DISS. ETH NO. 16070

Covalent DNA Display as a Novel Tool for Directed Evolution of Proteins *in vitro*

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Natural Sciences

presented by

Julian Bertschinger

Dipl. Natw., ETH Zürich

born 15.09.1976

citizen of Pfäffikon ZH

accepted on the recommendation of

Prof. Dr. Dario Neri, examiner

Prof. Dr. Peter Seeberger, co-examiner

2005

Table of Contents

1. Summary

Therapeutic proteins have become an increasingly important class of medicines during the last twenty years. After the first successes with recombinant proteins in replacement therapies, monoclonal antibodies are currently representing the second wave of innovation created by the biotechnology industry. Antibodies have proved useful as human protein therapeutics because they bind to drug targets with high affinity and specificity and exhibit a favorable pharmacokinetic profile. After a single injection, they can persist for a long time in the bloodstream, maintaining their biological activity for several weeks.

However, even blockbuster biologics suffer from some drawbacks. Nature did not evolve proteins for manufacture *ex vivo* and for this reason, many human proteins produced in recombinant form are difficult to produce. Poor expression yields, low solubility or immunogenicity are problematic aspects, which are often associated with therapeutic proteins. Additionally, thermal instability and aggregate formation are known to increase the immunogenic potential of therapeutic proteins. Therefore, there is a growing interest in the development of stable and highly soluble proteins of human origin, which might be less immunogenic in therapeutic applications and easier to produce in microbial expression systems.

The engineering of proteins for therapeutic and industrial applications with suitable properties will need a toolbox comprising a variety of efficient screening and selection methodologies for directed evolution of polypeptides *in vitro*. In order to select rare binding protein variants from a large repertoire of mutant polypeptides, a linkage between the protein (phenotype) and the encoding genetic information must be created to allow the efficient amplification of those genes, which code for a protein with favorable binding properties.

In this thesis, we present a novel method for the directed evolution of polypeptides *in vitro*, where a library of mutant proteins can be covalently coupled to its encoding DNA molecule, making such DNA-protein fusions amenable to affinity selection experiments under a wide range of experimental conditions. This methodology offers the possibility to work with large polypeptide libraries (10^{10}-10^{11} library members), which can be produced by PCR. Therefore, the two hallmarks of Darwinian evolution, which are the generation of genetic

diversity followed by selection of suitable phenotypes, can be rapidly repeated, thus allowing the efficient exploration of protein sequence space.

In the experimental part of this work, we present some proof-of-principle experiments investigating the robustness and efficiency of this selection technology. Using specific ligands in model selection experiments for the capture of globular proteins conjugated to their coding DNA, we observed enrichment factors up to 150-fold over non-binding protein-DNA fusions.

Owing to the covalent link between protein and DNA, this selection technology may be well suited for the engineering of second generation protein drugs which are thermostable, highly soluble, less immunogenic and efficiently expressed in bacteria.

1. Zusammenfassung

Während der letzten zwanzig Jahre sind therapeutische Proteine zu einer immer wichtigeren Klasse von Medikamenten geworden. Nach den ersten Erfolgen mit Substitutionstherapien repräsentieren nun monoklonale Antikörper die zweite Welle von innovativen Medikamenten, welche aus den Aktivitäten der biotechnologischen Industrie hervorgegangen sind. Antikörper menschlichen Ursprungs haben ihre Nützlichkeit als therapeutische Proteine bewiesen, weil sie Zielmoleküle mit hoher Affinität und Spezifität binden können und über ein vorteilhaftes pharmakokinetisches Profil verfügen. Nach einer einzigen Injektion in den Patienten können Antikörper über längere Zeit im Kreislauf zirkulieren und ihre biologische Aktivität über mehrere Wochen behalten.

Trotz den oben genannten positiven Eigenschaften weisen auch kommerziell sehr erfolgreiche Antikörper (so genannte Blockbuster) nachteilige Merkmale auf. In der Natur wurden Proteine nicht zur Herstellung *ex vivo* entwickelt, weshalb viele menschliche Proteine nur schwer rekombinant zu produzieren sind. Schlechte Expressionsausbeuten, niedrige Löslichkeit und das Potential, im Patienten eine gegen das therapeutische Protein gerichtete Immunantwort auszulösen, sind Probleme, welche häufig im Zusammenhang mit Biotherapeutika anzutreffen sind. Zudem ist bekannt, dass niedrige Stabilität sowie die Bildung von molekularen Aggregaten das immunogene Potential von therapeutischen Proteinen erhöhen. Deshalb ist ein wachsendes Interesse an der Entwicklung von stabilen, gut löslichen Proteinen menschlichen Ursprungs entstanden, um Medikamente zu entwickeln, welche weniger immunogen und einfacher in Mikroben zu exprimieren sind.

Um Proteine mit geeigneten Eigenschaften für medizinische oder industrielle Anwendungen zu gestalten, werden verschiedenste Screening- und Selektionsverfahren für die *in vitro* Evolution von Proteinen notwendig sein. Für die Selektion eines Proteins mit bestimmten, seltenen Bindungseigenschaften aus einer grossen Population von mutanten Proteinen, muss eine Art von Verbindung zwischen dem Protein (Phänotyp) und dem codierenden Gen (Genotyp) geschaffen werden, um die effiziente Vermehrung der Gene, welche ein Protein mit gewünschten Bindungseigenschaften codieren, zu gewährleisten.

In dieser Doktorarbeit stellen wir eine neue Methode zur zielgerichteten *in vitro* Evolution von Polypeptiden vor, mit welcher in einer Sammlung von mutanten

Proteinen jedes einzelne Polypeptid mittels einer kovalenten Bindung an sein codierendes DNA Molekül gekoppelt werden kann. Die dadurch entstehenden DNA-Protein Fusionen können für Selektionsexperimente benutzt werden, die unter einer Vielzahl von experimentellen Bedingungen stattfinden können. Diese Technologie ermöglicht das Arbeiten mit grossen Sammlungen von mutanten Polypeptiden (10^{10}-10^{11}), welche mit wenig Aufwand mittels PCR hergestellt werden können. Die hier präsentierte Methode erlaubt deshalb die rasche Wiederholung der zwei wichtigen Merkmale der Darwinschen Evolution – die Schaffung von genetischer Diversität gefolgt von der Selektion der geeigneten Phänotypen, um so die Evolution von Proteinen *in vitro* auf rasche Weise voranzutreiben.

Im experimentellen Teil dieser Arbeit werden wir Modellversuche präsentieren, deren Resultate die Robustheit und die Effizienz der vorgeschlagenen Methode darlegen. An einen immobilisierten Liganden spezifisch bindende DNA-Protein Fusionen konnten bis zu 150-fach effizienter selektioniert werden als nicht bindende DNA-Protein Fusionen.

Dank der kovalenten Bindung zwischen der codierenden DNA und dem entsprechenden Protein könnte diese Selektionstechnologie für die Entwicklung von Proteinmedikamenten, welche stabil, löslich, weniger immunogen und in grossen Mengen einfach exprimierbar sind, gut geeignet sein.

2. Introduction

2.1. Recombinant proteins as biopharmaceuticals

The advent of recombinant DNA technology has allowed the cloning of genes in living organisms such as bacteria and yeast for their efficient expression. Large-scale production of many proteins has become possible, thus opening several industrial applications, including the production of enzymes for food processing or the synthesis of chemicals and medicinal products.
Protein-based recombinant pharmaceuticals represent a growing portion of the pharmaceutical market. By the end of the year 2000, 84 recombinant biopharmaceuticals had gained market authorization in the United States and/or the European Union for use in humans (Walsh, 2000). Over the following three years, over 60 additional products have been approved and around 500 are undergoing clinical evaluation (Walsh, 2003). Approximately 250 million people have benefited from these "biotech" medicines, and so has the business of biotechnology. The annual global market for biopharmaceuticals is estimated at more than 30 billions USD, compared with 12 billion USD just three years ago (Robinson, 2002).

The first human recombinant protein (Humulin) was launched by Genentech with the help of its licensing partner Eli Lilly in 1982. Following the approval of Humulin and human growth hormone, the interferon and interleukin therapeutics entered the clinic. The three top-selling therapeutic proteins in 2003 [Procrit (4.0 billion USD, Johnson & Johnson), Epogen (2.4 billion USD, Amgen) and Novolin (2.2 billion USD, Novo Nordisk)] were approved by the FDA from 1989 until 1991 (Pavlou and Reichert, 2004). During the following decade from 1990 – 2000, not only unmodified but also engineered human recombinant proteins as well as monoclonal antibody-based products were successfully introduced onto the market.

2.1.1. Monoclonal antibody-based products

Following the success of recombinant proteins, therapeutic monoclonal antibodies (mAbs) represent the second wave of innovation created by the biotechnology industry during the past twenty years. Between 2001 and 2002, the value of the global market grew by 37% to 5.4 billion USD (Reichert and Pavlou, 2004). To

date, more than 15 mAbs, comprising four different types, have been approved by the FDA in the United States: three murine, five chimeric, eight humanized and one human (Table 2.1).

Depending on the antibody format used to design an antibody-based drug, a favorable pharmacokinetic profile can be achieved. Antibodies have proved useful as human protein therapeutics because they exhibit an appropriate pharmacokinetic profile. After a single injection, they can persist for a long time in the bloodstream, maintaining their biological activity for several weeks.

Table 2.1. Therapeutic mAbs approved in the United States

Company	US trade name	mAb type	indication	target	approval
Johnson & Johnson	Orhtoclone OKT3	murine	transplant rejection	CD3	1986
Centocor	ReoPro	chimeric	inhibition of blood clotting in caridivascular diseases	platelet surface receptor	1994
Biogen IDEC	Rituxan	chimeric	Non-Hodgkin lymphoma	CD20 on B-cells	1997
Protein Design Labs	Zenapax	humanized	transplant rejection	CD25	1997
Novartis	Simulect	chimeric	transplant rejection	CD25	1998
MedImmune	Synagis	humanized	Respiratorial sycytial virus	RSV	1998
Centocor	Remicade	chimeric	Crohn's diesease Rheumatoid arthritis	TNF-α	1998
Genentech	Herceptin	humanized	Breast cancer	Her-2	1998
Wyeth	Mylotarg	humanized-drug	Acute myleoid leukemia	CD33	2000
Millenium/LEX	Campath	humanized	Chronic lymphocytic leukemia	CD52	2001
Biogen IDEC	Zevalin	mouse-^{90}Y	Non-Hodgkin lymphoma	CD20	2002
Abott	Humira	human	Rheumatoid arthritis	TNF-α	2002
Genentech	Xolair	humanized	Asthma	IgE	2003
Corixa	Bexxar	mouse-^{131}I	Non-Hodgkin lymphoma	CD20	2003
Genentech	Raptiva	humanized	Psoriasis	CD11a	2003
Imclone Systems	Erbitux	chimeric	Colorectal cancer	EGF receptor	2004
Genentech	Avastin	humanized	Colorectal Cancer	VEGF	2004

Source: Reichert and Pavlou, 2004

The current global clinical antibody pipeline, which comprises more than 130 products in development and is dominated by humanized and fully human mAbs, will probably enable the industry to double the number of approved mAbs within the next five years.

2.2. Small globular proteins as antibody substitutes

Even blockbuster therapeutic antibodies suffer from drawbacks, such as the requirement for expensive mammalian cell production system and the need for

intravenous, intramuscular or subcutaneous injection. With molecular weights of around 150'000, whole IgGs are too large to be administered by any other route. In addition, antibodies – as almost any other protein - are readily digested in the gastrointestinal tract if deliverd orally.

An ideal drug would have the following qualities: it would have high affinity and exquisite specificity for its target; it could be manufactured in bacterial or yeast expression systems; it would be both very stable and highly soluble; it could be delivered to any part of the human body by any route of administration; and once there, the pharmacokinetic profile of the drug would be as such, that the drug has the desired therapeutic effect (Tomlinson, 2004). Achieving all these goals for protein drugs as well as small molecule drugs is a daunting challenge.

Virtually every therapeutic effect is the result of a specific drug-target interaction. Therefore, the ability to engineer the binding properties of a molecule – protein or small molecule – is very desirable even if the specific binding of a biological target is not sufficient *per se* to turn a binding molecule into a drug. Other properties such as adsorbtion or body uptake, distribution, metabolism, excretion, stability and ease of manufacturing are also recognized to be very important.

Over the last few years, several research groups and new biotech companies focused their research on the development of small globular proteins as substitutes for antibody-based drugs with improved therapeutic properties. Common to all approaches of finding a suitable protein scaffold for the generation of specifically binding protein therapeutics are the following steps: i) choosing a small human protein (domain) that is well expressed in bacteria or yeast and has good biophysical properties (stability, solubility), ii) creating a large library of protein mutants by introducing diversity in loop regions of the chosen scaffold, iii) using a genotype-phenotype display system (such as phage or ribosome display, see chapter 2.5.) to select a range of binders to a given therapeutic target, and iv) screening the binders obtained by the selection procedure for the desired biological activity.

Table 2.2. Selected companies using single-domain proteins as scaffolds to create drugs

Company	Protein scaffold	Species
Affibody (Sweden)	Protein A	Staphyl. aureus
Ablynx (Belgium)	Domain antibodies	Camelids
BioRexis (USA)	Transferrin	Human
Borean Pharma (Denmark)	C-type lectins	Human
Compound Therapeutics (USA)[a]	Trinectins	Human
Domantis (UK)	Domain antibodies	Human
Dyax (USA)	Kunitz domains	Human
Molecular Partners (Switzerland)	Ankyrins	Human
Pieris ProteoLab (Germany)	Lipocalins	Human

[a]On March 9, 2004, Compound Therapeutics announced the acquisition of the intellectual property of Phylos (USA).

However, in many cases, the mutations introduced in the protein scaffold to produce diversity often compromise the three-dimensional structure, stability and solubility of the protein scaffold, thus making the isolation of protein binders based on other folding scaffolds than the immunoglobulin fold difficult. The basis for the successful development of well designed protein drugs lies in the availability of both an efficient selection or screening technology in order to find suitable candidate proteins in the large repertoire created by mutagenesis, and a well behaving protein scaffold.

2.2.1. Domain antibodies (dAbs)

A domain antibody is either the variable domain of an antibody heavy domain (V_H domain) or the variable domain of an antibody light chain (V_L domain). Each dAb therefore contains three of the six naturally occurring complementary determining regions (CDRs) from an antibody, which are the highly diversified loop regions that bind to the target antigen (Holt *et al.*, 2003). Although it might seem surprising that three CDR regions are sufficient to confer antigen-binding specificity and high affinity, Darwinian evolution has itself arrived at the very same solution in camelids and sharks (Stanfield *et al.*, 2004), which produce antibodies comprising only a heavy chain (also designated as V_{HH} domains). The antigen-binding site of these antibodies consists of a single unpaired variable domain and bind a wide range of antigens, from small molecules such as haptens

to large proteins (Spinelli *et al.*, 2000). However, some of the CDR structures seen in dAbs with camelid origin are different from those in naturally paired V_H domains. A comparison of the structures of camelid CDRH1 and CDRH2 with the set of loops found in human and mouse antibodies has revealed significant differences in their main chain conformations (Decanniere *et al.*, 2000). In addition, the CDRH3 regions of naturally occurring camelid antibodies are, on average, longer than the CDRH3 regions of naturally paired V_H domains.

As with any protein of commercial value, dAbs need to be manufactured in high yield and must be extremely soluble and resistant to proteolysis, aggregation, denaturation and chemical degradation. The first murine V_H domain antibodies that were cloned, however, had relatively low expression levels and solubility (Ward *et al.*, 1989). Because the V_H domain was originally selected (in the mouse) for binding in the presence of a cognate V_L molecule, the removal of that V_L domain presumably resulted in a significant reduction in stability and/or the exposure of hydrophobic residues that would normally form part of the V_H-V_L interface. The observation of four highly conserved hydrophilic mutations at the surface of camelid domains, where the human V_H domain forms the interface with its pairing V_L domain, led to the idea of "camelisation" of human V_H domains by introducing the hydrophilic mutations found in camelid domains (Davies and Riechmann, 1994). However, camelisation only partially eliminated the tendency of the engineered V_H domains to aggregate, and the modified V_H domains were still poorly expressed and less thermodynamically stable than their wild-type (camelid) counterparts owing to β-sheet deformation (Riechmann, 1996). Further work elucidated the importance of CDR length and diversity for the solubility of dAbs. The CDRH3 loop of camelid domains is on average much longer (17 residues) than the human (12 residues) or the mouse (9 residues) CDRH3 loops, which causes the CDRH3 loops to cover the hydrophobic interface that would be formed with V_L (Desmyter *et al.*, 1996; Desmyter *et al.*, 2002). Consequently, the expression and solubilities of camelised and human V_H domains have been substantially improved by extending the length of the CDRH3 loop (Ewert *et al.*, 2002; Ewert *et al.*, 2003).

A camelid V_{HH} domain binding to carcinoembryonic antigen (CEA) was isolated by immunization of a dromedary and subsequent cloning of the repertoire of the variable domains of its heavy-chain antibodies in a phage display vector. Affinity selection of the resulting library yielded a V_{HH} domain with a dissociation constant in the sub-nanomolar range. This domain antibody was conjugated to a bacterial β-lactamase, and the immunoconjugate was used in a targeting strategy known

as antibody-dependent enzyme prodrug therapy (Bagshawe, 1987). The immuno-conjugate targeted solid tumours (~1% injected dose/g, tumour to blood ratio ~10:1 after 24 h) and was shown to induce cures and regressions in established tumour xenografts in mice (Cortez-Retamozo *et al.*, 2004).

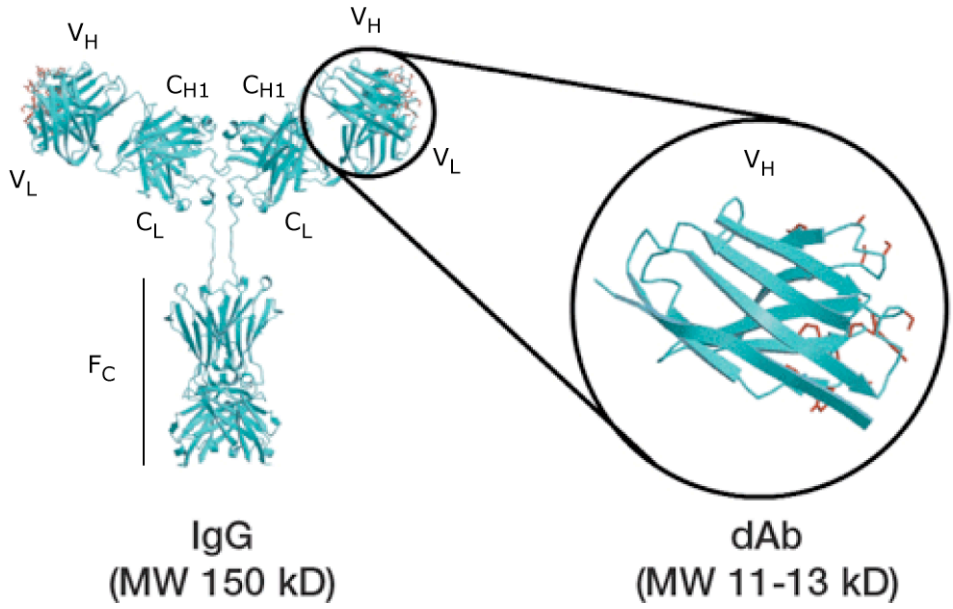

Figure 2.1. A human IgG molecule has both variable and constant regions, of which the former, present at the tips of the arms of the IgG molecule, confer antigen-binding specificity on the antibody. The constant regions building up the stem-like structure of the IgG (Fc fragment) recruits the effector functions of the immune system. The domain antibody (dAb) is only an isolated V_H or V_L domain. Each dAb thus contains only three of the naturally occurring six complementary determining regions (CDRs). The side chains of the CDRs are highlighted in red (adapted from Holt *et al.*, 2003).

Phage display has been used to select human, murine and camelid dAbs with high specificity and affinity to a range of antigens and antigen types, including binders against haptens, enzyme active sites and proteins as diverse as transcription factors and cytokines (Dumoulin *et al.*, 2002; Reiter *et al.*, 1999; Davies and Riechmann, 1995). Once a biologically active dAb has been identified, tens to hundreds of kilograms per year might be needed for use as a drug. Choosing an optimal expression system is therefore crucial to maximise the quality of purified product and to minimise cost of goods. The most commonly used system for expressing dAbs is periplasmic expression in *Escherichia coli*. Human V_H dAbs are produced in yields of up to 10 gml^{-1} in shake flask culture and human V_L dAb frameworks with expression levels of up to 17 mgl^{-1} have been identified (Holt *et al.*, 2003; Ewert *et al.*, 2003).

In a recent publication by Jespers *et al.*, aggregation-resistant domain antibodies were selected on phage by heat denaturation (Jespers *et al.*, 2004). Starting from the DP47d domain antibody (a typical human V_H dAb), which unfolds irreversibly and forms aggregates if heated above 55°C, a repertoire containing $1.6x10^9$ different mutants was cloned by diversification of the CDR loops. The library was multivalently displayed on phage and after three rounds of heat denaturation followed by selection on protein A, 179 out of 200 colonies secreted dAb phage that retained more than 80% of protein A-binding activity after heating. Twenty clones were sequenced and revealed many unique dAb sequences with a large variability in the CDR length and sequences, which shows that mutation located only in the CDR loops of the dAbs are sufficient to confer resistance to aggregation. Interestingly, the temperature of mid-point, heat induced reversible unfolding of the mutants was not higher than the temperature of mid-point, heat induced irreversible aggregation of the parent clone DP47d, which shows that the selected dAbs are not heat stable, but can unfold reversibly. The expression yields in *E.coli* (20-40 mgl^{-1}) of the mutants correlated with their increased resistance to aggregation.

Some progress has been made towards the design of human dAbs, which are soluble and can be expressed in *E.coli* with acceptable yields. However, the disulfide bond in the immunoglobulin fold of the dAb can lead (when fused to other proteins containing additional disulfide bonds or free cysteines) to the formation of disulfide bridges between non-cognate cysteines, and thus, cause misfolded and inactive proteins. Very recently, the engineering of an V_L domain devoid of disulfide bonds was shown to be stable and efficiently expressed in the cytoplasm of *Saccharomyces cerevisiae* (Colby *et al.*, 2004).

2.2.2 Lipocalins

The lipocalins represent a family of functionally diverse, small proteins that comprise 160–180 amino acid residues and have weak sequence homology but high similarity at the tertiary structural level. Members of this family have important biological functions in a variety of organisms, from bacteria to humans (Flower, 1996; Skerra, 2000). The majority of lipocalins are responsible for storage and transport of compounds with low solubility or chemical stability. An example of a human lipocalin is the retinol binding protein (RBP), of which the 3D structure has been elucidated by X-ray cristallography (Cowan *et al.*, 1990). RBP transports the poorly soluble vitamin A from the liver, where it is stored as a fatty

acid ester, to several target tissues. Common to the lipocalins is a conserved β-barrel structure formed by eight antiparallel strands and four loops, which are highly variable among different members of the lipocalin family and involved in the binding of the ligand (Figure 2.2.).

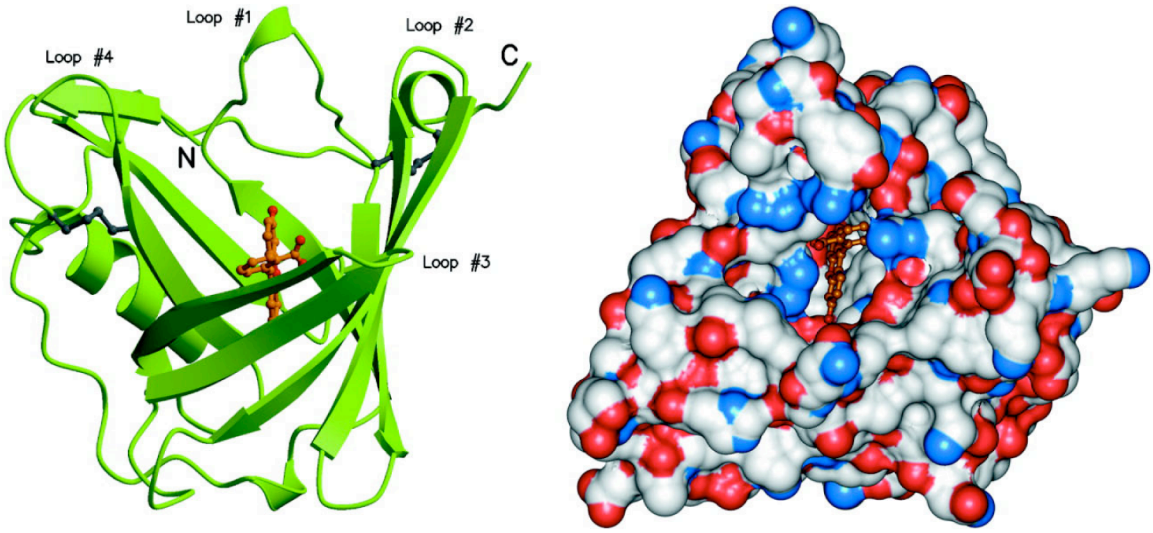

Figure 2.2. Crystal structure of an engineered lipocalin FluA binding to fluorescein. (*left*) Ribbon diagram of FluA (green) in complex with fluorescein (yellow, with oxygen atoms colored red). (*right*) Solvent-accessible surface of the polypeptide chain (C, white; O, red; N, blue), with fluorescein occupying the ligand pocket (adapted from Korndorfer *et al.*, 2003).

