (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 102, 16368–73

115 Diehl, F. *et al.* (2006) BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nature methods*. 3, 551–9

116 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437, 376–380

117 Stein, V. *et al.* (2007) A covalent chemical genotype-phenotype linkage for in vitro protein evolution. *Chembiochem : a European journal of chemical biology*. 8, 2191–4

118 Lee, Y.-F. (2002) Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. *Nucleic Acids Research*. 30, 4937–4944

119 Cohen, H. M. *et al.* (2004) Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein engineering, design & selection : PEDS*. 17, 3–11

120 Ghadessy, F. J. *et al.* (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America*. 98, 4552–7

121 Sepp, A. *et al.* (2002) Microbead display by in vitro compartmentalisation: selection for binding using flow cytometry. *FEBS letters*. 532, 455–8

122 Yonezawa, M. (2003) DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Research*. 31, 118e–118

123 Bernath, K. *et al.* (2005) Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. *Journal of molecular biology*. 345, 1015–26

124 Jongsma, M. a and Litjens, R. H. G. M. (2006) Self-assembling protein arrays on DNA chips by auto-labeling fusion proteins with a single DNA address. *Proteomics*. 6, 2650–5

125 Aharoni, A. *et al.* (2005) High-throughput screens and selections of enzyme-encoding genes. *Current opinion in chemical biology*. 9, 210–6

126 Bernath, K. *et al.* (2004) In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Analytical biochemistry*. 325, 151–7

127 Mastrobattista, E. *et al.* (2005) High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & biology*. 12, 1291–300

128 Thorsen, T. *et al.* (2001) Dynamic Pattern Formation in a Vesicle-Generating Microfluidic Device. *Physical Review Letters*. 86, 4163–4166

129 Umbanhowar, P. B. *et al.* (2000) Monodisperse Emulsion Generation via Drop Break Off in a Coflowing Stream.

130 Utada, A. S. *et al.* (2005) Monodisperse double emulsions generated from a microcapillary device. *Science (New York, N.Y.)*. 308, 537–41

131 Nisisako, T. *et al.* (2005) Controlled formulation of monodisperse double emulsions in a multiple-phase microfluidic system. *Soft Matter*. 1, 23

132 Dittrich, P. S. *et al.* (2005) A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices. *Chembiochem : a European journal of chemical biology*. 6, 811–4

133 Kintses, B. *et al.* (2010) Microfluidic droplets: new integrated workflows for biological experiments. *Current opinion in chemical biology*. 14, 548–55

134 Chen, Y. *et al.* (2008) Cell-free selection of RNA-binding proteins using in vitro compartmentalization. *Nucleic acids research*. 36, e128

135 Keppler, A. *et al.* (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature biotechnology*. 21, 86–9

136 Gronemeyer, T. *et al.* (2005) Adding value to fusion proteins through covalent labelling. *Current opinion in biotechnology*. 16, 453–8

137 Connolly, B. A. and Rider, P. (1985) Volume 13 Number 12 1985 Nucleic Acids Research Chemical synthesis of oligonucleotides containing a free sulphydryl group and subsequent attachment of thiol specific probes Nucleic Acids Research. 13, 4485–4502

138 Williams, R. *et al.* (2006) Amplification of complex gene libraries by emulsion PCR. *Nature Methods*. 3, 545–550

139 Miller, O. J. *et al.* (2006) Directed evolution by in vitro compartmentalization. *Nature Methods*. 3, 561–570

140 Schaerli, Y. and Hollfelder, F. (2009) The potential of microfluidic water-in-oil droplets in experimental biology. *Molecular bioSystems*. 5, 1392–404

141 Griffiths, A. D. and Tawfik, D. S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends in biotechnology*. 24, 395–402

142 Fallah-Araghi, A. *et al.* (2012) A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab on a chip*. 12, 882–91

143 Dressman, D. *et al.* (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*. 100, 8817–22

144 Zhang, H. *et al.* (2012) Massively parallel single-molecule and single-cell emulsion reverse transcription polymerase chain reaction using agarose droplet microfluidics. *Analytical chemistry*. 84, 3599–606

145 Barchanski, A. *et al.* (2012) Impact of Spacer and Strand Length on Oligonucleotide Conjugation to the Surface of Ligand-Free Laser-Generated Gold Nanoparticles. *Bioconjugate chemistry*. 23, 908–915

146 Ghadessy, F. J. and Holliger, P. (2004) A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein engineering, design & selection : PEDS*. 17, 201–4