The bilin-binding protein (BBP) from *Pieris brassicae* served as a model lipocalin for the initial studies to create artificial binding sites for defined ligand specificities. 16 residues at the center of the natural binding site for bilin, which is formed by four loops on top of the eight-stranded β-barrel, were subjected to random mutagenesis. A phage display library comprising $4x10^8$ individual mutant BBPs was used in affinity selection experiments with immobilised low molecular compounds as capture reagents. Thus, BBP variants that specifically recognise fluorescein (Beste *et al.*, 1999), digoxigenin (Schlehuber *et al.*, 2000), phtalic acid esters (Mercader and Skerra, 2002) and doxorubicin (Schlehuber and Skerra, 2005) were isolated. Observed affinities were in the nanomolar range and thus higher than those known for many lipocalins with their natural ligands. This new class of engineered binding proteins with antibody-like binding function emerged, which was designated as "anticalins". In order to extend the concept of anticalins, several human lipocalins were subjected to random mutagenesis and affinity selection experiments with low molecular weight molecules as immobilised ligands. Furthermore, anticalin libraries were generated to enable recognition of

macromolecular protein targets. Because proteins have larger molecular dimensions than small compounds, they cannot penetrate into the ligand-binding pocket of lipocalins, and thus, amino acid residues positioned at the tip of the loops of the lipocalin scaffolds were randomly mutated. Following this strategy, various anticalins based on human lipocalins with binding specificities for several therapeutic protein targets such as cytotoxic T-lymphocyte-associated antigen (CTLA)-4, CD4, CD22, CD25 and CD33 were isolated from a phage display library with affinities in the low nanomolar range (Schlehuber and Skerra, 2005). However, detailed data about this work has not yet been published and it has to be seen how these engineered proteins will behave in experiments with animal models.

2.2.3. Affibodies

The immunoglobulin binding domain of Staphylococcal protein A (SPA) is widely used as an immunochemical reagent for the purification or detection of antibodies and also serves as affinity tag in fusion proteins (Uhlen *et al.*, 1992). The Ig-binding region of wild-type SPA consists of five highly homologous three-helix bundle domains of approximately 58 amino acid residues each. Because SPA is known as a non-cysteine containing, highly soluble, proteolytically and thermally stable protein, an engineered version of one of the SPA domains, the so-called Z-domain, was chosen as scaffold for the development of novel affinity proteins (Nilsson *et al.*, 1987) designated as "affibodies".

Utilising structural data available for the complex between a native SPA domain and the F_c fragment of human IgG1, 13 positions distributed across two helices, located at the surface of the domain and involved in the F_c interaction, were chosen for random mutagenesis, and subsequently, for the creation of phage libraries (Nord *et al.*, 1995). From medium sized libraries comprising about $5x10^7$ individual clones, which were constructed using different degenerate codons at the selected positions, affibodies to a variety of protein targets were selected and characterised (Nord *et al.*, 1997; Nord *et al.*, 2001; Rönnmark *et al.*, 2002; Sandström *et al.*, 2003). In a recent paper (Wikman *et al.*, 2004), affibody ligands binding to HER2/neu-binding were selected by phage display from a library comprising about $3x10^9$ individual clones. After four rounds of phage display selection using recombinant biotinylated HER2-ECD (extracellular domain) immobilized on streptavidin-coated magnetic beads, one of the selected affibody

variants was demonstrated to bind with nanomolar affinity (≈50 nM) to the HER2-ECD molecule in biosensor binding studies. Furthermore, the selected radiolabeled affibody showed specific binding to HER2/neu expressed on the tumour cell line SKBR-3. From competition experiments using trastuzumab (a human antibody approved in 1998), the affibody was shown to recognise a different epitope.

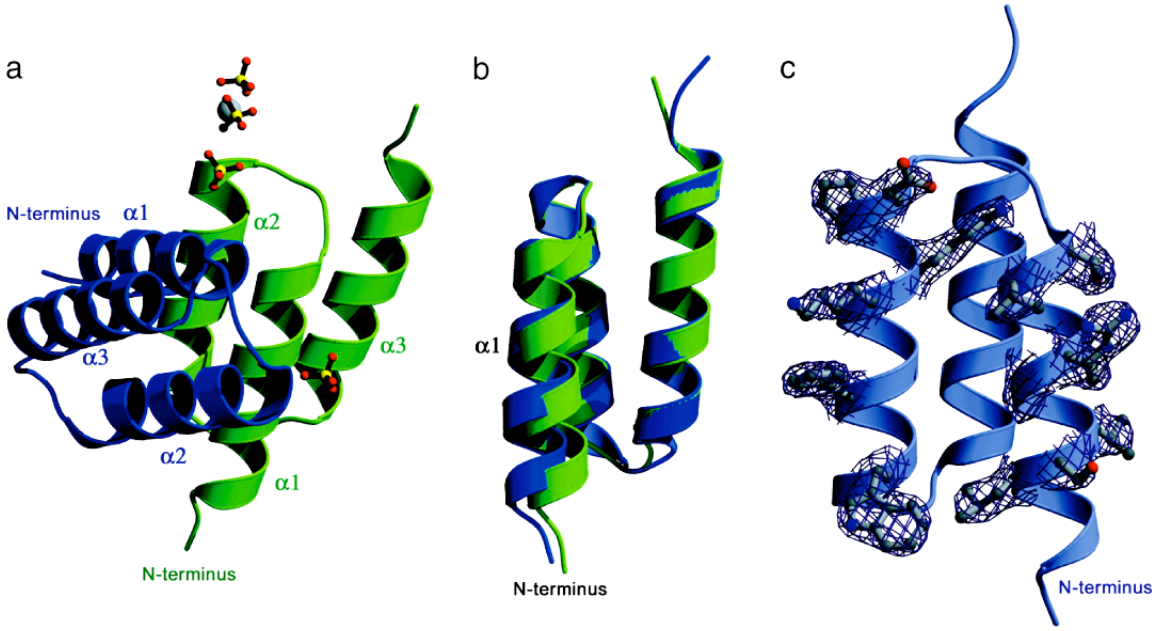

Figure 2.3. Structure of an *in vitro* evolved complex affibody (ZSPA-1) binding to the protein Z domain of Staphylococcal protein A. (*a*) Structure of the complex, the ZSPA-1 affibody in blue and protein Z in green. The ordered sulfate ions with partial occupancy and a putative magnesium ion from the mother liquor are also shown; however, they do not seem to influence the interaction surface. (*b*) Superposition of the two molecules. Notice the shift of helix 1 in the affibody (blue) compared with protein Z (green). (*c*) Electron density for all 13 mutated residues in the affibody; for clarity, only the electron density around the side chains is displayed (adapted from Hogbom *et al.*, 2003).

However, binding *per se* does not always confer a beneficial therapeutic effect to a protein, and thus, further experiments using suitable animal models will be needed in order to prove the potential of this affibody as biopharmaceutical. In addition, concerns about the potential immunogenicity of the affibody scaffold should be seriously considered, owing to its bacterial origin.

2.2.4. Trinectins

The tenth type III domain of human fibronectin (^{10}Fn3), a 94 residue (10 kDa) structural protein with an immunoglobulin-like fold was used as an antibody-mimic scaffold. High thermal stability (T_M = 90°C) and solubility (>15 mg/ml), high expression level in *E.coli* (50 mg/l), and the lack of disulfide bonds or unpaired cysteines in its structure (Xu *et al.*, 2002; Plaxco *et al.*, 1996) are attractive properties that make the ^{10}Fn3 domain a good scaffold candidate.

The first report on the use of the ^{10}Fn3 domain as a binding scaffold was published in 1998 (Koide *et al*., 1998), where a library of 10^8 distinct ^{10}Fn3 mutants was displayed on phage and used in affinity selection experiments with immobilised ubiquitin as a model antigen. After five rounds of phage display selection, clones were randomly picked for sequencing. A clone designated as Ubi4-FN3 dominated the population of selected protein mutants and was subjected to further analysis. Ubi4-FN3 was demonstrated to bind to ubiquitin in phage ELISA experiments. However, if expressed as a single domain protein in *E.coli*, Ubi4-FN3 exhibited low solubility at neutral pH and unspecific binding to the dextran matrices of the size exclusion chromatography column and the dextran coated biosensor chips used for the subsequent characterization of Ubi4-FN3.

In order to select ^{10}Fn3 domains binding to a target of interest with improved properties than Ubi4-FN3, an alternative library design in combination with a fully *in vitro* selection system called mRNA display (see chapter 2.5) was used for the isolation of ^{10}Fn3 variants, which bind to tumor necrosis factor-α (TNF-α) (Xu *et al.*, 2002). A large library with ~10^{12} individual members was created, which was displayed on mRNA (see chapter 2.5). These protein-mRNA fusions were used for affinity selection experiments with TNF-α as target antigen immobilized on sepharose. After ten rounds of selection, 7% of the fusion pool bound to TNF-α sepharose. Clones from the ninth and tenth round of selection were cloned in *E.coli*, sequenced and expressed. Affinity constants were determined by incubating the *in vitro* translated ^{35}S-methionine labeled proteins with biotinylated TNF-α at concentrations between 17 pM and 23 nM. The solution containing the ^{10}Fn3 mutants bound to TNF-α were sucked by vacuum through a membrane coated with streptavidin, therefore capturing protein complexes on the membrane. Binding was analysed by measuring the radioactivity on the

streptavidin coated membrane. In this way, K_D for the selected mutants were found in the range of 1-24 nM.

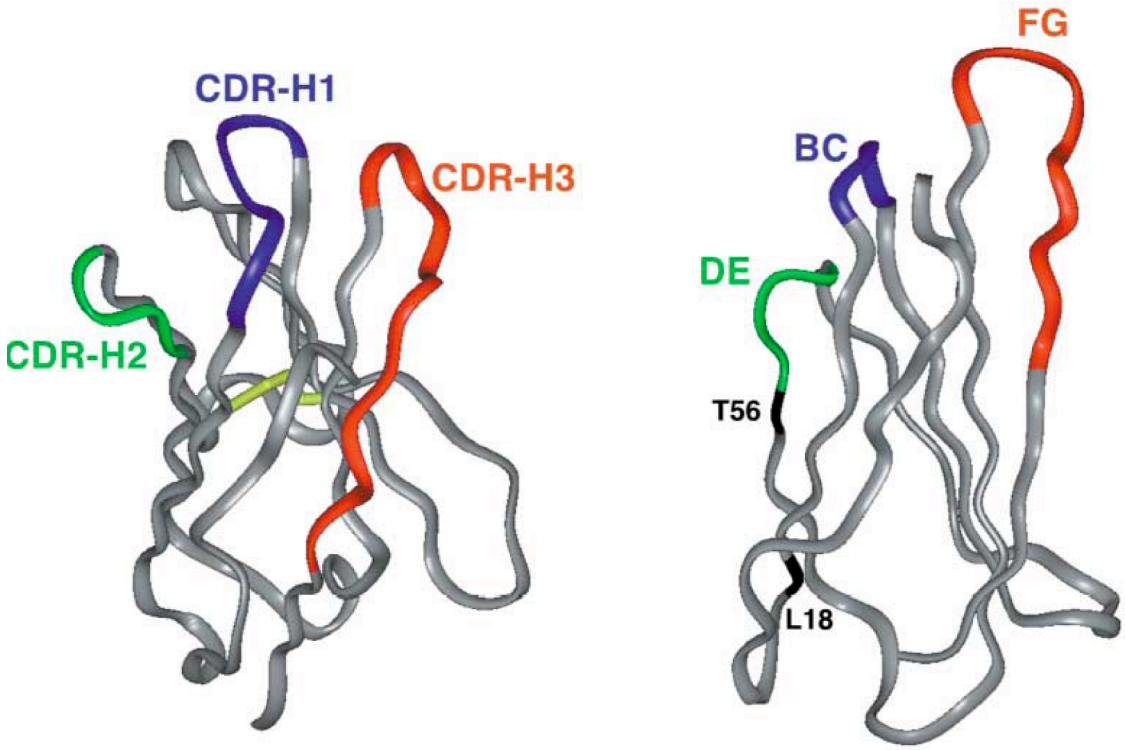

Figure 2.4. A comparison of the tertiary structures of a llama V_{HH} domain (*left*) (Spinelli *et al.*, 1996), and the wild-type, human ^{10}Fn3 domain (*right*) (Main *et al.*, 1992). Despite the lack of significant sequence identity, the V_{HH} and the ^{10}Fn3 fold into similar sheet sandwiches. The disulfide bond between Cys 22 and Cys 92 of the V_{HH} domain is shown in yellow; there are no disulfides in the ^{10}Fn3 domain. The complementarity-determining regions of the V_{HH} and the residues randomized in the ^{10}Fn3-based libraries (Xu *et al.*, 2002) are shown in color (adapted from Xu *et al.*, 2002).

In order to select ^{10}Fn3 variants with even higher affinity towards TNF-α, the population from round 8 or 10 was subjected to additional rounds of selection using soluble, biotinylated TNF-α at a concentration of 0.5 nM, therefore increasing the stringency of the selection procedure. Protein complexes were captured on streptavidin coated magnetic beads and the beads were extensively washed to remove weakly binding mutants. Characterization of the isolated clones revealed ^{10}Fn3 variants with K_D values around 100 pM, the best being 20 pM, and analytical gel filtration showed that the apparent molecular weight of

purified, soluble wild-type and mutant domains was consistent with the variants being monomeric. However, their thermodynamic stability was worse than that of the wild-type protein.

The dissociation constants published for the selected ^{10}Fn3 mutants toward TNF-α are very low. These values should be evaluated critically because first, the capture step of the biotinylated TNF-α-FN3 protein complexes was performed in a solid phase format, and second, TNF-α is a homotrimeric protein. Therefore, avidity effects probably played an important role in the determination of the dissociation constants, which implies that the real K_D values for the monomeric interaction between the ^{10}Fn3 mutants and TNF-α might be higher.
The use of engineered ^{10}Fn3 domains in therapeutic applications could be problematic. An immune response of the patient to the injected protein might generate cross-reactive antibodies binding to wild-type ^{10}Fn3, which is ubiquitous in the human body, thus leading to an autoimmune attack in the patient and/or formation of circulating immune complexes, with a potential kidney toxicity.

2.2.5. Ankyrins

Repeat proteins, such as ankyrin or leucine-rich repeat proteins, are ubiquitous binding molecules which are found in many biological processes and, unlike antibodies, occur intra- and extracellularly. They have developed a successful binding strategy because they feature repeating structural units, which stack together to form elongated protein domains (repeat domains) with an continous target-binding surface, which is variable and in size as the number of repeats can be varied (Forrer *et al.*, 2003). Residues on the surface of secondary structure elements and in loops can, depending on the type of repeat, contribute to the interaction surface. Each repeat contributes both to the stability of the domain and to the potential target-binding surface of the domain. Different classes of repeat proteins are known, such as leucine-rich repeats [e.g. ribonuclease inhibitors (Kobe and Deisenhofer, 1995) and internalins (Schubert *et al.*, 2002)], tetracorticopeptide repeats [e.g. cyclophilin (Taylor *et al.*, 2001)], armadillo/HEAT repeats [e.g. β-catenin (Huber and Weis, 2001) and importin (Vetter *et al.*, 1999)], and ankyrin repeats [ankyrin R (Michaely *et al.*, 2002)].

Ankyrin repeat proteins well illustrate the binding versatility of repeat proteins. They occur throughout all phyla and mediate protein-protein interactions in the

nucleus or cytoplasm, or while anchored to the membrane or when secreted into the extracellular space (Bork, 1993). Ankyrin repeat proteins are built from stacked, 33 amino acid repeats, each forming a β-turn that is followed by two antiparallel α-helices and a loop reaching the β-turn of the next repeat. In most known complexes, the β-turn and the first α-helix mediate the interactions with the target, and different numbers of adjacent repeats are involved in binding (Sedgewick and Smerdon, 1999). The reported target binding affinities of natural ankyrin repeat proteins are in the low nanomolar range (Malek *et al.*, 1998).

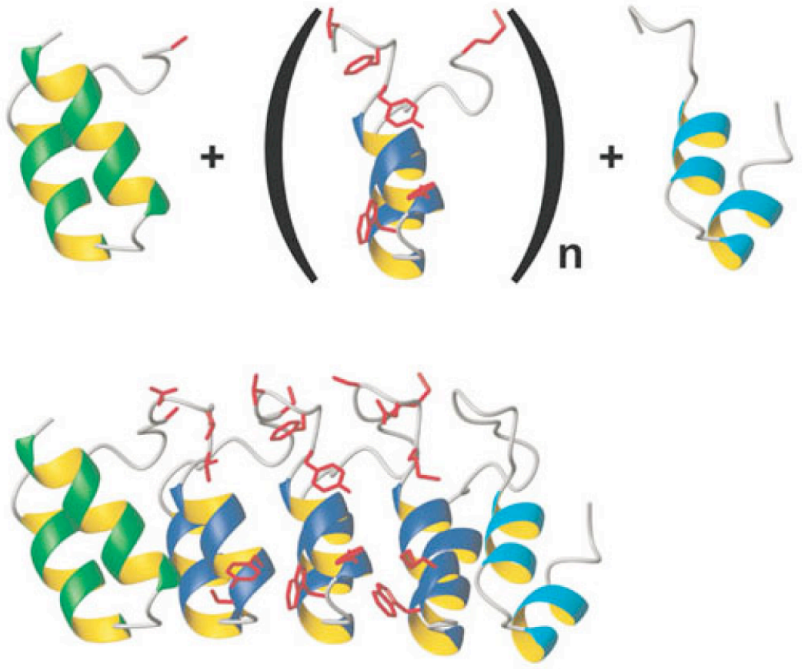

Figure 2.5. Schematic representation of the library generation of designed ankyrin repeat proteins (Binz *et al.*, 2004). (*upper part*) Combinatorial libraries of ankyrin repeat proteins were designed by assembling an N-terminal capping ankyrin (green), varying numbers of the designed ankyrin repeat module (blue) and a C-terminal capping ankyrin (cyan) (side chains of the randomized residues are shown in red). (*lower part*) Ribbon representation of the selected MBP binding ankyrin repeat protein (Binz *et al.*, 2004) (colors as above). This binder was isolated from a library consisting of a N-terminal capping ankyrin repeat, three designed ankyrin repeat modules and a C-terminal capping ankyrin repeat (adapted from Binz *et al.*, 2004).

Ankyrin repeat proteins were built and diversified to create a library from which ankyrin variants binding to maltose binding protein and two eukaryotic kinases were selected (Binz *et al.*, 2004). In the approach chosen by the authors, a consensus ankyrin repeat module consisting of six diversified potential interaction

residues and 27 framework residues was designed based on sequence alignments and structural analyses (Figure 2.5.). Varying numbers of this repeat module were cloned between capping repeats, which are special terminal repeats of ankyrin domains shielding the hydrophobic core. Two libraries were created with two and three, respectively, randomized ankyrin repeats in between an N-terminal and a C-terminal capping repeat. Using a library with more than 10^{10} individual members in combination with the ribosome display selection methodology (see chapter 2.5), ankyrin repeat variants binding to maltose binding protein and two eukaryotic kinases were isolated by performing four or five rounds of ribosome display selection. Dissociation constants were determined by surface plasmon resonance and found to be in the range of 2-20 nM. In addition, ELISA experiments showing that the selected ankyrin binders specifically recognized their target molecule and high expression levels of the ankyrin repeats in *E.coli* in shake flasks (up to 200 mg/l) make the ankyrin repeats an interesting scaffold for the development of binding proteins. Because the ankyrin repeat presented by Binz *et al.* were newly designed and are not found in nature in this format, their immunogenic potential should be investigated in detail.

2.3. Challenges in protein engineering

The advent of several selection and screening methodologies has allowed the isolation of proteins (mainly antibodies or fragments thereof) with predefined binding specificities for practically any protein or small molecule of interest. The selection of antibodies binding to a given target molecule was relatively easy because it is sufficient to introduce sequence diversity in the complementarity determining regions (CDRs) of the antibody scaffold in order to create a collection of binding specificities. However, when working with different scaffolds than immunoglobulin folds, library design is much more difficult due to following reasons. First of all, generating diversity by mutating stretches of a protein sequence or by inserting randomised sequences in a protein scaffold could influence the biophysical properties of the mutated protein, such as stability, solubility and folding. Secondly, even if a protein scaffold has been found to tolerate mutations at certain positions, it might not be possible to select a mutant protein, which specifically binds with high affinity to a target molecule owing to poor design of the potential interaction site. Therefore, it could be necessary to create a variety of libraries until an useful protein binder can be isolated.

Designing the binding specificity of a protein is not the only property of a protein that could be engineered, but interesting features such as stability, solubility, folding kinetics, and pharmacokinetics should also be considered when engineering a protein for therapeutic or industrial applications. In addition, the engineering of allosteric and catalytic proteins remains a formidable challenge due to the inherent complexity of catalytic and allosteric processes as shown in Figure 2.6.

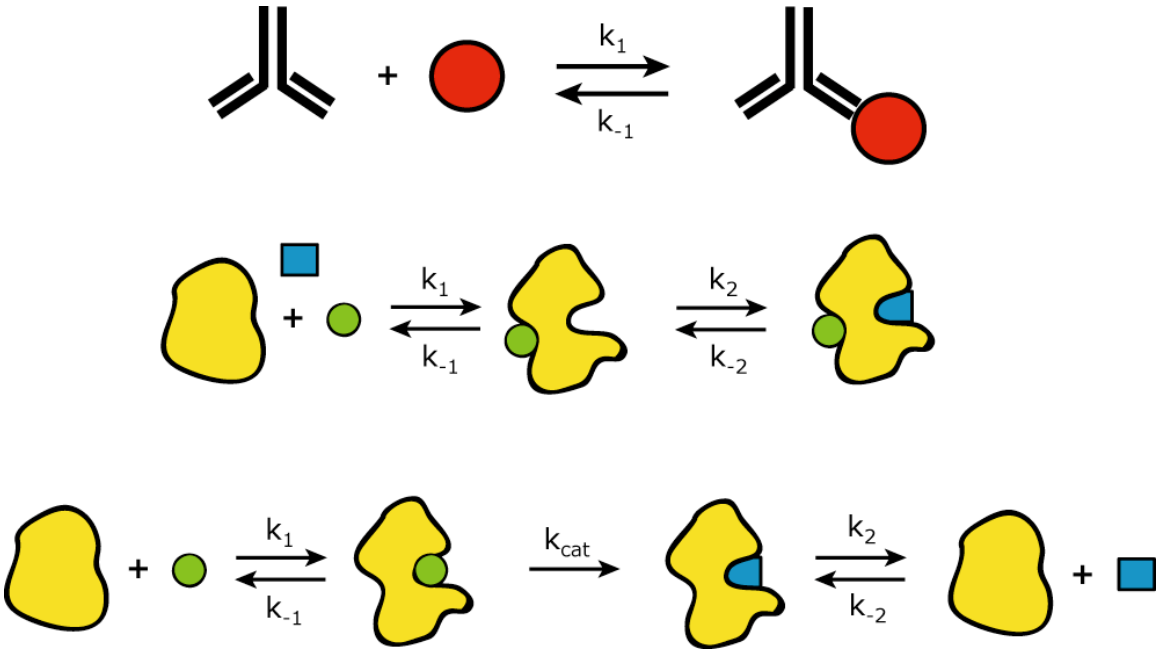

Figure 2.6. Engineering binding, allostery or catalysis. (*top*) For the isolation of a good binding protein (e.g. an antibody), the two rate constants k_1 (on rate) and k_{-1} (off rate) must be optimized. (*mid*) In the case of allostery, not only two but four rate constants must be engineered in order to obtain a protein, which binds blue ligand only in presence of green ligand. (*bottom*) Engineering catalysis is even more complex because direct selection or screening for binding is most often not suitable, and five individual rate constants must be optimized to isolate an efficient enzyme.

Therefore, recent research has been focused on the development of various methodologies, mainly for the engineering of proteins with binding or enzymatic activities. Common to all technologies is the generation of diversity of mutants, followed by a process of screening or selection to isolate interesting candidates. This procedure mimicking the natural Darwinian evolutionary process has the advantage that detailed structural information about the protein to be engineered is not needed. In contrast, so-called rational approaches, involving *in silico* modelling of the 3D structure of the protein to be engineered, require detailed

knowledge about structure-function relationships. However, proteins are extremely complex and usually in a delicate balance of functional flexibility and instability. In addition, proteins are fairly well designed, an thus, it is unfortunately far more easier to damage one than it is to improve it. And even if one was successful in designing one property (e.g. enhanced stability), it is often impossible to predict *ab initio* the cost to another (e.g. catalytic activity). These significant hurdles, which reflect our poor understanding of protein complexity can be overcome by *in vitro* evolution, if appropriate strategies for generation of diversity and screening or selection technologies are available (Arnold, 1998).

2.4. Screening methodologies

The identification and isolation of proteins exhibiting desired binding properties from large repertoires of mutant polypeptides is a daunting task. Due to the large number of different clones, individual members of the repertoire cannot be assayed one after the other due to time limitations. Therefore, methodologies have been developed, which allow the simultaneous evaluation of many different protein mutants in a short period of time. Screening methodologies represent procedures where all members of the repertoire are individually assessed in a biological assay in a highly parallel fashion, thus opening the opportunity to screen large numbers of protein candidates for desired activities. The researcher then chooses those proteins, which were positive in the screening assay.
In the field of protein engineering, the screening of a repertoire of polypeptides is complicated because, in many cases, the proteins to be screened cannot be stored for a longer time period. Therefore, they must be individually expressed in a suitable host organism harboring the genetic information just shortly before the proteins are assayed in a biological experiment. The expression of the mutant proteins is usually done in standard-sized multi-well plates, where each well contains one clone of the population. In order to facilitate the simultaneous assessment of large numbers of clones, colony filter screening assays have been proposed.

2.4.1. Iterative colony filter screening

In colony filter screening, clones (typically $<10^6$) are simultaneously assayed for their ability to generate the binding specificities of interest. A filter screening

assay was described (Gherardi *et al*, 1990) for the identification of the few clones secreting an antibody of given antigen specificity, out of several thousand hybridoma clones. This filter screening method was further developed (Skerra *et al*., 1991; Dreher *et al.*, 1991) by using a two-membrane system for detection of antigen binding by antibody Fab fragments secreted by bacterial colonies. The bacteria expressing the Fab fragments were grown on a porous master filter and secreted Fab fragments were captured on a second membrane, which was later used for detection of binding activity of the antibody fragments. After identification of antigen binding antibodies, the clones responsible for expression of the antibodies could be identified on the master filter. This screening methodology was tested in model screening experiments with two antibody clones. Further experiments showed that a modified version of this methodology could be used for the isolation of antibodies from a large library with millions of different clones (Giovannoni *et al.*, 2001) (Figure 2.7.).

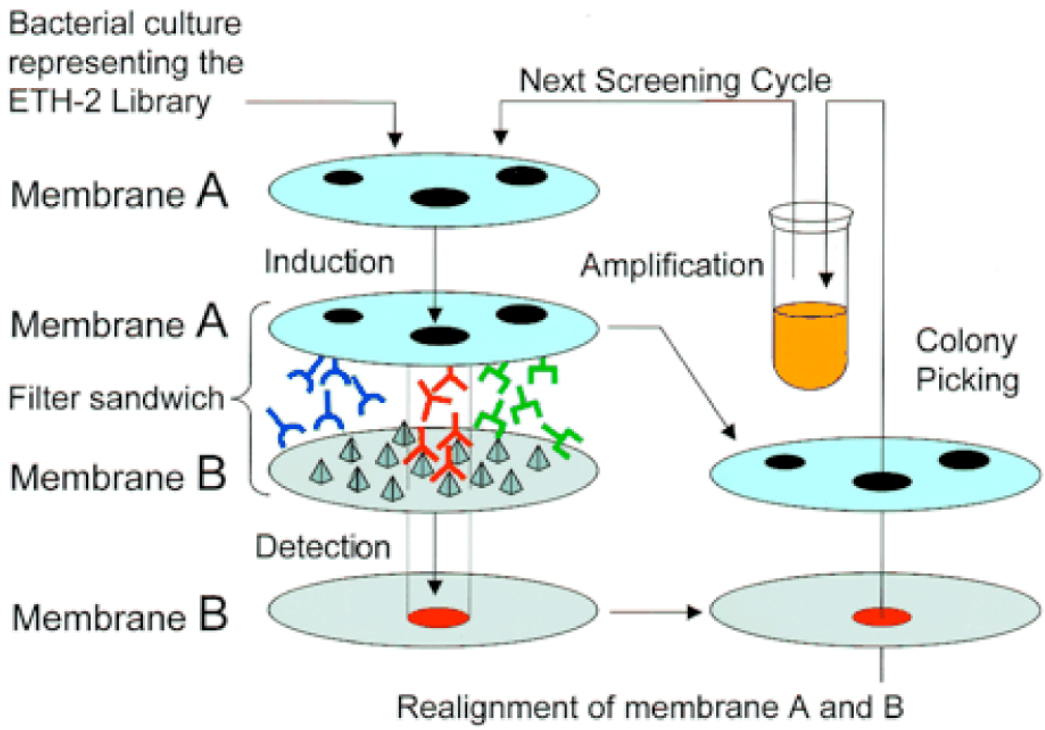

Figure 2.7. Schematic representation of the iterative colony filter screening method. Bacteria expressing potentially each a different antibody fragment were spread on a Durapore filter membrane (A). On the filter, placed on a solid medium that represses expression of the antibody fragments, colonies were visible after 8 h incubation at 37°C (here three different colonies are depicted schematically). A second filter membrane (B) was coated with the antigen of interest (represented by pyramids), laid on a solid medium capable of inducing the expression of scFvs, and in contact with membrane A. Antibody fragments that diffused from membrane A and that bound to the antigen (in red) were

captured on membrane B, and could be detected by an enzymatic colorimetric reaction. The corresponding colonies could be identified on membrane A, regrown and the procedure could be repeated until single positive colonies producing monoclonal antibodies were identified (adapted from Giovannoni *et al.*, 2001).

The authors used the naïve ETH-2 antibody library (Viti *et al.*, 2000), in the form of a pool of antibody secreting bacteria, to isolate monoclonal antibodies specific for the extra domain-B (ED-B) of fibronectin (Fattorusso *et al.*, 1999). One hundred million bacteria from the ETH-2 library, potentially expressing a different scFv antibody fragment, were grown on a porous master filter until a lawn of small confluent colonies was visible. The master filter was then laid on a capture filter embedded with the antigen (an ED-B containing fibronectin fragment), and bacteria on the master filter were induced to express the recombinant antibodies by placing the filter membranes on solid medium containing IPTG. Recombinant antibodies can diffuse from the master to the capture membrane, where the ones specifically binding the antigen are trapped onto the capture filter. A Flag-tag at the C-terminal extremity of the scFv antibody fragments allowed the detection of their binding to the antigen coated-capture membrane by an enzymatic colorimetric reaction. Due to the huge number of bacteria on the master filter, the first round of screening did not allow the direct identification of single positive bacterial colonies, but only of positive areas of confluent colonies that, by superposition of the two filters, could be rescued from the master membrane and grown to perform a further round of screening. Three rounds of screening were performed, plating 10^6 bacteria on a master filter for the second round and 1000 for the third round. Figure 2.8. shows the master (*A*) and the capture membrane after immunodetection (*B*) from the third round of screening. At this stage, colonies on the master filter were no longer confluent and approximately 30% of the plated colonies (filter A) produced scFvs that bound to filter B, therefore representing potential anti-ED-B monoclonal antibody fragments (Giovannoni *et al.*, 2001).

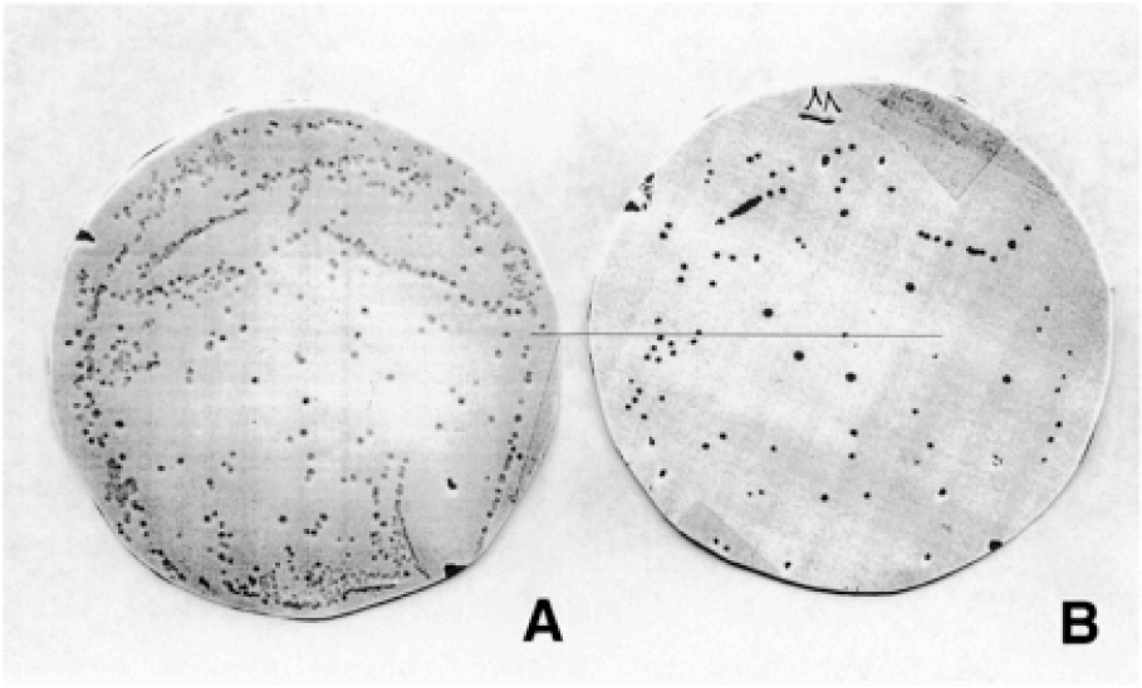

Figure 2.8. Anti ED-B antibodies identified after three rounds of colony filter screening. (*A*) The filter membrane, where 10^3 bacterial clones were plated and which were rescued from the second round of iterative filter screening. (*B*) The capture filter membrane, coated with the recombinant antigen, onto which antibodies secreted by the colonies grown on filter (*A*) could diffuse. The clones secreting antibodies binding to antigen were detected by enzymatic colorimetric reaction. Around 30% of the colonies in (*A*) result positive in (*B*) (adapted from Giovannoni *et al.*, 2001).

The positive clones at the third round of filter screening were regrown and the binding specificity of the corresponding antibodies was analysed by ELISA. 16% produced antibodies that bound specifically to the ED-B domain of fibronectin and the best clone, named scFv ME4C, was further characterised. The affinity of scFv(ME4C) towards the ED-B domain of fibronectin was measured by real-time interaction analysis using surface plasmon resonance detection, yielding a K_D = $1x10^{-7}$ M^{-1} (Giovannoni *et al.*, 2001).

Two general methodologies based on filter-sandwich assays were described by Heinis *et al.* (Heinis *et al.*, 2002) for isolating enzymatic activities from a large repertoire of protein variants expressed in the cytoplasm of *E. coli* cells. The enzymes were released by the freezing and thawing of bacterial colonies grown on a porous master filter and diffused to a second "reaction" filter that closely contacted the master filter. Reaction substrates were immobilized either on the filter or on the enzyme itself (which was, in the latter case, captured on the

reaction filter). The resulting products were detected with suitable affinity reagents. Biotin ligase was used as a model enzyme to assess the performance of the two methodologies. Active enzymes were released by the bacteria, locally biotinylated the immobilized target substrate peptide, and allowed the sensitive and specific detection of individual catalytically active colonies.

2.5. Selection methodologies

Even though iterative colony filter screening has been successfully used for the isolation of monoclonal human antibodies binding to ED-B from a population of 10^7 bacteria (Giovannoni *et al.*, 2001), there is an interest to work with protein libraries which are larger than 10^7 by several orders of magnitudes. It is now generally believed both for libraries of small organic molecules and proteins, that the expected number of hits with desired properties present in a library increases with the size of the library (Moore, 2003; Kurtzmann *et al.*, 2001).
For this purpose, selection methodologies have been described mimicking the natural process of evolution in a test tube in the laboratory. Common to all selection technologies is the linkage of the genetic information (genotype) with the encoded polypeptide (phenotype). This allows the amplification of the genetic information of the isolated protein mutants, leading to the survival of genotypes which code for favorable polypeptides. By repeating the two steps of generating genetic diversity by means of random mutagenesis followed by a procedure, which favors the survival of only those polypeptides, which fulfill the requirements of the researcher, proteins can be evolved *in vitro*.

There are two main strategies to link a genotype to the corresponding protein phenotype. The first possibility is to link the protein physically to the DNA encoding it. This type of linkage is fundamentally different from a second avenue for linking phenotype and genotype, based on compartmentalisation. This concept is commonly found in nature, where genes, the corresponding encoded proteins and the products of their activity, are all held together in a cell. Most procedures applied in protein engineering use the physical linkage between genotype and phenotype.

2.5.1. Phage display

For the last decade, phage display has been one of the most widely used selection methodology (Winter *et al.*, 1994). In phage display, the polypeptides to be engineered are displayed on the surface of filamentous bacteriophage by fusion to the minor coat protein pIII, whereas the genetic information encoding the displayed protein is packaged inside the phage coat. In 2003, the first phage derived human monoclonal antibody was approved by the FDA for the treatment of rheumatoid arthritis.

Filamentous phage (*Inovirus*) comprise a family of organisms that have only about 10 genes and grow in well-characterised hosts, the Gram-negative bacteria. The relative simplicity of these viruses and the ease with which they can be genetically manipulated have made them a powerful tool in biotechnology. Many different strains have been isolated. The wild-type virion is a rod-shaped structure of about 6 nm in diameter and 800-2000 nm long (depending on the strain), made up of a sheath of several thousand α-helical proteins in a helical array around a single-stranded circular DNA molecule at the core (Marvin, 1998). A few other proteins form the cap at each end of the virion. Upon infection, viral DNA enters the cell and is converted to a double-stranded replicative form by host enzymes. Progeny DNA is replicated by a rolling circle mechanism, and is assembled with the viral protein pV into a intracellular nucleoprotein complex. Protein pV coating the viral DNA is replaced by capsid protein at the cell membrane, where the virions are exported.

When foreign DNA is inserted as a separate gene in a viral intergenic region, the virus-derived plasmid can become a cloning vector. When the foreign DNA is fused in frame to a gene encoding a viral coat protein, the virion can display foreign peptides on its surface. The viral coat protein pIII has mainly been used as a fusion partner for the display of peptides and proteins on the surface of filamentous phage (Smith, 1985). pIII is present at the tip of the phage as a capping protein in probably three to five copies per phage particle. Folded antibody fragments can be displayed on phage as pIII fusions, either as single chain Fv fragments (V_H and V_L domains are connected on the same polypeptide chain by a flexible polypeptide spacer) or Fab fragments (one chain is fused to pIII and the other is secreted into the periplasm) (Winter *et al.*, 1994). In this way, a physical connection between the displayed protein (phenotype) and the genetic information packaged inside the phage particle is created. If a repertoire

of protein mutants is displayed on phage, rare phage particles displaying a protein variant which binds to an immobilized target molecule can be enriched by affinity selection (Figure 2.9.).

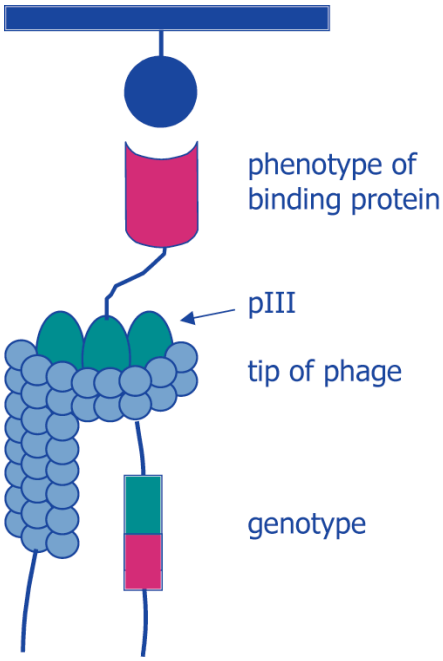

Figure 2.9. Proteins can be displayed on the surface of filamentous phage as N-terminal fusions of the minor coat protein pIII. The genetic information (genotype) encoding the fusion protein is packaged inside the phage particle, therefore creating a physical link between phenotype and genotype. If a repertoire of protein mutants is displayed on phage, rare phage particles displaying a protein variant which binds to an immobilized target molecule (*blue*) can be enriched by affinity selection.

The pIII fusion and other proteins of the phage can be encoded entirely within the same phage replicon or on different replicons. When two replicons are used, the pIII fusion is encoded on a phagemid, a plasmid containing a phage origin of replication and an antibiotic resistance gene. Phagemids can be packaged into phage particles by superinfection with helper phage such as M13K07 or VCSM13 that provide all the phage proteins needed for phage assembly (including pIII), but due to a defective origin of replication is itself poorly packaged in competition with the phagemids (Winter *et al.*, 1994).

The pIII fusion is often proteolysed, as shown by Western blotting of the phage proteins and detection with anti-pIII antibody (Silacci *et al.*, 2005). This is expected to give a population of phage particles, each displaying zero, one, two, three (or perhaps even four and five) antibody fragments. The average valency of display in the population is further reduced by use of helper phage, in which the

helper pIII competes for incorporation into the phage particle. Such phage have been estimated on average to display less than a single fusion protein per particle; they have been called "monovalent" phage (Winter *et al.*, 1994). Helper phage, called hyperphage, that lack pIII have been designed to rescue phage particles that incorporate only the pIII fusion encoded by the phagemid (Rondot *et al.*, 2001). These phage particles are therefore multivalent.

Phage displaying a repertoire of antibodies can be selected by binding to antigen coated plates, columns, cells or to biotinylated antigen in solution followed by capture onto solid phase streptavidin. After washing, selected phage can be eluted under alkaline conditions and used for the infection of *E.coli*, which will produce new phage particles that can be subjected to further rounds of selection (Figure 2.10.). In this way, enrichment factors of about 100 can lead to a million fold enrichments after only three rounds of selection.

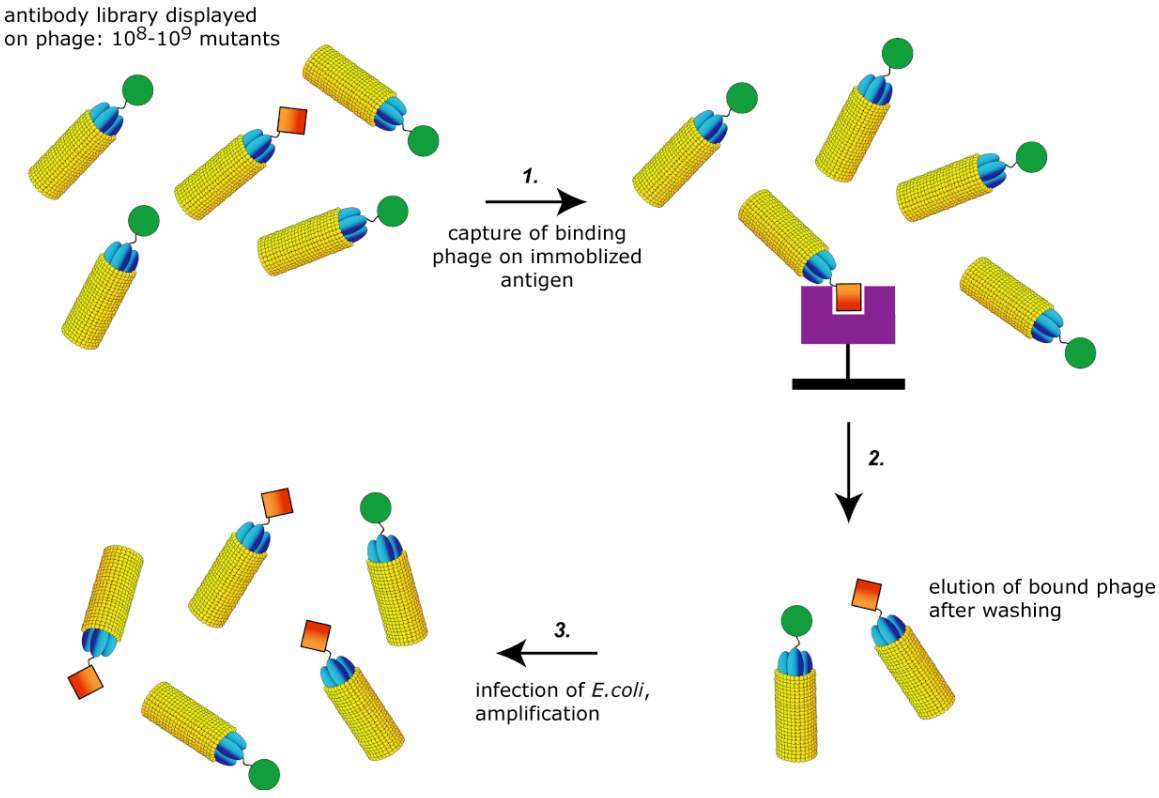

Figure 2.10. Selection cycle for the isolation of binding proteins from phage display libraries. A library of proteins displayed on the phage surface (e.g. antibody fragments) are used as input for the selection. Phage displaying a binding protein are captured on immobilised target molecules, and after washing, bound phage can be eluted. This phage population is then propagated in bacterial cultures after infection of *E.coli* cells and can be used for further rounds of selection.

The efficiency of selection is likely to depend on many factors, including the kinetics of dissociation during washing, and whether multiple antibody fragments on a single phage can be simultaneously engage with antigen immobilised on solid support. For example, antibodies with fast dissociation kinetics (and weak binding affinities) should be retained by use of short washes, multivalent display and a high coating density of antigen at the solid phase. The high density should not only stabilize the phage through multivalent interactions, but favor rebinding of phage that has dissociated. Nevertheless, it appears that binding affinities for single antibody fragments of 10^5 M^{-1} are barely sufficient to hold multivalent phage to solid phase (Winter *et al.*, 1994).
In contrast, the selection of antibodies with slow dissociation kinetics (and good binding affinities) should be promoted by use of long washes, monovalent phages and low coating density of antigen. For the elution of phage displaying an antibody with very high affinity ($>10^{10}$ M^{-1}), harsh conditions such as low or high pH must be applied.

Not only antibody fragments, but also other proteins and enzymes have been successfully displayed and selected on phage, such as domain antibodies (dAbs) (van den Beucken *et al.*, 2001), lipocalins (Beste *et al.*, 1999), affibodies (Wikman *et al.*, 2004), biotin ligase and trypsin (Heinis *et al.*, 2001), as well as subtiligase (Atwell and Wells, 1999). However, not every protein can be displayed on phage as fusion to pIII. The main two reasons for inefficient display on phage are proteolytic cleavage of the pIII fusions and poor incorporation of the pIII fusion protein into the phage coat, because of a competition with the pIII protein of the helper phage. In order to obtain a sufficient number of phage particles displaying the fusion protein, high titers of phage are needed, thus limiting the diversity of protein mutants that can be accessed for selection. A strategy to improve the display of proteins on phage was published by Jestin *et al.* (Jestin *et al.*, 2001). In this work, optimized leader peptides were selected, which are cloned upstream of pIII fusion proteins and serve as a signal for the export of pIII fusions to the bacterial periplasm.
An additional drawback of phage display is the need for transformation of bacterial cells by electroporation when cloning a library because the limiting factor in making large primary libraries is the efficiency of introduction of plasmid or phage DNA into bacteria. This limits the size of phage display libraries to 10^{10} – 10^{11} (Rauchenberger *et al.,* 2003).