147 Anderson, J. C. *et al.* (2007) Environmental signal integration by a modular AND gate. *Molecular systems biology*. 3, 133

148 Alper, H. *et al.* (2005) Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America*. 102, 12678–83

149 Kelly, J. R. *et al.* (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of biological engineering*. 3, 4

150 Song, H. and Ismagilov, R. F. (2003) Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *Journal of the American Chemical Society*. 125, 14613–9

151 Hoogenboom, H. R. (2005) Selecting and screening recombinant antibody libraries. *Nature biotechnology*. 23, 1105–16

152 Bandwar, R. P. and Patel, S. S. (2002) The Energetics of Consensus Promoter Opening by T7 RNA Polymerase. *Journal of Molecular Biology*. 324, 63–72

153 Deane, C. M. (2002) Protein Interactions: Two Methods for Assessment of the Reliability of High Throughput Observations. *Molecular & Cellular Proteomics*. 1, 349–356

154 Shao, K. *et al.* (2011) Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection. *PloS one*. 6, e24910

155 Sumida, T. *et al.* (2009) Bicistronic DNA display for in vitro selection of Fab fragments. *Nucleic acids research*. 37, e147

# Appendix

## Patents

**Paul, S**, Stang, A, Ueberla K 2012.T7 promoter variants and methods of using the same. European Patent Application .12 177 111.7, filled July 2012. *patent pending*

## Publications

**Paul, S**, Stang, A, Lennartz, K, Tenbusch, M, Ueberla, K. "Selection of a T7 promoter mutant, with enhanced *in vitro* activity by a novel multi-copy bead display approach for *in vitro* evolution" *Nucleic Acid research* **2012,** *In press.*

### Manuscript in preparation

**Paul, S**, *et al.* "Identification and characterization of mimotopes in protein scaffolds to broadly neutralization HIV antibody 2F5 through multi-copy bead display system"

### Other Publications

Asokan, M, Lone, I.N., Mukthey, A.B., **Paul, S**, *et al* (April 2012) "Evident stabilization of the clinical profile in HIV/AIDS as evaluated in an open label clinical trial using a polyherbal formulation*", Indian Journal of Medical research,* **2012**, *In press.*

Kashi, V.P, Jacob, R.A, **Paul, S**, Nayak, K, Satish, B, Swaminathan, S, Satish, K.S,Ranga, U. "HIV-1 Tat-specific IgG antibodies in high-responders target a B-cell epitopein the cysteine-rich domain and block extracellular Tat efficiently" *Vaccine,* **2009,** 27(48),6739-47

## Conference contributions

**Paul, S**, Stang, A, Lennartz, K, Tenbusch, M, Ueberla, K. Selection of a T7 promoter mutant, with enhanced *in vitro* activity by a novel multi-copy bead display approach for *in vitro* evolution. EMBO conference series: Chemical Biology, Heidelberg, Germany September 2012

**Paul S,** Stang A, Ueberla K. Development of a novel bead display approach for the identification of mimotopes for broadly neutralizing HIV antibody (2F5). FoRUM tagung, Bochum, Germany November 2009

# Curriculum vitae

## Personal data

| | |
|---|---|
| Name: | Siddhartha Paul |
| Date of Birth | 24.11.1979 |
| Place of Birth | Kamrup (India) |
| Nationality | India |

## Education

**Since 2008:** PhD, Ruhr University Bochum,Germany.
*"Development of a multi-copy bead display approach and its application in in vitro evolution of proteins and nucleic acids"* ( Prof. Dr. med. Klaus ueberla)

**2005 - 2007** Junior Research Fellow, Jawaharlal nehru centre for Advanced scientific Reseach.(Prof. Udaykumar Ranga)

**2002 - 2004** M.Sc. Biochemistry, University of Mysore, Mysore, India.
*"Differential expression of matrix metalloproteinase (MMP-2) and its partial purification from tumor samples"* (Prof. M Karuna Kumar)

**1999 - 2002** B.Sc. Chemistry, Botany, Microbiology, Bangalore University, Bangalore, India.

**1996 -1998** Higher Secondary, Swadeshi Academy, Guwahati, India.

**1986 - 1996** High School, Don Bosco Boys High School, Guwahati, India.

## Awards and Fellowships

**2008 - 2012** Fellowship from Department of Medicine, Ruhr University Bochum, Germany.

**2005 - 2007** Junior Research Fellow, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore Department of Biotechnology, Government of India