2.5.2. Yeast display

Yeast display of antibody fragments has proven to be an efficient methodology for the directed evolution of single chain Fv (scFv) antibodies for increased affinity and thermal stability and, more recently, for the display and screening of a non-immune scFv and immune Fab libraries. A major strength of yeast display as a antibody discovery platform is the ability to characterize the binding properties such as the affinity and epitope binding characteristics of a clone without the need for subcloning, expression and purification of the scFv. An additional strength of yeast display of antibody fragments is its compatibility with fluorescent-activated cell sorting (FACS). FACS (Ibrahim and van den Engh, 2003) has become an indispensable tool in biology with recent technical advances allowing fine resolution between cell types by measuring many optical parameters in parallel. Both yeast and bacterial surface display (see chapter 2.5.3.) could be seen as screening and not as selection methodologies because, by using flow cytometry, each clone of the library is screened one by one. However, when working with large libraries displayed on the surface of yeast cells or bacteria, the huge number of cells cannot be handled anymore by a flow cytometer. If this is the case, magnetic beads coated with antigen are used for the first one or two rounds of selection followed by screenings with FACS (Feldhaus and Siegel, 2004).

Yeast display uses the a-agglutinin yeast adhesion receptor to display recombinant proteins on the surface of *Saccharomyces cerevisiae* (Boder and Wittrup, 1997). In *S. cerevisiae*, the a-agglutinin receptor acts as an adhesion molecule to stabilise cell–cell interactions and facilitate fusion between mating *a* and α haploid yeast cells. The receptor consists of two proteins, Aga1 and Aga2. Aga1 is secreted from the cell and becomes covalently attached to b-glucan in the extracellular matrix of the yeast cell wall. Aga2 binds to Aga1 through two disulfide bonds, and after secretion remains attached to the cell via Aga1 (Lu *et al.*, 1995). The yeast display system takes advantage of the association of Aga1 and Aga2 proteins to display a recombinant protein or scFv on the yeast cell surface (Schreuder *et al.*, 1996). It has been proposed that as a eukaryote, *S. cerevisiae* offers an advantage of posttranslational modifications and processing of mammalian proteins, and therefore, is better suited for expression and secretion of human derived antibody fragments than a host such as *Escherichia coli* (Boder and Wittrup, 1997). The gene of interest is cloned into a vector as an in frame fusion with the AGA2 gene. Expression of both the Aga2 fusion protein

from the vector and the Aga1 protein in the host strain is regulated by *GAL1*, a tightly regulated promoter, that does not allow any detectable scFv expression in absence of the inducer molecule (galactose). In fact, this degree of regulation permits expansion of a scFv library 10^{10}-fold without any recognisable changes in either the percentage of antibody expression or the frequency of specific clones within the library (Feldhaus *et al.*, 2003). Upon induction with galactose, the Aga1 protein and the Aga2–scFv fusion protein associate within the secretory pathway, and the epitope-tagged scFv antibody is displayed on the cell surface at 10,000–100,000 copies per cell (Feldhaus and Siegel, 2004) (Figure 2.11.).

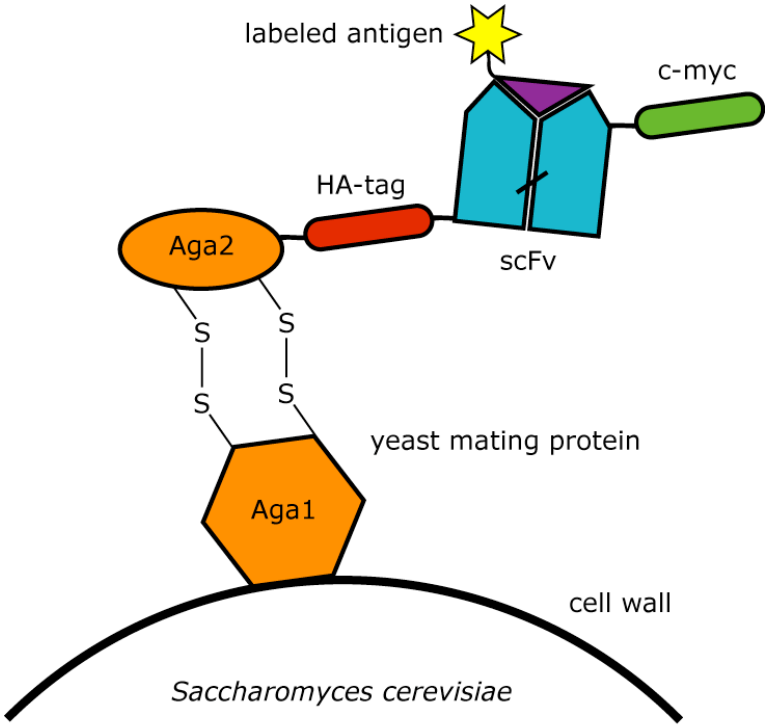

Figure 2.11. The scFv Aga2 fusion protein surface expression system. Aga1 is bound to a cell wall glucan and connected by a disulfide bond to Aga2. The protein to be displayed is cloned in frame with Aga2 protein. Using suitable antibodies, N-terminal hemagglutinin (HA) tag and C-terminal c-myc tag allow the monitoring of fusion protein expression. By addition of labeled antigen, yeast cells displaying antibody fragments binding the antigen can be isolated by affinity purification (e.g. biotinylated antigen) or FACS (e.g. fluorescently labeled antigen).

ScFv antibody expression on the yeast cell surface can be monitored by flow cytometry with fluorescently labeled antibodies recognizing either the C-terminal c-myc and N-terminal HA epitope tags encoded by the display vector. The extracellular surface display of scFv by *S. cerevisiae* also allows the detection of

appropriately labeled antigen–antibody interactions by flow cytometry. Flow cytometric representation of a highly enriched, antigen-binding scFv library displayed on yeast is shown in Figure 2.12. Antigen-binding interactions with the scFv antibody are easily visualised by either direct or indirect fluorescent labeling of the antigen of interest. In a suitable concentration range, the intensity of the fluorescent signal for antigen is related to the binding affinity (Figure 2.12.).

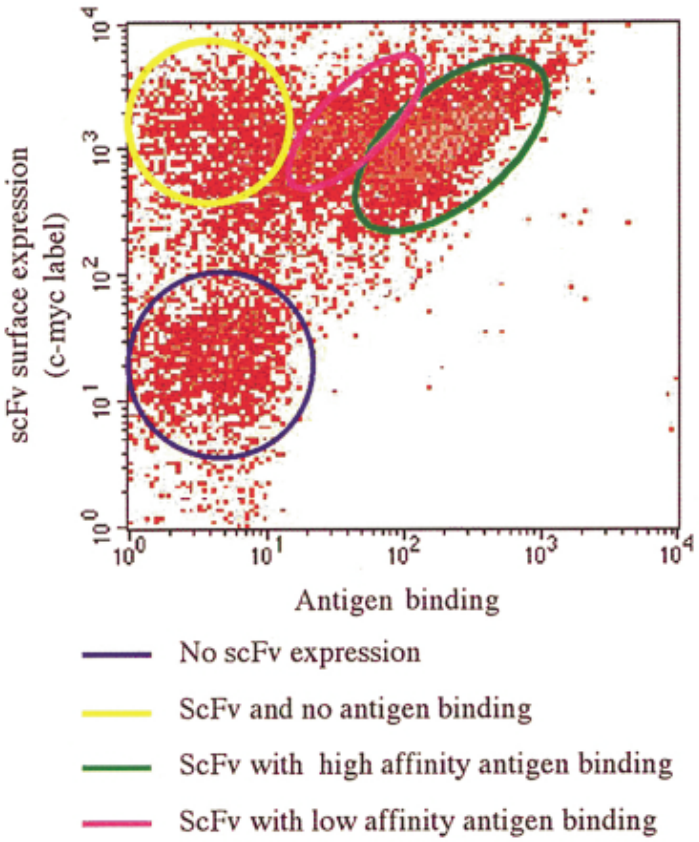

Figure 2.12. Flow cytometric analysis of yeast-displayed antibody fragments in a population, which was greatly enriched for antigen binding. Yeast cells were labeled with biotinylated antigen/streptavidin-phycoerythrin (detection of bound antigen), and anti-c-myc/anti-mouse-FITC (monitoring of scFv display level) labels. The subpopulation of yeast cells in the blue circle displays only few scFv fragments, which do not bind a significant amount of antigen. Despite good display of scFv, the subpopulation in the yellow circle does not bind to the antigen. The subpopulation in the red circle represents clones with high levels of scFv display, but rather poor affinity towards the antigen. The most interesting subpopulation is present in the green circle, where the yeast cells display many scFv with high binding affinity to the antigen (adapted from Feldhaus *et al.*, 2003).

Yeast display has some unique strengths as a platform for affinity reagent discovery and optimization. Enrichment of clones displaying antibodies binding to the target molecule can be achieved by multiple rounds of enrichment on a cell

sorter, magnetic or flow cytometry based. Using FACS, the discrimination between clones with different affinities for the antigen can be performed during selection, which facilitates the separation of high affinity clones from low affinity clones (see Figure 2.12.). In addition, FACS allows the selection to be performed in solution, and thus, antigen has not to be immobilised on a solid support, which could cause strong avidity effects. Selections in solution allow the investigator to precisely control the concentration of the antigen and establish a lower affinity threshold preventing the accumulation of low affinity clones. However, in cases where the antigen is not monovalent, strong avidity effects may come into play due to dense display of scFv on the cell wall of the yeast cell (Feldhaus and Siegel, 2004).

When working with large antibody libraries comprising 10^9 members, for most antigens, a combination of two rounds of selection using magnetic particles followed by two rounds of flow cytometric sorting will generally allow recovery of clones of interest. The large size of the non-immune library imposes the use of an alternative method to enrich antigen-binding clones in the initial rounds of the selection in order to reduce the overall complexity of the library prior to using FACS. For example, 10^{10} cells (10x library coverage) are required to have statistical confidence of representing 90% of the clones in a 10^9 diverse library at least once. This number of cells becomes prohibitive for flow cytometry with even the fastest of cell sorters; however, sorting the expected 10^7–10^8 output from the second round of selection is easily accomplished (Feldhaus and Siegel, 2004).

Clone characterisation is a time-consuming and labor-intensive step in any antibody discovery process. Characterisations may include: dissociation constant determination (K_D), determination of off-rate (k_{off}) as well as on-rate (k_{on}) constants and stability analysis. Yeast display is well suited for these analytical tasks, as the binding properties of multiple individually isolated scFv fragments can be rapidly and quantitatively determined directly on the yeast surface using flow cytometry, obviating the need for subcloning, expression, and purification. Affinities are determined using equilibrium-based kinetic analysis by measuring the degree of binding over a wide range of antigen concentrations. The off rate constant (k_{off}) is determined by measuring the degree of antigen binding remaining after saturation at various time points in the presence of excess volume and/or excess (100x) of unlabeled antigen. With these two measurements, the k_{on} rate constant can be calculated. Stability can be determined by measuring the ability to bind antigen after exposure of the scFv displayed on the cell surface to multiple temperatures (Orr *et al.*, 2003).

An additional possibility of the yeast platform is the ability not only to identify scFv antibodies that can recognize a particular antigen, but also antibody pairs that will work well in ELISA-type ''sandwich'' assays. In this type of selection, an initial antibody (referred to as ''detection Ab'') is already available to the antigen of interest. This detection antibody can then be used to identify novel yeast clone(s) displaying a suitable ''capture'' antibody on their surface, one that can bind the antigen at an epitope that is not sterically blocked in the presence of the already known detection antibody.

There are several examples of antibodies with extremely high affinities, which have been affinity matured by yeast display. Boder *et al.* (Boder *et al.*, 2000) used a scFv fragment called 4-4-20 recognising the hapten fluorescein for affinity maturation. The 4-4-20 scFv antibody fragment open reading frame was mutagenised by error-prone DNA shuffling (Stemmer, 1994). Fluorescently labeled clones exhibiting slowed antibody-hapten dissociation kinetic constants were identified and isolated by flow cytometry with optimal screening and sorting conditions. Up to 20 improved clones were selected randomly for individual measurement of the dissociation rate constant k_{diss}, and 10 improved clones exhibiting the widest range of k_{diss} values were selected for further analysis. The complete collection of isolated mutants was then recombined by modified DNA shuffling, together with further error-prone PCR mutagenesis. This cycle of mutagenesis and screening was repeated three times, resulting in mutant scFvs with dissociation rate constants over four orders of magnitude slower than the k_{diss} of the parental scFv 4-4-20. These *in vitro* evolved single-chain antibody variants have a very low antigen binding equilibrium dissociation constant (K_D = 48 fM) and slower dissociation kinetics (half-time >5 days) than those for the streptavidin-biotin complex.

In a recent study, the directed evolution of an anti-carcinoembryonic antigen (CEA) scFv with a 4-day monovalent dissociation half-time at 37°C was published (Graff *et al.*, 2004). A scFv antibody fragment designated as MFE-23 (Chester *et al.*, 1994) was first humanised by replacing several amino acid residues (hMFE-23) and then affinity matured by tow rounds of mutagenesis and screening of yeast surface-displayed libraries. Several variants of hMFE-23 were isolated which showed 10-, 100-, and 1000-fold improvement in the off-rate over the original scFv. The greatest improvement corresponded to a half-life for binding to CEA of 4-7 days at 37°C (versus 10 min for the parental antibody hMFE-23). This is the

slowest reported dissociation rate constant engineered for an antibody against a protein antigen (Graff *et al.*, 2004).

Using yeast display, a single-domain intracellular antibody devoid of disulfide bonds was engineered (Colby *et al.*, 2004b). In a former publication, a scFv antibody specific for the first 20 amino acid residues of huntingtin (htt) protein was reduced to a single variable light-chain (V_L), to enable intracellular expression and mild inhibition of the Huntington's disease caused by the aggregation of htt protein (Colby *et al.*, 2004a). This V_L domain antibody was then engineered to potently inhibit htt aggregation and cytotoxicity by removing the disulfide bond to make intrabody properties independent of redox environment present during expression (Colby *et al*., 2004b). First, the cysteine residues that form the intradomain disulfide bond were mutated, which resulted in an unexpectedly large decrease in the intrabody's affinity for its antigen. Iterative rounds of mutation and screening by FACS were then applied to improve the intrabody's affinity for the huntingtin protein, and the ability to block htt aggregation correlated with antigen-binding affinity in the absence of disulfide bonds.

In summary, yeast display has been used for affinity maturation and the isolation of monoclonal antibodies from naïve libraries. The ability to measure biophysical properties such as stability, expression levels and affinity to the antigen directly on the yeast surface, makes yeast display probably one of the cleanest methodologies for affinity maturation of antibodies. However, the transformation of yeast cells limits library size to about 10^9 individual members, and yeast cells are less resistant to harsh experimental conditions if compared with filamentous phages (Winter *et al.*, 1994).

2.5.3. Bacterial display

Several fusion protein strategies for the display of relatively short peptides on the surface of Gram-negative bacteria have been described (Georgiou *et al.*, 1997). Short peptides of less than 60 amino acid residues can be displayed on the cell surface when fused into surface exposed loops of outer membrane proteins (OMPs) from bacteria. However, it was quickly realised that the insertion of peptides longer than 60 amino acids perturbed the overall conformation and assembly of the carrier protein, therefore interfering with the localization of the

fusion proteins (Georgiou *et al.,* 1997). These limitations have been addressed by constructing an Lpp-OmpA hybrid display vehicle consisting of the N-terminal outer membrane localization signal from the major lipoprotein (Lpp) fused to a domain from the outer membrane protein OmpA. OmpA mediates the display of proteins fused to the C-terminus of the Lpp-OmpA hybrid. Lpp-OmpA fusions have been used to successfully display on the surface of *E.coli* several proteins with varying size, including scFv antibody fragments (Francisco *et al.*, 1993). An anti-digoxin scFv displayed as Lpp-OmpA fusion on the surface of *E.coli* could be enriched from a pool of irrelevant control cells by FACS with fluorescently labeled hapten (Francisco *et al.*, 1993).

In a recent paper (Bessette *et al.*, 2004) a large library ($5x10^{10}$ clones) was constructed composed of random 15-mer peptide insertions constrained within a flexible, surface exposed loop of the *Escherichia coli* outer membrane protein A (OmpA). The library was screened for binding to five unrelated proteins, including targets previously used in phage display selections: human serum albumin, anti-T7 epitope mAb, human C-reactive protein, HIV-1 GP120 and streptavidin. Two to four rounds of enrichment were needed to enrich peptide ligands specifically binding to each of the target proteins.

In a different approach, a "display-less" system was developed for proteins that bind to small molecules, including peptides and oligonucleotides, called periplasmic expression with cytometric screening (PECS) (Chen *et al.*, 2001). Libraries of scFvs were expressed in soluble form in the periplasmic space of *E. coli* by means of a suitable signal sequence, and a fluorescently tagged ligand was added to the population of *E.coli* cells. Normally, the outer membrane of Gram-negative bacteria restricts the diffusion of molecules larger than ~650 Da. However, the authors found that an appropriate combination of growth conditions and bacterial strains permitted molecules as large as 10 kDa to equilibrate through the outer membrane while periplasmically expressed proteins are retained within the periplasma of the cells. Specific binding of the fluorescent conjugate by the cognate periplasmic protein resulted in increased fluorescence, allowing the cells to be isolated from the rest of the library by fluorescence-activated cell sorting (FACS). Thus, the cell envelope could serve as a dialysis bag that retained protein–ligand complexes. Applying this methodology, an scFv antibody fragment with sub-nanomolar affinity to digoxigenin could be isolated.

In an attempt to combine bacterial display with phage display technology and to circumvent the limitations imposed by PECS (large antigens such as proteins cannot be used in PECS because conditions that allow the accessibility of high molecular weight species to the recombinant scFv also would result in the destruction of the scFv linkage to the cell), Harvey *et al*. developed a technology designated as anchored periplasmic expression (APEx) (Harvey *et al.*, 2004).

In APEx, proteins are expressed in the periplasm, tethered to the inner membrane of *E. coli* via lipidation of a small N-terminal 6-aa fusion or as a C-terminal fusion to the N-terminus of the M13 phage pIII coat protein (g3p), for flow cytometric analysis of clones selected by phage display without further subcloning. After chemical enzymatic permeabilization of the bacterial outer membrane, *E. coli* cells expressing anchored scFv antibodies can be specifically labeled with fluorescent antigens, ranging in size up to at least 240 kDa, and analyzed by flow cytometry (Figure 2.13.).

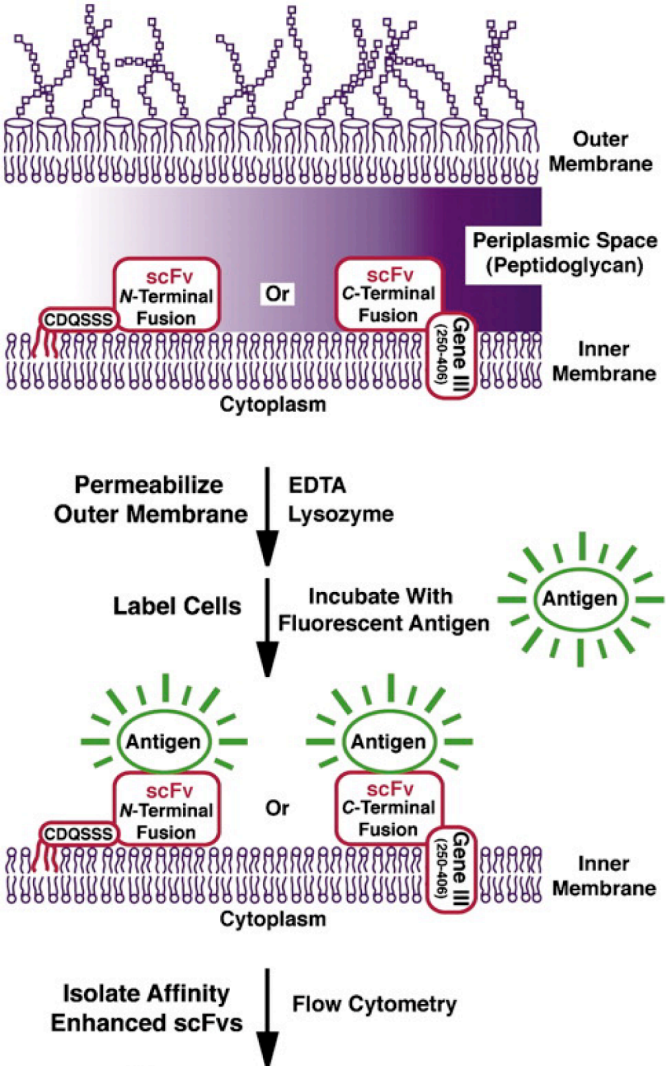

Figure 2.13. Schematic representation showing the principle of APEx. scFv antibody fragments are anchored to the inner membrane of an *E.coli* cell either as N-terminal fusion to a short peptide or a C-terminal fusion to the minor coat protein pIII of filamentous phage. After permeabilisation of the outer membrane, the displayed scFvs are accessible from the surrounding solution, and thus, can be labeled with fluorescent antigen (adapted from Harvey *et al.*, 2004).

By using APEx, antibodies with improved ligand affinities, including an antibody fragment to the protective antigen (PA) of *Bacillus anthracis* with an 200-fold affinity gain were isolated. Furthermore, fusions between GFP and antigen were expressed endogenously and captured by periplasmically anchored scFv. Thus, after a washing step, cells that express both the fluorescent antigen and an APEx-anchored scFv were highly fluorescent and could be readily sorted from cells that expressed either only an scFv or GFP-antigen fusion alone. This feature might be

particularly useful when searching genomes for interacting pairs of proteins. A disadvantage of APEx is the low viability of *E.coli* cells, which have been depleted of their outer membrane, after flow cytometry. Therefore, the genetic information coding for the selected scFvs must be amplified by PCR and cloned again for a subsequent round of screening or selection.

Bacterial display, in contrast to phage or yeast display, has not found widespread use, probably due to problems with the display of proteins and steric interference with the lipopolysaccharide layer of *E.coli*, which may impede the binding of displayed proteins to large macromolecular antigens such as proteins.

2.5.4. Ribosome display

The screening and display methodologies mentioned so far all have in common that library size is limited to $\sim 10^{10}$ individual variants due to the initial transformation of living cells with library DNA. In addition, the cloning of libraries larger than 10^9 clones can consume a considerable amount of time and work. In order to circumvent this problem, selection methodologies have been proposed which, by translating the genetic information into polypeptides outside living cells, can take place fully *in vitro*. This not only opened the possibility to work with very large repertoires ($>10^{12}$ members), but also shortened significantly the time needed to generate sequence diversity, thereby allowing to repeatedly introduce new mutations after each round of selection.

Ribosome display was first described by Mattheakis *et al.* (Mattheakis *et al.*, 1994), who displayed short peptides on polysomes. Considerable improvements were needed to allow the display and selection of folded proteins (Hanes and Plückthun, 1997), such as scFv antibody fragments with dissociation constants as low as 10^{-11} M (Hanes *et al.*, 1998). In ribosome display, a DNA library coding for particular proteins, for instance scFv fragments of antibodies, is transcribed *in vitro*. The mRNA is purified and used for *in vitro* translation. As the mRNA lacks a stop codon, the ribosome stalls at the end of the mRNA, giving rise to a ternary complex of mRNA, ribosome and functional protein. After *in vitro* translation, the ribosomal complexes, which are stabilised by high concentrations of magnesium ions and low temperature, are directly used for selection either on a ligand immobilised on a surface or in solution, with the bound ribosomal complexes subsequently being captured with, e.g., magnetic beads. The mRNA incorporated

in the bound ribosomal complexes is eluted by addition of EDTA, purified, reverse-transcribed and amplified by PCR. During the PCR step, the T7 promoter and the Shine–Dalgarno sequence are reintroduced by appropriate primers. Therefore, the PCR product can be directly used for further selection cycles. Ribosome display is schematically depicted in Figure 2.14.

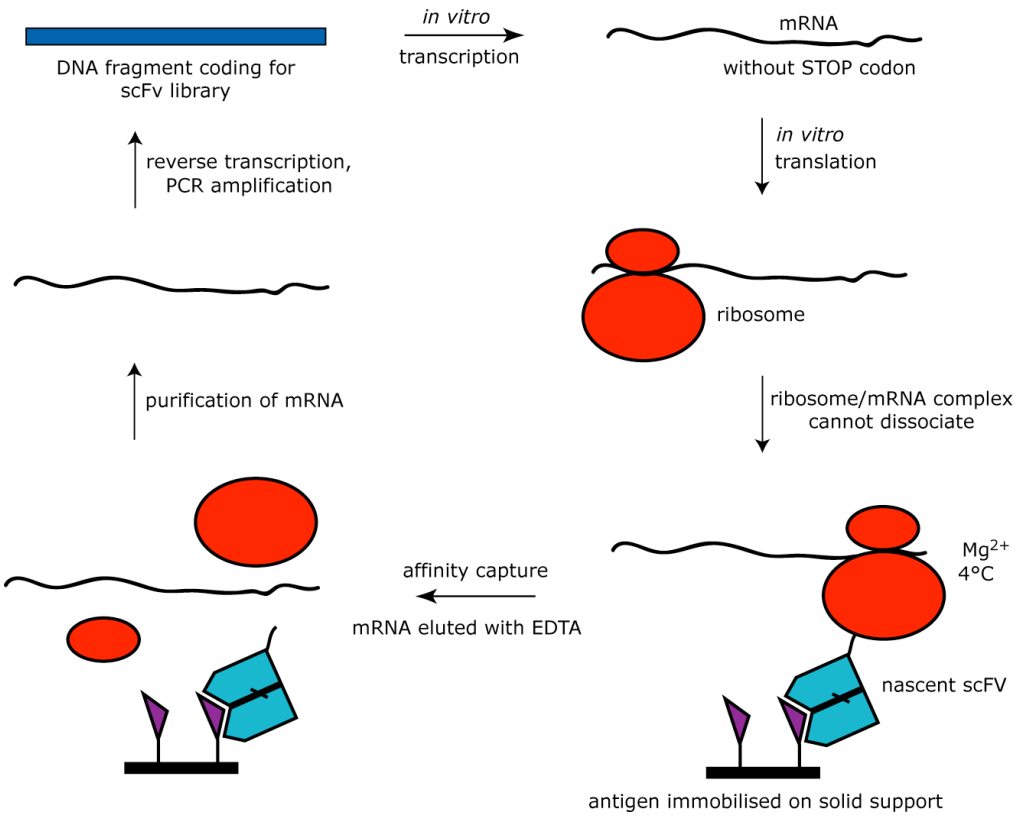

Figure 2.14. Schematic representation of the selection cycle of ribosome display. Linear DNA fragments coding for a protein library (here scFv variants) are transcribed *in vitro* and then purified before subsequent translation *in vitro*. After having reached the end of the mRNA during translation, the ribosome is unable to dissociate from the mRNA because the stop codon is missing. The resulting ternary complex comprising the ribosome, mRNA and the nascent polypeptide can be stabilised by high concentrations of magnesium ions and at low temperature, therefore creating a stable linkage between the mRNA (genotype) and the encoded polypeptide (phenotype). Ribosomes displaying a binding protein can be isolated by affinity selection on immobilised antigen, and the genetic information is amplified by reverse transcription and PCR after elution of selected mRNA molecules by addition of EDTA.

The DNA construct used for ribosome display does not contain a stop codon after the open reading frame, and thus, a quite stable mRNA–ribosome–polypeptide ternary complex can be formed during *in vitro* translation, provided the

complexes are stabilized by high Mg^{2+} concentration and low temperature after translation. The 5'-untranslated region of the ribosome display construct contains the T7 promoter for efficient transcription by T7 RNA polymerase. The 5'-untranslated region of the RNA is derived from gene 10 of phage T7 and contains the Shine–Dalgarno sequence for efficient initiation of *in vitro* translation. This 5'-untranslated region is capable of forming a stable stem-loop structure on the mRNA level. The 5'-stem-loops were shown to protect mRNA against degradation by RnaseE (Schaffitzel *et al.*, 1999). A second stem-loop structure, present at the 3'-end of the mRNA, is of particular importance to protect mRNA from degradation by 3'-5'-exonucleases in the *E. coli* S30 extract.
The protein coding sequence consists of a protein library and a spacer tether, which is necessary for sufficient spatial separation of the nascent polypeptide chain from the ribosome and thus enables the polypeptide to fold correctly while being linked to the mRNA–ribosome complex. Also, sufficient space and flexibility are provided by this tether for the protein to recognise and bind to the given target molecule

In vitro transcription and translation can be carried out either in coupled or in separate reactions. Proteins containing disulfide bridges, such as immunoglobulins, in general, fold correctly only under oxidizing conditions, such that the crucial intradomain disulfide bridges can be properly formed. It should be noted that T7 RNA polymerase requires β-mercaptoethanol to maintain its stability. Therefore, transcription and translation should be carried out separately for scFv fragments of antibodies (Hanes and Pluckthun, 1997) and other proteins containing functionally important disulfide bridges. In contrast to coupled transcription/translation reactions, where a constant level of mRNA is maintained through continuous transcription, particular attention needs to be paid towards RNases if the two steps have to take place separately. To date, five of the 20 *E. coli* ribonucleases have been shown to contribute to mRNA degradation (Schaffitzel *et al*., 1999), many if not all of which are very likely to be present in the *E.coli* S30 cell extract. Thus, mRNA stability is regarded as the limiting factor for *in vitro* translation, and mRNA half life can vary from as short as 30 sec to up to 20 min in *E. coli*, depending on mRNA secondary structure and RNase activity.

Translation arrest in ribosome display occurs by fourfold dilution of the translation mixture in an ice-cold buffer containing 50 mM Mg^{2+}. This magnesium concentration is used throughout the entire selection process to stabilise the mRNA– ribosome–protein complexes. Selections can be performed both with

ligands immobilised on a surface or in solution, if a tagged ligand is used, which can then be captured, e.g., with magnetic beads. Immobilization of protein antigens on plastic surfaces, however, may lead to partial unfolding of the protein due to hydrophobic interactions with the plastic. The high stability of the ribosomal complexes at 4°C allows for intensive washing of the bound ribosomal complexes with ice-cold Mg^{2+}-containing buffer. The selected mRNA molecules are then eluted with EDTA-containing buffer. EDTA dissociates the ribosomal complexes by chelating Mg^{2+}. The eluted mRNA is then purified with commercial mRNA purification kits, including DNaseI treatment to remove any DNA template bound nonspecifically, which may still be present after *in vitro* transcription, translation, selection and elution. The purified mRNA is utilised for RT-PCR. The Shine–Dalgarno and T7 promoter sequences at the 5'-end are reintroduced by subsequent PCR steps with appropriate primers. Thus, only intact mRNA is reverse-transcribed and PCR-amplified. The PCR-amplified DNA can now be directly transcribed *in vitro*. The mRNA can then be used for the next selection cycle (Schaffitzel *et al.*, 1999).

As a first validation of the ribosome display technology for whole proteins, a model system of two antibody scFv fragments was used. A 10^9-fold enrichment was achieved by five cycles of ribosome display, with an average enrichment of about 10^2 per cycle (Hanes and Pluckthun, 1997). All selected scFvs had acquired genetic mutations during the five cycles of ribosome display; the range was 0–4 amino acid changes from the original sequence per clone because Taq DNA polymerase without proof-reading activity was used for PCR.
Next, immune libraries were used as the starting material, and it was demonstrated that scFv antibody fragments could be selected and evolved using the *E. coli* ribosome display system. Three rounds were necessary to isolate a family of scFv fragments binding to a peptide variant of the GCN4 leucine zipper (Hanes *et al.*, 1998). Most of the isolated scFvs had again acquired mutations, with 0–5 amino acid changes from their consensus sequence, which is the sequence present in the library before the diversification occurred during selection and the most likely progenitor scFv. The best scFv obtained by ribosome display had a dissociation constant of 40 pM, as measured in solution.

Owing to poor expression levels of scFv fragments in *E.coli* lysates, the efficiency of ribosome display with scFv fragments was investigated. Interestingly, in a direct comparison (Hanes *et al.*, 1999), the *E. coli* system was found to be more efficient for the display of the model scFv constructs tested than the rabbit

reticulocyte system. This argues against the occasionally proposed idea that eukaryotic proteins would be displayed more efficiently in an eukaryotic environment, and it seems that other factors determine the efficiency of selection.

Next, a very large synthetic antibody scFv library, HUCAL-1, of $2x10^9$ independent members (Knappik *et al.*, 2000), was used directly as the starting material (Hanes *et al.*, 2000). This naïve library was applied for six rounds of ribosome display selection using insulin as the antigen. In three independent experiments, different scFv families with different framework combinations were isolated. Since the library was completely synthetic (Knappik *et al.*, 2000), the starting scFv sequences were known and any mutations could be directly identified as being generated during the ribosome display procedure by non-proofreading polymerases in the PCR steps. In summary, this procedure mimics to a certain degree the process of somatic hypermutation of antibodies during secondary immunization.

In a recent publication, ribosome display was successfully used for the isolation of binders based on the ankyrin repeat scaffold (Binz *et al.*, 2004; see chapter 2.2.5.), which bound to their targets with nanomolar affinities.

Ribosome display has been shown to work especially well for affinity maturation of scFvs and the selection of binders based on ankyrin repeats. Additionally, very large libraries can be created within short time by using PCR.
However, only few reports are published, where antibodies were selected from naïve libraries (Hanes *et al.*, 2000), most probably due to folding problems of scFv fragments in the reducing environment of *E.coli* extract. Ribosome display implicitly depends on a large fraction of the ribosomes translating a full protein sequence, thus allowing the protein to fold properly. However, such full-length translation rarely happens *in vitro* on all ribosomes. Therefore, a certain fraction of molecules displayed are not full-length. This in turn may limit the enrichment factor, as the incomplete synthesis products would be more likely not to fold, and, consequently, to bind nonspecifically. While additional pre-screening steps can be employed to maximize the fraction of full-length molecules, a direct improvement of translational efficiency might be more advantageous because it would increase the total number of full-length molecules (Lipovsek and Pluckthun, 2004).

Throughout the selection process, high Mg^{2+} concentrations and low temperature are needed in order to maintain the integrity of the ternary ribosome-mRNA-polypeptide complexes, and therefore, selection conditions can only be varied to a certain extent, which might hamper the isolation of proteins that are intended to be stable under harsh conditions, e.g. high temperature or extremes of pH. Since mRNA is used as carrier of genetic information, care should be taken to avoid contamination with RNAse molecules.

2.5.5. mRNA display and cDNA display

The mRNA display peptide- and protein-selection system [it has also been referred to as mRNA–protein fusions (Roberts and Szostak, 1997), *in vitro* virus and *in vitro* virus virion (Nemoto *et al.*, 1997) and PROfusion technology (Weng *et al.*, 2002)] provides an alternative method to ribosome display that can be applied to both ligand discovery and interaction analysis problems. In this approach, encoded peptide and protein libraries are covalently fused to their own encoding mRNA. Fusion synthesis is possible because the mRNA molecule can act as both template for peptide synthesis and peptide acceptor if it contains a 3'-puromycin molecule. Puromycin serves as a chemically stable, small-molecule mimic of aminoacyl tRNA (Figure 2.15.).

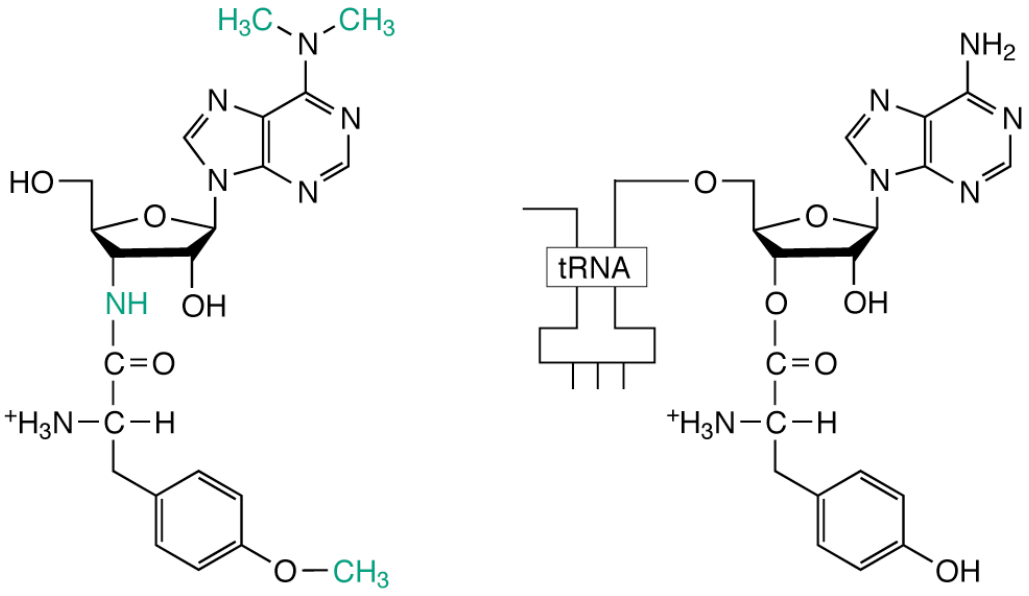

Figure 2.15. Puromycin (*left*) is a small molecule analog of tyrosyl tRNA (*right*). Differences between the two molecules are highlighted in green.

In a ribosome display selection cycle, a synthetic oligonucleotide containing a 3'-puromycin is ligated (Roberts and Szostak, 1997) or hybridised (Tabuchi *et al.*, 2001) to the 3'-end of an mRNA and the product is translated in rabbit reticulocyte lysate. After initiation of translation by the ribosome, the polypeptide encoded by the mRNA molecule is synthesised until the ribosome stalls at the end of the mRNA because the stop codon normally found at the end of an open reading frame is missing. The puromycin at the 3'-end of the mRNA enters then the ribosome at the A-site and a peptide bond between the nascent polypeptide and puromycin is formed. Therefore, the protein is covalently linked to its encoding mRNA, and this mRNA-protein fusion is amenable to subsequent selection by affinity purification. For most applications, a DNA molecule complementary to the polypeptide encoding mRNA is synthesized to stabilise the genetic information and to avoid the binding of single-stranded mRNA as an aptamer to the target molecule used for affinity purification. After affinity capture and washing, the genetic information of the isolated protein variants can be amplified either by reverse transcription followed by PCR (if cDNA has not been synthesised previously) or, after hydrolysis of RNA at high pH, by standard PCR protocols (Figure 2.16.).

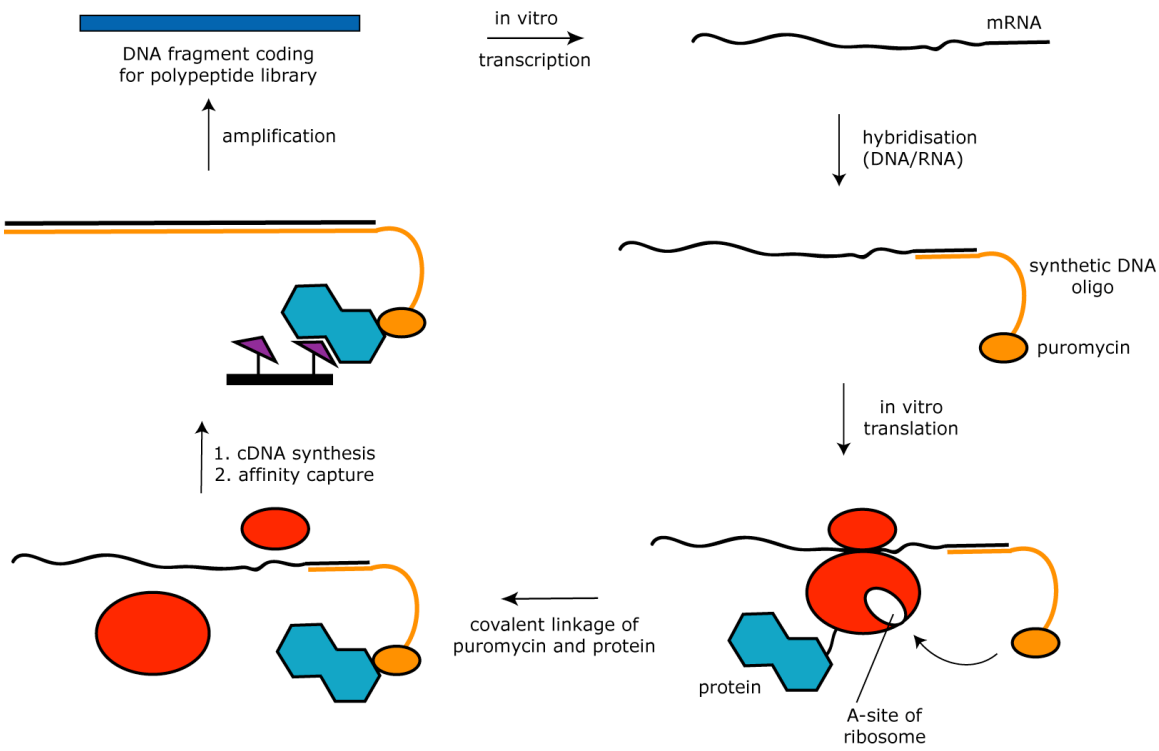

Figure 2.16. DNA fragments coding for a polypeptide library are transcribed *in vitro*. A synthetic DNA oligonucleotide containing puromycin at its 3'-end hybridised to the 3'-end

of the mRNA molecule, allows the formation of a covalent bond between the nascent protein and the puromycin. After cDNA synthesis and affinity capture, the genetic information of the selected proteins is amplified by PCR.

In the majority of mRNA-display experiments, polypeptides with relatively short chain lengths (10–110 amino acids) have been used. Larger proteins have also been studied but these typically form fusion products with somewhat reduced efficiency (Takahashi *et al.*, 2003). Early applications of mRNA display used randomized linear peptides as libraries, and well-characterized proteins known to bind peptides as targets. A peptide library with 27 randomized positions yielded a family of sequences that bound the anti-c-Myc antibody 9E10 (Baggio *et al.*, 2002). The selected clones contained sequences that were either homologous or identical to a 10-residue stretch of the 32-residue c-myc antigen. Another study started with a library of linear polypeptides with 88 randomized positions (Cho *et al.*, 2000), and identified 20 different sequences that bound streptavidin (Wilson *et al.*, 2001). A 38-residue fragment of one of the sequences selected by mRNA display bound streptavidin with a K_D of 2.5 nM. Low nanomolar affinity is more common in antibodies than in linear peptides selected *in vitro*, probably because significant entropy is lost upon binding a flexible peptide, but less entropy is lost in forming a protein–protein complex. The nanomolar affinity of linear peptides selected by mRNA display reflects probably the high complexity of the starting sequence diversity as well as the structural properties of the target molecules, which contain a groove in their surface, which is well suited for the accommodation of peptides (Lipovsek and Pluckthun, 2004).

Affinity reagents derived from constrained peptides may have an advantage over linear peptides in some applications, both because they are more similar in structure to globular proteins, and because of their smaller loss of entropy upon target binding. A disulfide-constrained library based on EETI-II, a knottin trypsin inhibitor, was constructed by randomising the six residues of the trypsin-binding site, and mRNA display was used to select new trypsin-binding peptides. The selected peptides were highly homologous or identical to wild-type EETI-II; their dissociation constants from trypsin ranged from 16 (for the wild type) to 82 μM (Baggio *et al.*, 2002).

In a selection of ATP-binding polypeptides from a library of 80 randomized residues a novel family of polypeptides was isolated, each with two invariant CXXC motifs (Keefe and Szostak, 2001). The highest-affinity clone had a K_D of 100 nM.

The first application of mRNA display to the selection of protein binders used immunoglobulin-like domains, like the tenth human fibronectin type III domain (^{10}Fn3) (see also chapter 2.2.4.). A library of ^{10}Fn3 variants was produced by randomizing the 21 residues in the loops analogous to antibody CDRs (Xu *et al.*, 2002). Using mRNA display, a selection against tumor necrosis factor alpha (TNF-α) yielded numerous Fn3-like proteins that bound the target with 1–24 nM affinity, and affinity maturation using error-prone PCR improved the affinity to 20 pM.

mRNA is, as well as ribosome display, a selection technology taking place fully *in vitro*. Therefore, genetic diversity can rapidly be generated and large libraries can be used (10^{12}-10^{13}). However, there are some points, where care should be taken (see below) and a still unanswered question in mRNA display is how exactly the purification of protein–puromycin–DNA–mRNA adduct from the ribosome can be performed (Liu *et al.*, 2000; Lipovsek and Pluckthun, 2004). After translation, the protein folds outside of the ribosomal tunnel to a globular domain. At the other end of the tunnel, the puromycin–mRNA/DNA reagent reacts with the polypeptide. Thus a folded domain sits at one end of the tunnel, while the long mRNA/DNA is connected to the peptide at the other end. Whereas the purification is performed under conditions expected to dissociate the ribosome, no direct evidence is yet available that an ''opening'' of the tunnel takes place. Alternative explanations are that (*i*) the mRNA/DNA passes through the protein exit tunnel, or (*ii*) the protein denatures and goes ''backwards'' through the tunnel. If such denaturation of the displayed protein is required, that might limit the application of mRNA display to proteins with robust refolding properties (Lipovsek and Pluckthun, 2004).

A further delicate point in mRNA display is the ligation process. For efficient ligation of the puromycin containing oligonucleotide to the mRNA, two factors must be taken in to account. First, the mRNA template to be ligated should be devoid of secondary structure at its 3'-end as this can completely abolish ligation. Second, it is important to purify and desalt the template, the splint (connecting oligonucleotide) and the puromycin oligonucleotide, since high concentrations of salt can interfere with ligation (Liu *et al.*, 2000).
The ligated mRNA-puromycin can then be used for *in vitro* translation and synthesis of mRNA-protein fusions. However, by the end of a standard translation reaction, only a small amount of the *in vitro*-synthesised protein has been converted to its mRNA-protein fusion. The yield of mRNA-peptide fusions can be

increased by posttranslational addition of Mg^{2+} and K^{+}, use of a flexible linker of the correct length (puromycin oligonucleotide), and long incubations (12–48 h) at room temperature (with metal ions) or low temperatures (-20°C) (without metal ions) (Liu *et al.*, 2000). Once synthesised, mRNA-protein fusions are purified before the next step in the selection cycle, which is the affinity purification. Alternatively, the selection step could also be performed directly using the fusion/lysate mixture. However, there are several disadvantages to this. For example, as already mentioned above, RNA aptamers that bind the selection target could be selected rather than peptides. Further, RNA and protein could synergistically collaborate to provide the desired function. Therefore, mRNA-protein fusions are purified in order to generate a cDNA/mRNA hybrid fusion product using reverse transcriptase before the selective step (Liu *et al.*, 2000).

In summary, even though successful selections using mRNA display have been reported, the procedure remains quite laborious and complicated.

2.5.6 *CIS* display

A common limitation of selection methodologies using mRNA as library-encoding nucleic acid is the physical instability of mRNA molecules and their sensitivity to RNAses (Odegrip *et al.*, 2004). An alternate approach termed *CIS* display was developed, which uses double-stranded DNA as carrier of genetic information (Odegrip *et al.*, 2004).

This system exploits the *cis*-activity of a DNA replication initation protein RepA of the R1 plasmid (Masai *et al.*, 1983; Nikoletti *et al.*, 1988). Three replication functions, copB, copA and RepA are encoded on the basic replicon of the R1 plasmid. RepA protein, identified as a 33 kDa protein, was shown to be required for initiation of R1 plasmid replication *in vitro* (Masai and Arai, 1988). A unique feature of R1 plasmid replication is the apparent *cis*-activity of RepA protein. Newly synthesised RepA protein is preferentially utilised by nearby origins of replication sequences (*oriR*) on the same template from which it was synthesised, and other templates are not activated (Masai and Arai, 1988). The DNA sequences coding for RepA protein and *oriR* are separated by approximately 170 base pairs. This short DNA sequence in between RepA and *oriR* was designated as *CIS* sequence. After removal of the *CIS* sequence, the *cis*-activity of RepA was lost (Masai and Arai, 1988). *CIS* contains a rho-dependent transcription

terminator, which terminates RepA transcription. Detailed analysis showed that transcription terminator activity is necessary but not sufficient for efficient replication (Masai and Arai, 1988). The *CIS* sequence may contain additional information, which helps to maintain a stable mRNA/DNA hybrid, so that the newly synthesised RepA protein on the ribosome is kept close to the template DNA. RepA protein seems to be no longer available for reutilisation by a second template once it has bound to *oriR in cis* and has initiated replication (Masai and Arai, 1988). The consensus model for *cis*-activity is that the *CIS* element, which contains a rho-dependent transcriptional terminator, causes the host RNA polymerase to stall. This delay allows nascent RepA polypeptide emerging from translating ribosomes to bind transiently to *CIS*, which in turn directs the protein to bind to the adjacent *ori* site (Praszkier and Pittard, 1999; Praszkier *et al.*, 2000).

By genetically fusing peptide libraries to the N terminus of the RepA protein, a direct linkage of peptides to the DNA molecules that encode them was achieved, and thus, the link between genotype (DNA) and phenotype (RepA fusion protein) that is needed in evolutionary methodologies is established (Odegrip *et al.*, 2004). *CIS* display selections begin with the construction of a peptide encoding DNA library followed by *in vitro* transcription and translation (in an *E.coli* lysate) to form a pool of protein–DNA complexes. The library pool is incubated with an immobilised target, non-binding peptides are washed away, and the retained DNA molecules that encode the target-binding peptides are eluted and amplified by PCR, to form a DNA library ready for the next round of selection. After three to five rounds of selection, recovered DNA is cloned into an appropriate expression vector for the identification of individual target binding peptide sequences (Figure 2.17.).

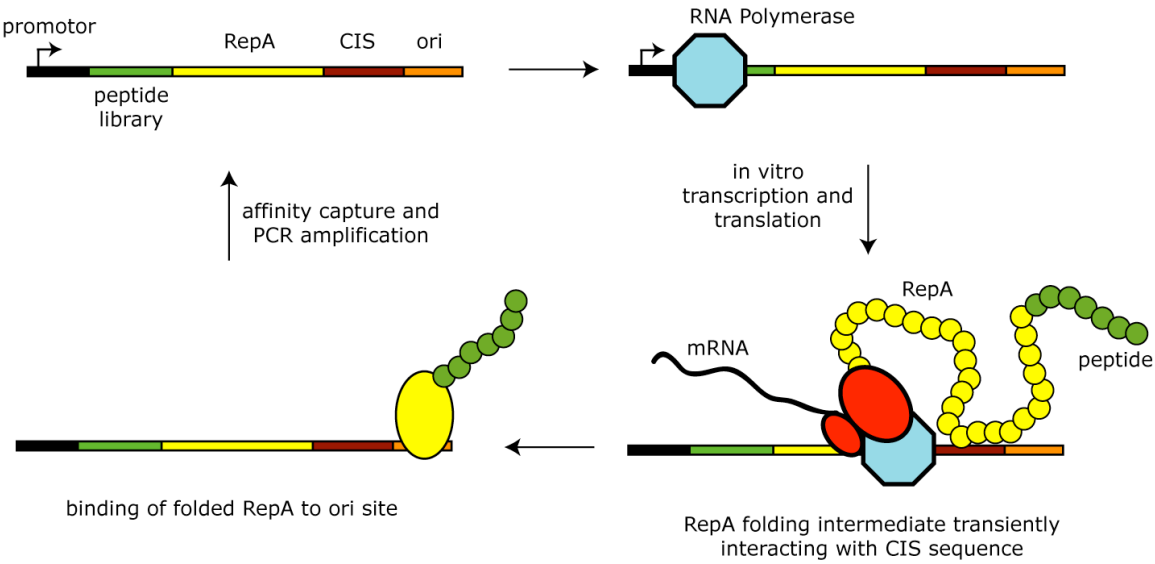

Figure 2.17. Schematic representation of a selection cycle of *CIS* display. A library of DNA templates comprising a promotor, a randomised peptide sequence as N-terminal fusion to RepA, a *CIS* sequence and an origin of replication (*ori*) is used for *in vitro* transcription and translation. During transcription, newly synthesised mRNA is simultaneously translated and, due to the stalling RNA polymerase at the *CIS* sequence, RepA folding intermediate transiently interacts with *CIS* sequence. Finally, completely folded RepA-peptide fusion binds to the *ori* site on the DNA template, thus forming a protein-DNA complex, which is amenable to affinity selection experiments and subsequent amplification of isolated DNA molecules by PCR.

To demonstrate *cis*-activity of RepA, two linear DNA constructs coding either for the 14-aa V5 peptide tag (Southern *et al.*, 1991) or the human 105-aa human Cκ Ig constant domain, were mixed in a equimolar ratio and translated *in vitro*. The resulting protein-DNA complexes were subjected to affinity selection against two different antibodies. If anti-Cκ antibody was used, only DNA molecules coding for the Cκ-RepA fusion were isolated. In contrast, if anti-V5 antibody was taken for affinity selection, DNA templates encoding the V5-peptide-RepA fusion were selected (Odegrip *et al.*, 2004). Furthermore, an 18-mer peptide library fused to RepA comprising 10^{12} individual peptide sequences was constructed. From this library, peptides binding to anti-Flag antibody, anti-p53 antibody and lysozyme were selected and their specificity was assayed by ELISA using a panel of different antigens.

The use of DNA to encode the displayed peptides provides advantages over RNA-based *in vitro* selection methods. Very large libraries can be rapidly constructed and screened without separating transcription and translation steps and without

purification of the protein–DNA complexes before selection. Also, control of divalent cation concentration is not required, and the complexes do not require incubation under sterile or ribonuclease-free conditions. Since the RepA-DNA interaction is not covalent (in contrast to mRNA display), the stability of RepA–DNA complexes was examined by incubating *in vitro* expressed Cκ-RepA with an anti-Cκ antibody for 48 h before elution and recovery of associated DNA. In this experiment, no reduction in DNA yield was observed compared with that obtained using standard selection conditions (Odegrip *et al.*, 2004).

2.5.7. Microbead display

As already mentioned at the beginning of chapter 2.5., Darwinian evolution relies on the linkage of genotype to phenotype. Early in the earth's history this linkage could have been achieved if RNA, or more primitive molecules, came into existence that could catalyse their own replication and function as both gene and enzyme (Joyce, 1989). Without some form of compartmentalisation, however, the potential benefits of other molecules encoded by the gene would be lost. It is possible, that self-reproducing compartments (perhaps membrane-bound vesicles) could have been the first living thing and only later captured nucleic acids with which they replicated symbiotically (Deamer, 1997). In nature, compartmentalisation has provided a very general way of linking genotype and phenotype that has allowed the evolution of a vast range of molecules with diverse activities.

In 1998, the application of *in vitro* compartmentalisation (IVC) for the selection of nucleic acids coding for an enzyme with beneficial properties was reported for the first time (Tawfik and Griffiths, 1998). The concept of *in vitro* compartmentalisation (IVC) has been used in a number of further publications (Ghadessy and Holliger, 2004) for the selection of binding (Sepp *et al.*, 2002; Doi and Yanagawa, 1999; Yonezawa *et al.*, 2003) and catalytically active proteins (Ghadessy *et al.*, 2001; Lee *et al.*, 2002; Griffiths and Tawfik, 2003; Cohen *et al.*, 2004; Doi *et al.*, 2004; Ghadessy *et al.*, 2004; Bernath *et al.*, 2005).

In order to combine the concept of *in vitro* compartmentalisation with FACS for the isolation of binding proteins, a novel methodology was published, based on creating repertoires of microbeads, each displaying a gene and the protein that it encodes. These beads were selected by the binding properties of the polypeptide

variants present on their surface (Sepp *et al.*, 2002) (Figure 2.18.). Biotinylated Flag-HA genes, encoding a 30-residue peptide with a Flag epitope tag at the N-terminus and an HA epitope at the C-terminus, or Flag-folA genes, encoding DHFR with an N-terminal Flag epitope tag were attached to 1-μm diameter streptavidin-coated beads coated with anti-Flag antibodies at one gene/bead. Microbeads displaying the translated protein were then created by translation in an water-in-oil emulsion, and the beads were subsequently labeled with peroxidase-conjugated anti-HA antibody and fluorescein tyramide signal amplification (TSA).

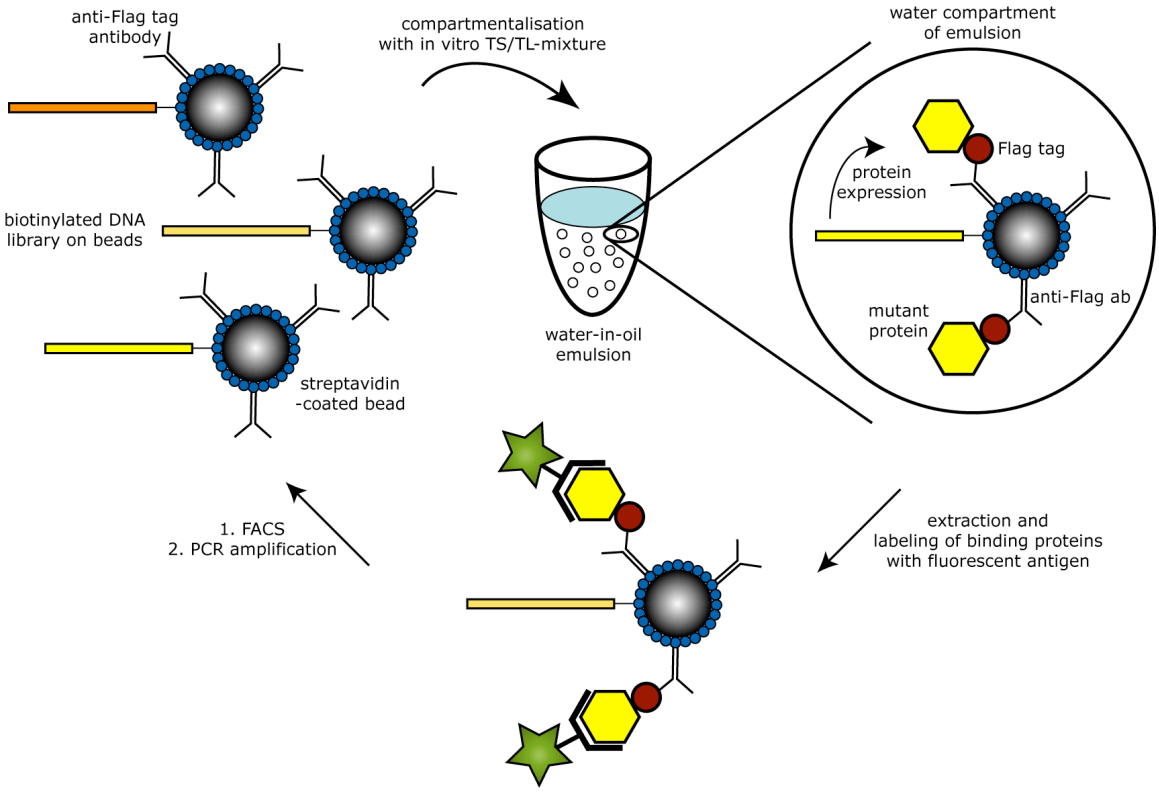

Figure 2.18. Creation of microbead display libraries by IVC and selection for binding using flow cytometry. A repertoire of genes encoding protein variants, each with an N-terminal Flag epitope tag and a C-terminal potential binding domain, are linked to streptavidin-coated beads (carrying antibodies that bind the epitope tag) at, on average, one gene per bead. The beads are compartmentalised in a water-in-oil emulsion to give, on average, one bead per compartment, and transcribed and translated *in vitro* in the compartments. Consequently, in each compartment, multiple copies of the translated protein become attached to the gene that encodes it via the bead. The emulsion is broken, and the microbeads carrying the display library isolated. The beads are incubated with labeled ligand molecules, and beads carrying binding protein variants on their surface can be isolated by flow cytometry (adapted from Sepp *et al.*, 2002).

The beads displaying Flag-HA peptides could easily be distinguished from the beads displaying Flag-folA by flow cytometry: the mean fluorescence of beads carrying the Flag-HA genes was 57 times that of beads with Flag-folA genes. Biotinylated Flag-HA genes and biotinylated Flag-folA genes were mixed in a ratio of $1:10^6$ and attached to beads coated with anti-Flag antibodies at one gene/bead, as above. Microbeads displaying the translated protein were then created by translation in an emulsion and the beads were labeled with peroxidase-conjugated anti-HA antibody and fluorescein TSA (Sepp *et al.*, 2002). Genes encoding a peptide containing an HA epitope tag were enriched to near purity within tow rounds of selection using flow cytometry, indicating ~1'000-fold enrichment per round.

Microbead display libraries can be selected by flow cytometry, which has a variety of practical advantages for the selection of ligand binding. Ligand binding equilibria and dissociation kinetics can be determined directly on the beads without need for cloning and purification of the selected polypeptides. Indeed, flow cytometry has been used to select an extremely high affinity anti-fluorescein scFv antibody (K_D = 48 fM) from libraries displayed on yeast (Boder *et al.*, 2000). Throughput of flow cytometry is relatively high (up to 100 000 s^{-1}; Sepp *et al.*, 2002) but flow cytometry does impose an upper limit of $\sim 10^9$ on the size of libraries that can be selected. The small size of the compartments means that very large gene libraries could potentially be selected. However, to do so it is probably necessary to select the microbead display libraries by affinity purification, as it is done with other display technologies.

Having a stronger interaction than the one between Flag tag and anti-Flag antibody to display the translated polypeptides on the beads would be preferable, because FACS opens the opportunity to isolate binders with extremely high affinities. Therefore, if a microbead library was screened for very low k_{off} values of the displayed proteins, the Flag tag/anti-Flag antibody interaction (K_D = 15 nM; Sepp *et al.*, 2002) could impose an upper threshold for selectable k_{off} values.

2.5.8. STABLE: a protein-DNA fusion system

A methodology was described, where a DNA is attached to the protein which it encodes through a stable linkage in a compartmentalised *in vitro* transcription-translation system (Doi *et al.* 1999; Yonezawa *et al.*, 2003). As already

mentioned above, such emulsions can be used as cell-like compartments that keep the DNAs and the proteins they encode together (Tawfik and Griffiths, 1998). When an *in vitro* synthesised polypeptide is directly attached to its encoding DNA in each compartment, the resulting protein-DNA fusion molecules can be recovered from the emulsion and be subjected to affinity selection. Streptavidin (STA)-biotin complex was chosen as a connector for the protein-DNA fusion. The proposed technology was termed **STA**-biotin **l**inkage in **e**mulsions ("STABLE"), because (*i*) the STA-biotin conjugate is very stable, and (*ii*) DNA is more stable than RNA as a genetic material (Figure 2.19.).

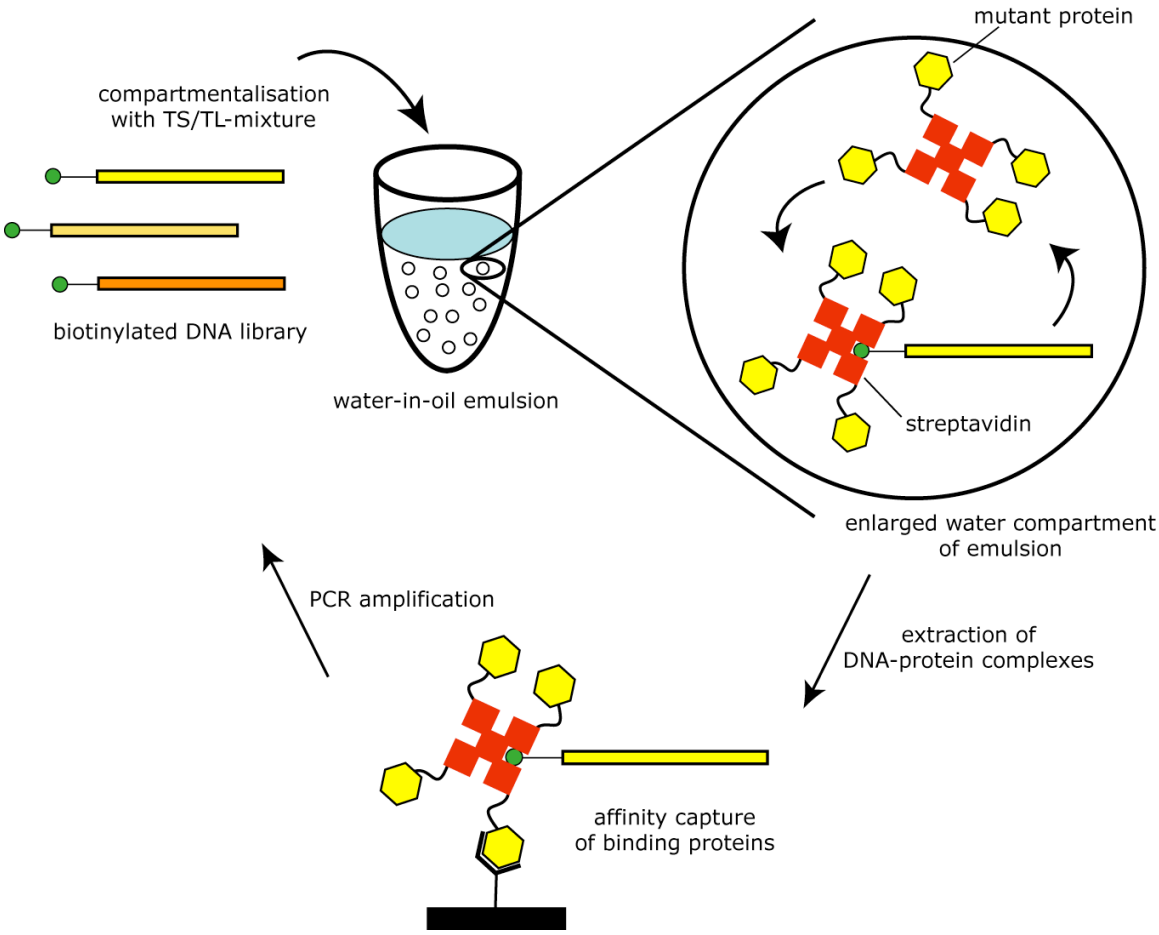

Figure 2.19. Schematic representation of the DNA display selection procedure (STABLE). A DNA library encoding streptavidin-fused polypeptides is labeled with biotin and compartmentalised in water-in-oil emulsions containing an *in vitro* transcription/translation (TS/TL) mixture. In each compartment, streptavidin-fused peptides are synthesised and attached to the template DNA via biotin labels. DNA-peptide conjugates are recovered from the emulsion and subjected to affinity selection on an immobilised target molecule. After washing and elution, the DNA portion of the isolated molecules is amplified by PCR. DNA is then subjected to the next round of selection or can be identified by sequencing.

In their original publication (Doi *et al.*, 1999), the authors used an *E.coli* S30 lysate transcription/translation mixture. Using this expression system, only about 1% of input DNA associated with streptavidin fusions due to the extremely inefficient expression of the streptavidin fusions in *E.coli in vitro* transcription and translation system. In a further study, an optimised protocol was published (Yonezawa *et al.*, 2003), where expression was improved by using a wheat germ *in vitro* transcription and translation system and by synthesis of a streptavidin gene with reduced GC content (from 68 to 54%). DNA-protein complexes were formed with an efficiency of more than 95%.

To test whether the improved DNA display system was effective in screening peptide ligands, an anti-Flag epitope antibody was used as bait for selection. The DNA-displayed, randomised decapeptide library was captured on agarose beads coated with anti-Flag antibody. Unbound protein-DNA complexes were washed away, and the remaining DNA molecules displaying binding polypeptides fused to streptavidin were eluted with an excess of Flag peptide. The simplicity of the selection procedure allowed two rounds of affinity selection per day. After each round, DNA molecules were amplified, and the enriched population was translated *in vitro* and analysed by dot blot analysis using the anti-Flag antibody. The fraction of peptides capable of binding to the anti-Flag antibody increased in each round and reached saturation after the fifth round of selection. Sequence analysis of randomly chosen clones from round five revealed that 69 out of 71 clones could had sequences that were similar to the wild-type Flag-tag epitope (Yonezawa *et al.*, 2003).

The use of a DNA genotype offers some advantages over an RNA genotype because the chemical stability of DNA permits screening for peptides under conditions where RNA could be degraded. These include peptide ligands for receptors that are localised on membranes or whole cells. The stable DNA display would be beneficial in off-rate selections of high affinity binders over a long period of time, and in the directed evolution of novel proteins that work under harsh conditions. Despite the successful isolation of peptides binding to anti-Flag antibody it remains to be seen whether the methodology can also be applied for the selection of folded proteins.

2.5.9. Endonuclease P2A mediated display

In a very recent publication (Reiersen *et al.*, 2005), a direct and covalent linkage of template DNA to its encoded protein has been created by exploiting the *cis*-nicking activity of the endonuclease P2A from the bacteriophage P2. This phage replicates by attaching its early gene product P2A to its own DNA. The first genetic evidence for the *cis*-activity of P2A *in vivo* was obtained when it was observed, that wild-type P2A expressed from phage genome was not able to complement inactivating mutations in P2A (Reiersen *et al.*, 2005).
Endonuclease P2A initiates rolling circle replication of P2 phage *in vivo*. Its catalytic tyrosine residue (Y454) makes a single-stranded specific nick at the viral origin of replication and forms *in vivo* transiently a covalent bond with the 5'-phosphate group of the coding strand (Figure 2.20.).

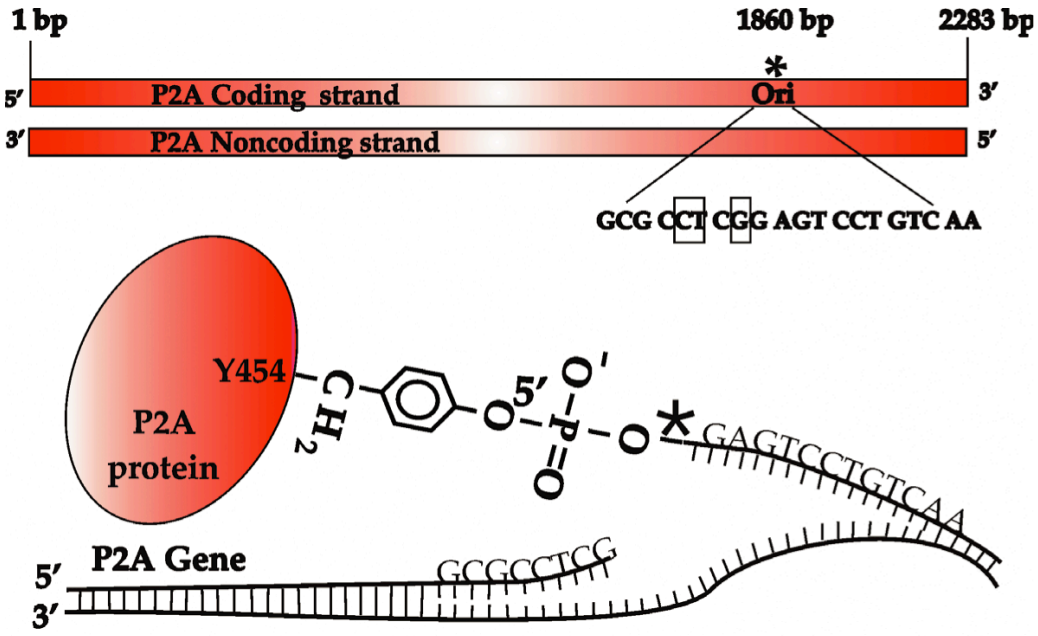

Figure 2.20. The endonuclease P2A (761 residues, 86.3 kDa) makes a single-stranded site-specific nick at *ori* of replication, located inside its own gene at position 1860, and becomes covalently attached (via Y454) to the 5'-phosphate of its own DNA (adapted from Reiersen *et al.,* 2005).

Single chain antibodies are genetically fused to the P2A protein creating a protein-DNA fusion amenable to affinity purification (Figure 2.21.) using standard selection strategies. Specifically binding protein-DNA fusions can be captured, eluted and amplified by PCR.

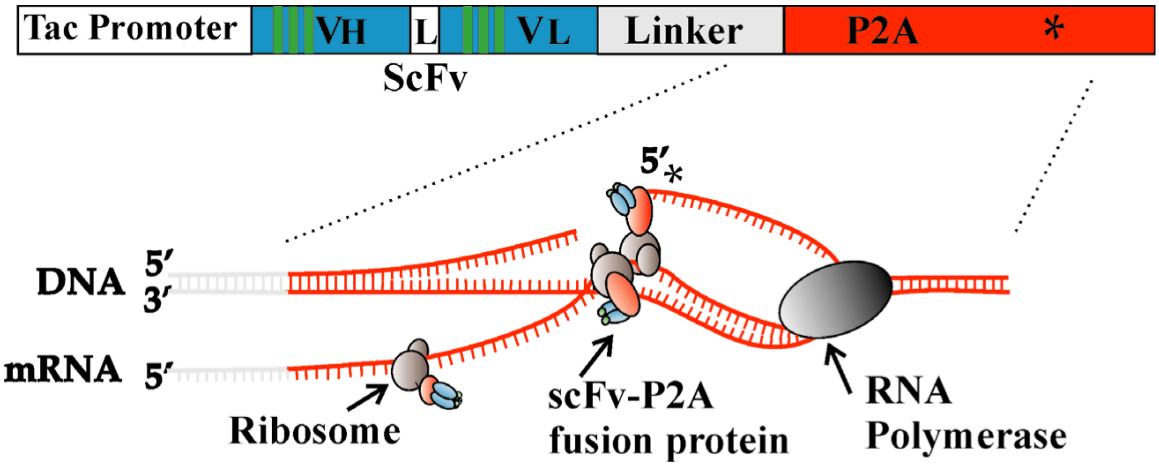

Figure 2.21. Endonuclease P2A is used to connect antibodies or antibody fragments to their encoding information. In prokaryotes, the transcription and translation are coupled, and the start of translation normally takes place before the transcription is finished. A model for the covalent cross-linking of the P2A-antibody fusion is shown schematically. P2A-scFv protein is synthesised by the ribosomal machinery during transcription. Due to the endonucleolytic activity of P2A, a covalent bond can be formed between DNA and encoded protein (adapted from Reiersen *et al.,* 2005).

In order to prove the feasibility of the proposed display technology, the authors performed capture experiments using two well characterised scFv fragments as N-terminal fusion partners for P2A. In the best experiment shown, DNA encoding a scFv-P2A fusion was recovered about 10-fold more efficiently if a scFv-binding ligand was immobilised on the solid support than compared to the situation, where "naked" solid support without ligand was applied.
Furthermore, spiked selections with anti-phOx-scFv-P2A DNA mixed with a 25'000-fold excess of scFv-P2A fusions from a library, which was isolated from human lymphocytes, were performed. After three rounds of selection with the hapten phOx (4-ethoxymethylene-2-phenyl-2-oxazoline-5-one) as bait, 82 out of 3040 clones were positive as analysed by filter screening. The enrichment from each round of selection decreased from round to round (round 1: 50-fold; round 2: 8-fold; round 3: 2-fold), indicating that P2A has also a significant *trans*-activity.

In an additional experiment, the human antibody library isolated from lymphocytes of a donor who had been previously immunised against tetanus toxoid, was used for selection experiments. Tetanus toxoid was taken as selection target, and the serum titre of the donor against tetanus toxoid was found to be

about 1:25'000. After one round of selection, some clones were found to bind weakly to tetanus toxoid in ELISA.

Expression levels of P2A-scFv in *E.coli* lysate were low. It was estimated, that only about 3% of the input DNA molecules were able to produce full-length P2A-scFv protein (Reiersen *et al.*, 2005). Therefore, the practical use of the proposed technology to isolate scFv antibody fragments seems to be limited, unless further improvements become available.

2.6. Covalent DNA display

After reviewing a large variety of different selection methodologies, it can be concluded that an ideal *in vitro* system for the directed evolution of proteins should comprise the following features:

- rapid generation of large genetic diversity (library).
- facile and irreversible linkage between phenotype (protein) and genotype (genetic material).
- high expression levels of protein mutants in the format used for the selection process.
- a carrier of genetic information, which is stable under a wide range of conditions (e.g. high and low salt concentrations, extremes of temperature and pH, denaturing agents).
- control of the number of proteins to be linked to the genetic material (valence of display).

We recently proposed a technology (Bertschinger and Heinis, 2003; Bertschinger and Neri, 2004) in which a library of linear DNA molecules is co-packaged with an *in vitro* transcription/translation mix in the compartments of a water-in-oil emulsion (Tawfik and Griffiths, 1998) (Figures 2.22. and 2.23.).

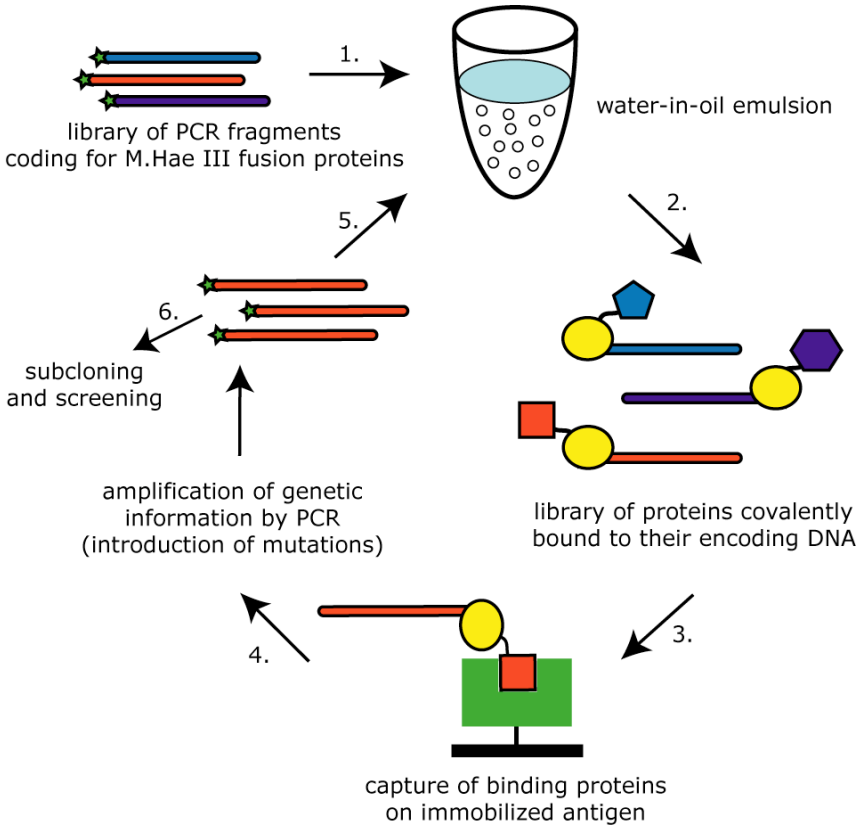

Figure 2.22. Concept of the selection method. A library of linear DNA molecules is co-packaged with an *in vitro* transcription/translation mix into a water-in-oil emulsion (1). Ideally, one compartment contains one DNA molecule. The DNA molecules each code for a Hae III DNA-methyltransferase fusion protein and contain a mechanism-based inhibitor for the covalent cross-linking of the DNA-methyltransferase fusion proteins (see also Figures 2.23. and 2.24). After *in vitro* expression and formation of the DNA–protein complexes, the water phase is extracted from the emulsion (2) and DNA molecules displaying a protein with desired binding properties are selected from the pool of DNA–protein fusions by affinity panning (3). The genetic information of selected DNA–protein fusions is amplified by PCR (4) and can either be used for a further round of selection (5) or for cloning and characterisation of the selected mutants (6).

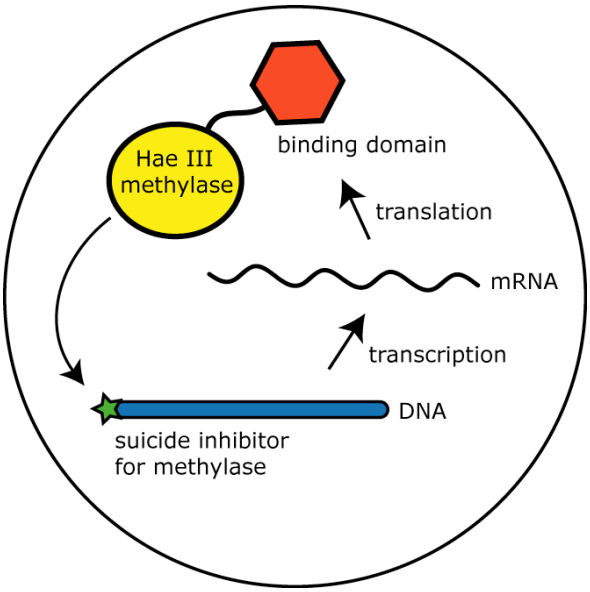

Figure 2.23. Enlarged view of a water compartment of the water-in-oil emulsion. The DNA molecule is transcribed into mRNA, which is translated into a fusion protein consisting of two domains: the N-terminal Hae III DNA-methyltransferase (yellow) and a C-terminal potential binding domain (red). Owing to the catalytic activity of the Hae III DNA-methyltransferase, the fusion proteins form a covalent bond via the modified methylation target sequence 5'-GGFC-3' (F = FdC = 5-fluorodeoxycytidine) with their encoding DNA molecule, which is present in the same compartment of the water-in-oil emulsion.

Experimental conditions are chosen so that, in most cases, one compartment contains one DNA molecule. A covalent linkage between DNA and the encoded polypeptide is achieved by using fusion proteins which contain the Hae III DNA-methyltransferase domain of *Haemophilus aegypticus*, an enzyme which is able to form a covalent bond with DNA fragments containing the sequence 5'-GGFC-3' (F = 5-fluoro-2'-deoxycytidine) (Chen *et al.*, 1991) (Figure 2.24). M.Hae III fusion proteins expressed in each compartment of the water-in-oil emulsion are extracted from the emulsion, and DNA molecules displaying a protein with the desired binding specificity are selected from the pool of DNA–protein fusions by affinity selection. The genetic information of the selected DNA–protein fusions is amplified by polymerase chain reaction (PCR) and can be used either for a further round of selection or for cloning and characterisation of the selected mutants (Figure 2.22.).

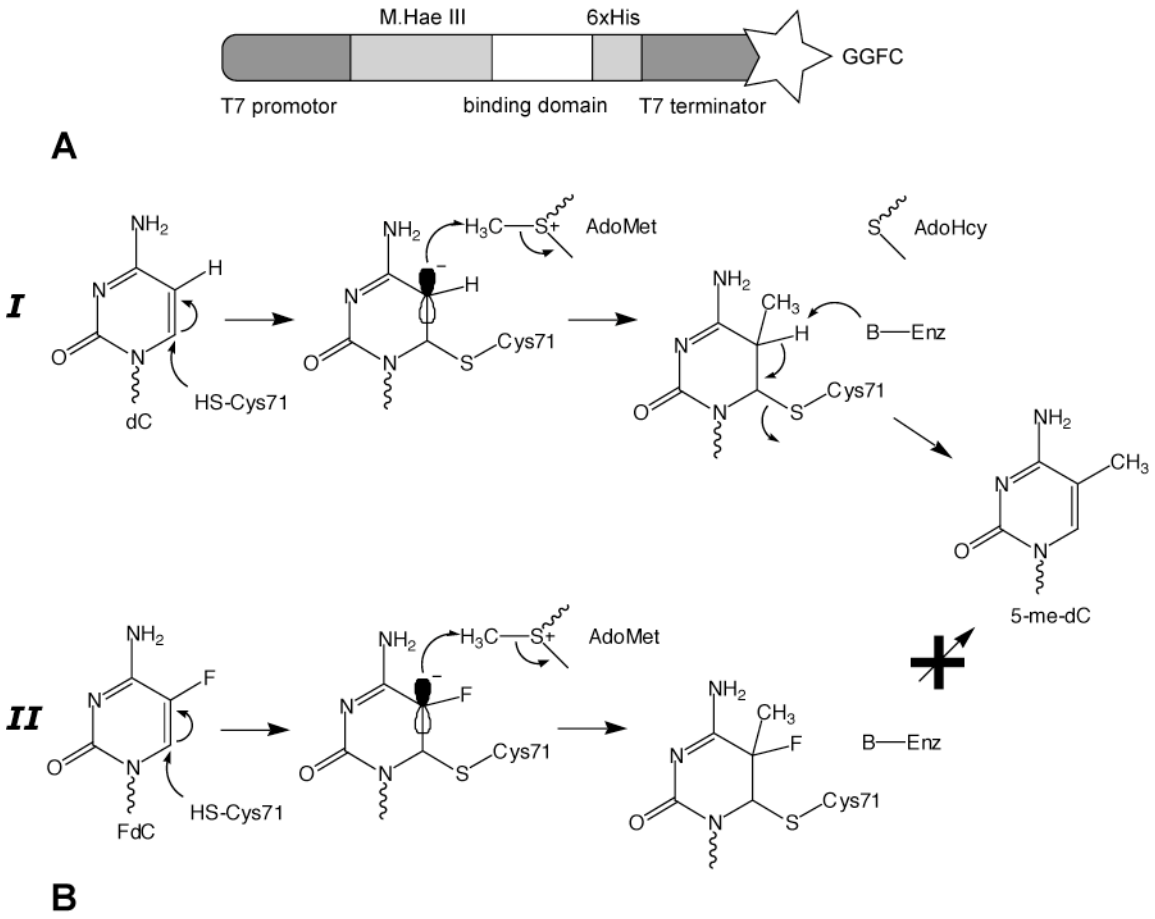

Figure 2.24. (*A*) Schematic representation of DNA template used for the *in vitro* expression of M.Hae III fusion proteins. The sequences coding for Hae III DNA-methyltransferase and the potential binding domain are flanked by T7 regulatory sequences (T7 promoter and T7 terminator) for efficient expression. The modified base 5-FdC is introduced into the DNA by PCR at the downstream end of the DNA T7 terminator sequence. (*B*) Chemical reaction mechanisms that lead either to the methylation of deoxycytidine (dC) (I) or to the covalent linkage of M.Hae III with the modified methylation sequence 5'-GGFC-3' (F = FdC = 5-fluorodeoxycytidine) (II). I: Methylation is initiated by nucleophile attack of an active site thiol group (Cys71) at C6 of cytosine. S-Adenosylmethionine (AdoMet) acts as the methyl donor. The methylated dihydrocytosine intermediate undergoes beta-elimination to generate a 5-methylcytosine and an active enzyme. II: If the nucleotide 5-fluorodeoxycytidine is inserted in the methylation sequence (5'-GGFC-3'), the dihydrocytosine beta-elimination is blocked. The enzyme remains covalently bound to the DNA substrate.

2.6.1. Theoretical considerations on compartmentalisation

In principle, emulsions can be produced from any suitable combination of immiscible liquids. In our case, these liquids are water and oil, which can be mixed to form an emulsion by applying mechanical energy to force the phases together. In order to avoid the spontaneous separation of water and oil back into discrete phases, the emulsion must be stabilised by the addition of surfactants. Therefore, by mixing a certain amount of water with an excess of oil phase (e.g. ~20x) in presence of stabilising surfactants, a stable emulsion can be produced consisting of tiny water compartments surrounded by the oil phase, called water-in-oil emulsion (Tawfik and Griffiths, 1998). For the successful application of water-in-oil emulsions in a selection methodology, the aqueous microcapsules must be stable, have a suitable size, and biochemical reactions such as transcription and translation must be feasible within these water compartments (Nametkin *et al.*, 1992).

For the creation of large functional diversity on a protein level, the genetic information present in each of the water compartments must be expressed efficiently. Both *in vitro* transcription and translation become less reliable at sub-nanomolar DNA concentrations. Because of the requirement of ideally only one DNA molecule to be present in each microcapsule, this sets a practical upper limit on the possible microcapsule size (Table 2.3.). Additionally, the number of compartments in a given volume strongly increases with decreasing diameter, and therefore, the diversity, which can be sampled in a selection experiment, increases as well.

Table 2.3. Influence of water compartment diameter on the number of microcapsules per volume and the effective concentration of one molecule (e.g. DNA) per compartment.

Diameter (μm)	No. of compartments/50 μl[a]	Concentration (M)[b]
0.1	$9.6 \cdot 10^{13}$	$3.2 \cdot 10^{-6}$
0.3	$3.5 \cdot 10^{12}$	$1.2 \cdot 10^{-7}$
0.5	$7.7 \cdot 10^{11}$	$2.5 \cdot 10^{-8}$
1.0	$9.6 \cdot 10^{10}$	$3.2 \cdot 10^{-9}$
2.0	$1.2 \cdot 10^{10}$	$4.0 \cdot 10^{-10}$

[a]In a typical selection experiment, 50 μl of water phase are mixed with 950 μl of oil phase.

[b]Effective concentration of a single molecule present in one compartment.

An ideal diameter for a water compartment would in the range of 1-2 μm, because the prokaryotic transcription/translation machinery has evolved to operate in compartments (cells) of approximately this size, in which single genes are present at approximately nanomolar concentrations.

The microcapsule size not only has to be small enough for the creation of large libraries, but it must also be sufficiently large to accommodate all of the required components of the biochemical reactions that are needed to occur within the microcapsule. For example, *in vitro*, both transcription reactions and coupled transcription-translation reactions require total nucleoside triphosphate concentration of about 2 mM. In order to transcribe a gene to a single short RNA molecule of 500 bases in length, this would require a minimum of 500 molecules of nucleoside triphosphate per microcapsule. In order to constitute a 2 mM solution, this number of molecules must be contained within a microcapsule of volume 4.17×10^{-19} liters, which if spherical would have a diameter of 93 nm (Griffiths and Tawfik, 2002). Furthermore, particularly in the case of reactions involving translation, it is to be noted that the ribosomes necessary for the translation to occur are themselves approximately 20 nm in diameter. Therefore, the lower limit for microcapsules is a diameter of around 0.1 μm (Griffiths and Tawfik, 2002).

As already mentioned, water-in-oil emulsions must be stable to be of practical use in selection experiments. If water compartments fused, an exchange of phenotypes (proteins) would occur leading to an incorrect assignment of genotype and phenotype. Compartment size and stability can be controlled by

using appropriate surfactants at optimal concentrations. Surfactant molecules have a polar head group and an apolar tail, which leads them to the water-oil interphase in the emulsion. When working with a given amount of water phase, higher concentrations of surfactants result in smaller microcapsules, because the urge of the polar head groups of the surfactants to be in the vicinity of water pushes the system to maximise the water-oil phase interphase, therefore decreasing microcapsule size (Oldfield and Freedman, 1989). However, a monodisperse emulsion can only be formed by energy input, such as ultrasound or vigorous stirring. By adjustment of the stirring speed and velocity, particle size can be readily controlled (Tawfik and Griffiths, 1998).

Oil phase containing mineral oil and the surfactants Tween 80 (4.5% v/v) and Span 80 (0.5% v/v) was used for the production of remarkably stable water-in-oil emulsions (Tawfik and Griffiths, 1998), which did not interfere with transcription or translation taking place within the water compartments. These emulsions were even stable when performing PCR inside the water droplets (Ghadessy *et al.*, 2001). Thus, water-in-oil emulsions have proved to be a useful tool for the assignment of genotpye and phenotype by compartmentalisation.

2.6.2. Methyltransferase III from *Haemophilus aegypticus*

The enzymatic addition of methyl groups to DNA is an essential element of genomic function in organisms ranging from bacteria to mammals (Razin *et al.*, 1984). In prokaryotes, DNA methylation directs the mismatch repair and restriction-modification systems, which correct errors and prevent transformation by non-self DNA, respectively. DNA-methyltransferases can transfer a methyl group from the cofactor S-adenosyl-L-methionine (SAM) to duplex DNA. The class of DNA methyltransferase enzymes that direct methylation of the 5-position of C [DNA-(cytosine-5)-methyltransferases)] are designated as DCMtases. DCMtases carry out a chemically improbable reaction: substitution of an unactivated vinyl hydrogen with a methyl group, in neutral aqueous solution. They do this by using a catalytic mechanism that involves an intermediate, in which the enzyme is covalently bound to the substrate DNA (Figure 2.24.).

Extensive sequence comparisons of DCMtases (Lauster *et al.*, 1989; Posfai *et al.*, 1989) have revealed the presence of an absolutely conserved Pro-Cys motif, of which the Cys-SH was shown to be the catalytic nucleophile (Chen *et al.*, 1991;

Verdine and Norman, 2003). The structure of the DCMtase from *Haemophilus aegypticus* (M.Hae III) was elucidated in complex with double-stranded DNA as substrate (Reinisch *et al.*, 1995) (Figure 2.25.). In order to freeze the short lived DNA-enzyme complex, the methylation target sequence of M.Hae III, 5'-GGCC-3', was modified to 5'-GG**F**C-3', whereas F stands for 5-fluoro-2'-deoxycyitidine. When attempting to methylate this modified substrate sequence, M.Hae III remains covalently bound to its DNA substrate (Figure 2.24.).

A remarkable feature of M.Hae III structure is the extrusion of the substrate cytosine from the DNA helix and its insertion into a deep concave pocket on the enzyme. This "base-flipping" mechanism, more accurately described as nucleotide-flipping, enables the enzyme to gain access to the substrate without interference from the surrounding duplex; equally important, the concave active site pocket provides a desolvated microenvironment within which catalysis can be efficiently performed. M.Hae III (as other DCMtases) are composed of two domains. The larger domain, which contains the active site, has the Rossman fold common to many nucleotide-binding proteins. The smaller domain, which makes extensive sequence-specific contacts in the major groove of the DNA, is highly unusual in the sense that it contains little if any repetitive secondary structure (α-helix or β-sheet) (Verdine and Norman, 2003).

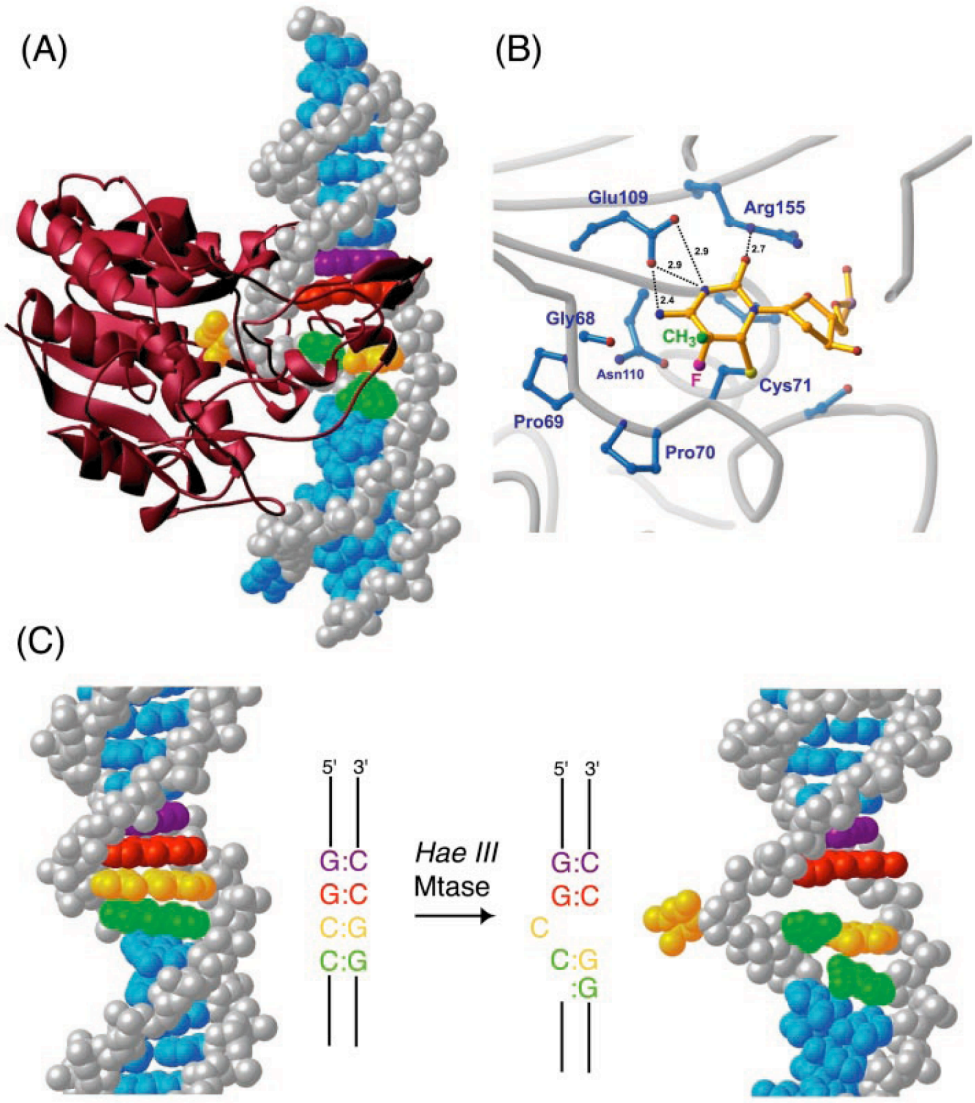

Figure 2.25. The structure of the *Hae* III methyltransferase trapped on an oligonucleotide containing 5-fluoro-2'-deoxycytidine. (*A*) Overall structure of the trapped protein-DNA complex. The protein is depicted as a ribbon trace, and the DNA is in space-filling model, with the bases of the recognition site color-coded; bases outside the recognition site are blue. Note that substrate cytosine (yellow) is extrahelical and inserted into an active site pocket on the protein. (*B*) Close-up view of the *Hae* III methyltransferase active site. The protein backbone is depicted as a tube model in gray, with particular side chains shown in blue. The substrate cytosine is shown as a gold framework, with the 5-methyl group in green and 5-fluoro group in magenta. Note the covalent addition of Cys 71 to C6 and the key interaction of Glu 109 with N3 of the substrate. (*C*) Space-filling model of canonical B-form DNA (*left*) and the DNA in the crystal structure of the trapped *Hae* III-DNA complex (*right*) (adapted from Verdine and Norman, 2003).

M.Hae III is one of the smallest methylases known and has a molecular mass of 37.7 kDa. Besides its canonical target sequence 5'-GGCC-3', M.Hae III was shown

to methylate with lower efficiency non-canonical sites in DNA (Cohen *et al.*, 2002). M.Hae III was cloned and produced in *E.coli* (Chen *et al.*, 1991), and additionally, it was expressed in its active form in *in vitro* transcription-translation mixture within the water compartments of a water-in-oil emulsion (Tawfik and Griffiths, 1998). Therefore, it seemed attractive to us to use M.Hae III as a cross-linking domain to create a covalent linkage between protein and its encoding DNA.

3. Results

3.1. Characterisation of water-in-oil emulsions

A basic requirement for covalent DNA display is the ability to produce a water-in-oil emulsion with individual compartments, whose size and stability (no fusion with other compartments) can be controlled. The water-in-oil emulsions were prepared by stirring a mixture of oil phase and water phase. The latter contained the DNA template and *in vitro* transcription/translation mixture. The time and the velocity of stirring as well as the amount of surfactants in the oil phase are critical parameters for the shape of the size distribution of the water compartments of the water-in-oil emulsion. Water in-oil-emulsions were prepared as described in the literature (Tawfik and Griffiths, 1998) with minor modifications, and the size distribution of the water compartments of the emulsions was determined by laser diffraction.

3.1.1. Reproducibility of emulsion preparation

Three emulsions were prepared independently by mixing 50 μl of water phase with 950 μl of oil phase on a magnetic stirrer and the corresponding size distributions were measured by laser diffraction (Figure 3.1.). The average diameter of the water compartments was about 1 μm with 67.5% (±3.5% s.d., n=3) of the water volume being in compartments with diameters ranging from 0.71 to 1.84 μm (area under the curve).

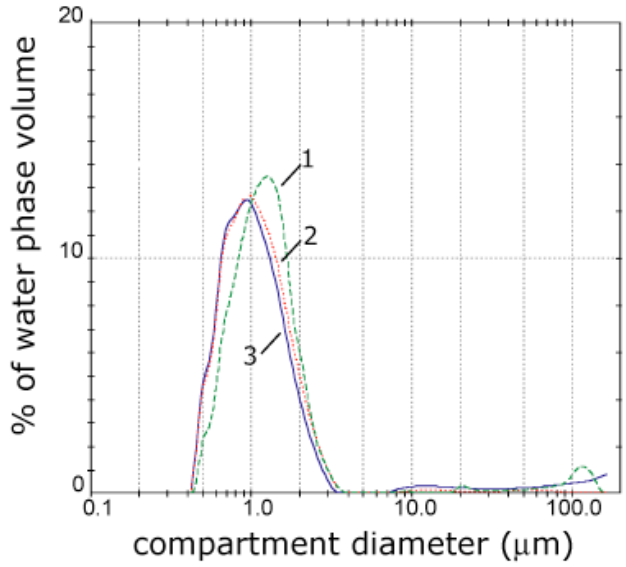

Figure 3.1. Reproducibility of emulsion preparation: three emulsions were independently prepared as described above (curves 1, 2 and 3). The size distributions were very similar.

3.1.2. Stability of water-in-oil emulsions

In order to determine the stability of the water-in-oil emulsions, we checked whether the size distribution of the water compartments changed over time. One emulsion was prepared and the size distribution of the water compartments was determined immediately using only a small fraction of the emulsion. The remaining part of the emulsion was left for 96 h at room temperature until the second measurement was performed (Figure 3.2.). The profiles of the two curves obtained at the different time points did not differ significantly from each other, indicating a remarkable stability of the water-in-oil emulsions. These results are supported by data published in the literature, where water-in-oil emulsions stabilised by the addition of surfactants were stable even at temperatures of more than 90°C (Ghadessy *et al.*, 2001).

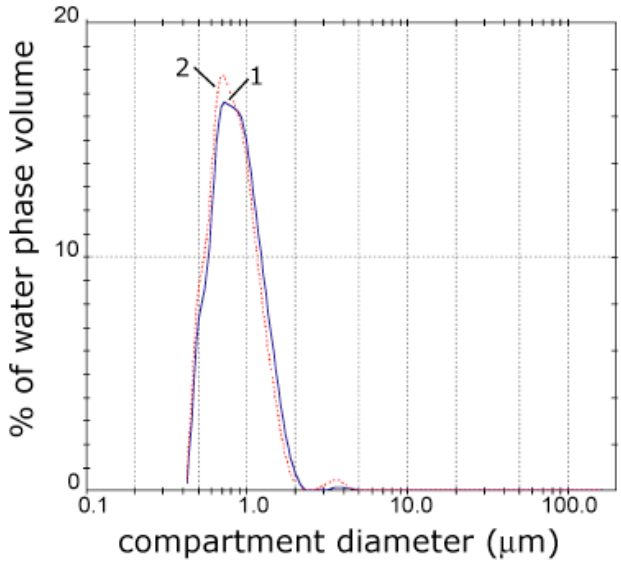

Figure 3.2. Stability of water-in-oil emulsions: an emulsion was prepared and size distributions of water compartments were measured after storage at room temperature for 0 h (1) and 96 h (2), respectively.

3.2. Formation of covalent DNA-protein adducts

We could show that M.Hae III can be linked covalently to DNA fragments which contain the modified substrate sequence for the DNA-methyltransferase (5'-GGFC-3') by incubating DNA (2 nM), S-adenosylmethionine (SAM, 80 μM) and M.Hae III (38 nM) for different time periods (Chen *et al.*, 1991). The concentrations of the reactants were chosen to be in the nanomolar range in order to mimic the conditions for cross-linking DNA with protein, assuming that one DNA molecule with about 15 molecules of M.Hae III are present in one compartment of a water-in-oil emulsion after transcription and translation. Reactions were stopped by heat inactivation of the DNA-methyltransferase (15 min at 70°C) and the outcome of the experiment was analysed on a denaturing TBE gel containing 8 M urea (Figure 3.3.). The heating step and the denaturing conditions during gel electrophoresis should allow the disruption of all non-covalent protein-DNA complexes.

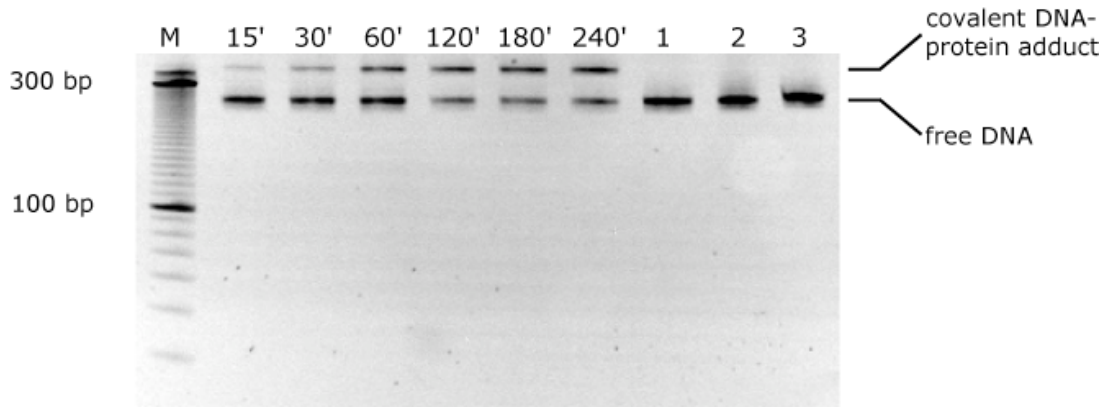

Figure 3.3. Covalent coupling of M.Hae III with a 268 bp DNA fragment containing the modified methylation site 5'-GGFC-3'. DNA (2 nM), M.Hae III (38 nM) and S-adenosylmethionine were incubated at 37°C for different time periods (given in min above the gel). M.Hae III was inactivated by heating (70°C, 15 min) and the samples were loaded on a denaturing polyacrylamide gel (10%, 8 M urea) and DNA was stained with SYBR green II. DNA molecules covalently bound to M.Hae III are shifted upwards as compared to free DNA (indicated in the figure). Three controls are shown on the right: 1: without S-adenosylmethionine; 2: without M.Hae III; 3: only 268 bp DNA fragment; M: 10 bp marker.

After a reaction time of 120 min, about 50% of the DNA molecules were covalently associated with one protein of M.Hae III DNA methyltranferase. The yield of DNA-protein adducts did not increase if the reaction was allowed to proceed for up to 240 min. If DNA containing the suicide inhibitor and Hae III DNA-methyltransferase were incubated without the cofactor SAM, no covalent protein-DNA adducts were detected. This result is in accordance with the reaction mechanism described in the literature (Figure 2.24.) (Chen *et al.*, 1991).

3.2.1. M.Hae III forms multivalent complexes with DNA

Hae III DNA-methyltransferase not only binds covalently to the modified target site 5'-GGFC-3' but has also been shown to interact non-covalently with DNA in different ways. Using a hemimethylated, double-stranded substrate DNA, the dissociation constants of wild-type M.Hae III and two mutants for the substrate (K_{dS}) were measured in presence of 80 μM S-adenosylhomocysteine (SAC) by electrophoretic mobility shift assays (EMSA) (Chen *et al.*, 1993). Wild-type M.Hae III bound the hemimethylated DNA substrate with a dissociation constant of 0.34 nM, whereas the C71A and the C71S mutants had a K_{dS} of 0.35 and 0.48 nM,

respectively. In presence of S-adenosylmethionine, the affinity of the C71A mutant for the hemimethylated target sequence was measured to be 20 pM, implying a very strong affinity of wild-type M.Hae III for its hemimethylated target sequence. In contrast, fully methylated target sequences were bound by both wild-type and mutant proteins no more tightly than nonspecific DNA ($K_{dS} \approx$ 12.5 nM). Unmethylated DNA sites were bound less tightly by C71A ($K_{dS} \approx 0.36$ nM) as compared to hemimethylated DNA. Since methylation of viral DNA would allow it to escape inactivation by a restriction endonuclease, evolution would be expected to select for DNA-methyltransferases that act on hemimethylated sites rather than on unmethylated sites. The data obtained from the EMSA experiments support this hypothesis.

As a consequence, M.Hae III can bind at various sites to DNA fragments with different affinities. Assuming (in the presence of 80 μM SAM) all target sites of M.Hae III to be fully methylated, non-covalently associated M.Hae III would be bound to DNA with an K_{dS} of about 12.5 nM.
In order to analyse the non-covalent binding of M.Hae III to DNA, a short DNA fragment (268 bp) containing one modified target sequence 5'-GGFC-3' was incubated with M.Hae III for 2 h at 37°C. In control experiments, either M.Hae III or the cofactor SAM was omitted. After incubation, the samples were divided in two and subsequently either heated to 70°C or kept on ice for 15 min. The experiment was analysed on a native 6% TBE gel and DNA was visualised by SYBR green I dye (Figure 3.4.). When the sample was not heated before loading on the gel, several M.Hae III molecules (1 – 9) were associated with one DNA, thus forming a DNA ladder of which each step represented DNA molecules complexed with a certain number of M.Hae III proteins. On the contrary, if the samples were heated, only one step of the ladder remained on the gel, proving that covalent cross-linking of DNA with M.Hae III can occur only in presence of the cofactor SAM.

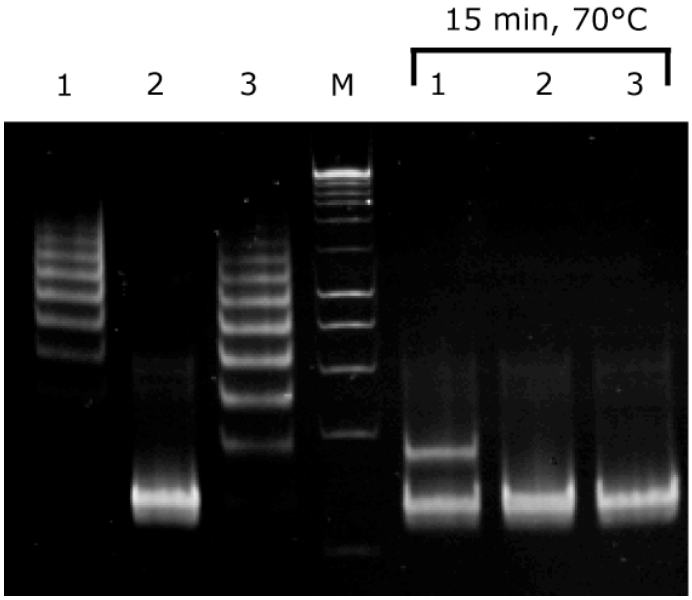

Figure 3.4. A short DNA fragment (268 bp) was incubated with M.Hae III and SAM for 2 h at 37°C (lanes 1). In control experiments, either M.Hae III (lanes 2) or SAM (lanes 3) were omitted. Before loading the samples on a native 6% TBE gel, one half of each sample was heated to 70°C for 15 min. DNA was visualized by SYBR green I dye. M = DNA marker (Smart Ladder, Eurogentech). Without heating of the samples and in presence of enzyme, DNA molecules were associated with several M.Hae III molecules and thus, forming a DNA ladder. If the samples had been heated before loading on the gel, only one M.Hae III molecule covalently bound to DNA could be detected (lane 1, right half of the gel). If M.Hae III (lane 2, right half) or cofactor SAM (lane 3, right half) was omitted, no covalent DNA-adduct was formed.

3.3. Expression of M.Hae III fusion proteins

In order to use M.Hae III as a cross-linking domain to create a covalent bond between a randomised polypeptide fused to M.Hae III and its encoding DNA, M.Hae III as well as fusion proteins thereof are required to be efficiently expressed in *Escherichia coli* as well as in *in vitro* transcription-translation mixture. For this purpose, M.Hae III-6xHis tag was cloned in *E.coli* and linear DNA templates coding for a panel of M.Hae III fusion proteins were prepared.

3.3.1. Cloning and expression of M.Hae III-6xHis tag in *E.coli*

The gene coding for M.Hae III was cloned in an *E.coli* expression vector (pQE-12) using the restriction sites EcoRI and HindIII. By using these restriction sites,

M.Hae III is expressed with an 6xHis tag to facilitate purification. Transformed *E.coli* cells (clone pJB18/1) were grown in 100 ml medium in shake flasks and protein expression was induced by the addition of IPTG. Proteins were isolated from the cytoplasm of the bacteria using a Ni-coated resin for affinity purification. M.Hae III was expressed at ~10mg/L of bacterial culture and was shown to be catalytically active (data not shown).

3.3.2. Preparation of DNA templates for *in vitro* expression

Five DNA templates were assembled by PCR coding for M.Hae III-6xHis tag, M.Hae III-Flag tag, M.Hae III-CaM-6xHis (calmodulin), M.Hae III-EDB-6xHis (extra-domain B of fibronectin) and GST-M.Hae III-6xHis (glutathione-S-transferase of *Schistosoma japonicum*). All polypeptides except GST were C-terminal fusions to M.Hae III and connected by the following linkers: 6xHis tag: GGGS; Flag tag: GGGSG; M.Hae III-GGSGAS-(ED-B/CaM)-GGGS-6xHis tag. DNA templates directly used for *in vitro* expression were not purified over agarose gel, because traces of remaining agarose can inhibit protein expression.

3.3.3. Expression of M.Hae III fusions in solution

Each M.Hae III fusion protein was expressed using 200–250 ng of DNA template in a volume of 25 μl of *in vitro* transcription/translation mixture, and after incubation at 30°C for 3.5 h, expression levels of the fusion proteins were analysed by taking a 2 μl volume for western blotting. Proteins were detected using either anti-His tag–horseradish peroxidase (HRP) immunoconjugate (for M.Hae III-6xHis tag, -CaM-6xHis tag, -EDB-6xHis tag and GST-M.Hae III-6xHis tag) or anti-Flag antibody (M2) in combination with a secondary anti-mouse-HRP detection reagent (for M.Hae III-Flag tag) (Figure 3.5.). All constructs except GST-M.Hae III-6xHis were well expressed and appeared to have the correct size, without any detectable proteolytic fragments. Expression yields for M.Hae III-6xHis tag was estimated from signal intensities on western blots to be around 300 ng/μl in *in vitro* transcription and translation mixture, which would give an expression yield of total 15 μg protein in a standard 50 μl *in vitro* transcription-translation reaction (data not shown). Taking into account the molecular weight of M.Hae III-6xHis (~40 kDa), 300 ng/μl of M.Hae III protein corresponds to a protein concentration of 7500 nM.

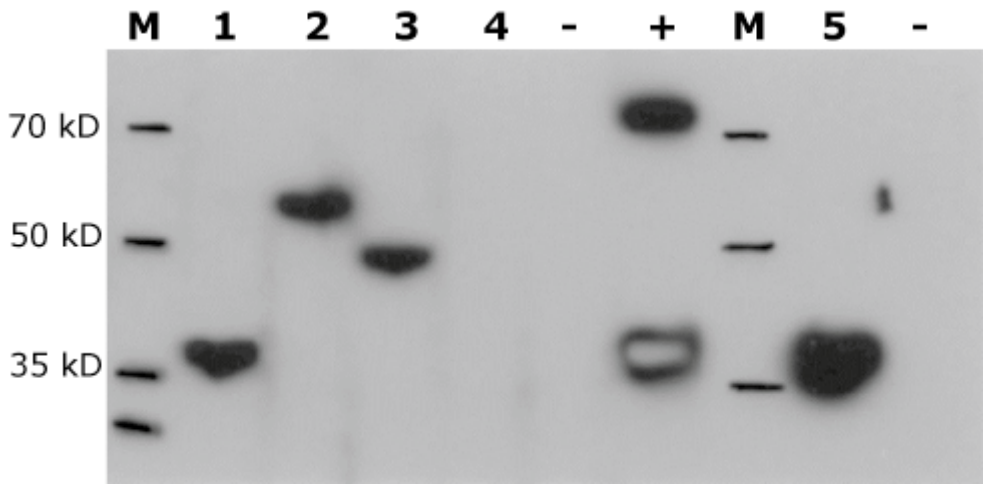

Figure 3.5. *In vitro* expression of M.Hae III fusion proteins. Linear DNA templates (200-250 ng) coding for different M.Hae III fusion proteins were incubated in an *in vitro* transcription-translation mix. A 2 μl volume was taken for western blot analysis [detection: anti-His-HRP or anti-Flag (M2)/anti-mouse-HRP]. M, marker; 1, M.Hae III-6xHis tag (total molecular weight, 38 kDa); 2, M.Hae III-calmodulin-6xHis (56 kDa); 3, M.Hae III-ED-B-6xHis (extra-domain B of fibronectin) (49 kDa); 4, GST-M.Hae III-6xHis tag (*S. japonicum*) (64 kDa); 5, M.Hae III Flag tag (38 kDa); +, ~1.5 μg recombinantly expressed M.Hae III-6xHis tag (38 kDa); –, transcription-translation mix. All peptides/proteins except GST were C-terminal fusions to M.Hae III.

The fusion proteins expressed *in vitro* all were catalytically active as shown by an activity assay for M.Hae III. A DNA fragment (1094 bp) containing a NotI restriction site (5'-GCGGCCGC-3') was incubated for 1 h at 37°C with commercially available M.Hae III or 0.1 μl of *in vitro* transcription-translation mixtures containing freshly synthesised M.Hae III fusion proteins. In the presence of M.Hae III, the 5'-GGCC-3' sequence comprised in the NotI restriction site should be methylated, thus protecting the DNA fragment from NotI digestion. After heat inactivation of M.Hae III, $MgCl_2$ (0.1 M) and NotI were added, and the samples were incubated for an additional hour at 37°C. As shown in Figure 3.6., DNA was protected from NotI digestion in all cases tested but the GST-M.Hae III fusion construct. These results correspond well to the expression data obtained by western blotting (Figure 3.5.).

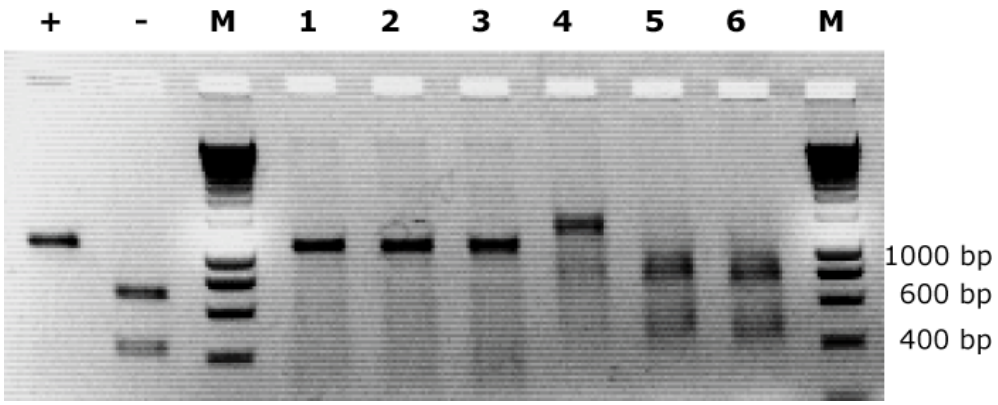

Figure 3.6. Assay for the catalytic activity of M.Hae III fusion proteins expressed *in vitro*. Only unmethylated DNA can be digested by NotI restriction enzyme yielding two fragments (679 bp and 415 bp). Samples were analysed on an agarose gel stained with ethidium bromide. +: commercially available M.Hae III; -: without enzyme; 1-6: 0.1 μl of *in vitro* expression mixture. 1: M.Hae III-6xHis tag; 2: M.Hae III-Flag tag; 3: M.Hae-CaM-6xHis tag; 4: M.Hae EDB-6xHis tag; 5: GST-M.Hae III; 6: transcription/translation mixture only; M: DNA marker.

3.3.4. Expression of M.Hae III fusions in emulsion

In order to determine expression yields of M.Hae III fusion proteins in the compartments of a water-in-oil emulsion, 300 ng of each DNA template was incubated in emulsion for 3 h at 30°C (50 μl water volume). The proteins were extracted from the emulsion in a volume of 100 μl, and 10 μl of this solution was used for western blot analysis (Figure 3.7.). Protein detection was performed as above (Figure 3.5.).

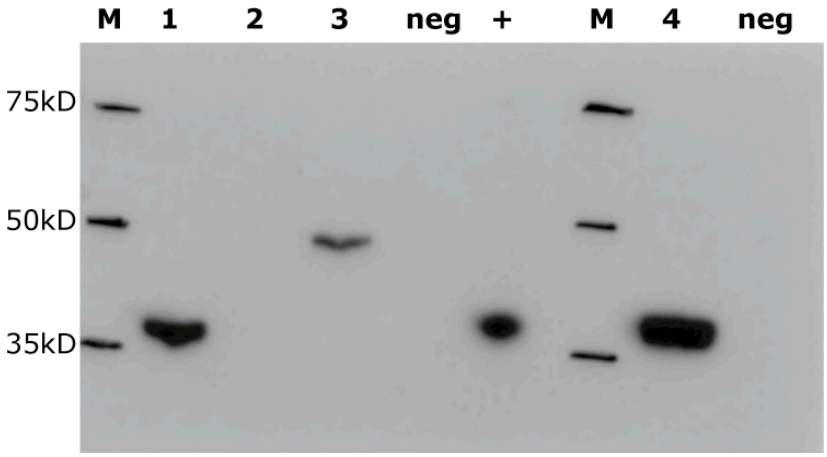

Figure 3.7. *In vitro* expression of M.Hae III fusion proteins in emulsion. Linear DNA templates coding for different M.Hae III fusion proteins (300 ng) were incubated in an *in vitro* transcription-translation mix. A 2 μl volume was taken for western blot analysis [detection: anti-His-HRP or anti-Flag (M2)/anti-mouse-HRP]. M, marker; 1, M.Hae III-6xHis tag (total molecular weight, 38 kDa); 2, M.Hae III-calmodulin-6xHis (56 kDa); 3, M.Hae III-ED-B-6xHis (extra-domain B of fibronectin) (49 kDa); 4, M.Hae III-Flag tag (38 kDa); +, ~150 ng recombinantly expressed M.Hae III-6xHis tag (38 kDa); neg, transcription-translation mix.

Expression levels in emulsion were markedly lower compared to expression of the M.Hae III fusion proteins in solution. From the intensities of the bands on the western blot (Figure 3.7.), the amount of M.Hae-6xHis tag protein was estimated to be ~300 ng (compare with + = 150 ng). Therefore, the total amount of M.Hae III-6xHis tag expressed in emulsion was around 3 μg (10 μl of 100 μl were loaded on gel), which corresponds to 20% of the yield achieved by expression in solution (see above), implying an enzyme concentration of about 1500 nM. This concentration can be expected to promote efficient cross-linking of M.Hae III and DNA. Unfortunately, expression of M.Hae III-calmodulin-6xHis tag was not detectable.

3.4. Stability of linear DNA in transcription/translation mixtures

In order check for the formation of covalent M.Hae III-DNA adduct formed in *in vitro* transcription and translation mixture, DNA template coding for M.Hae-6xHis tag was incubated with recombinantly produced M.Hae III protein in *in vitro* transcription-translation mixture. After incubation at 37°C, the reaction was analysed on a polyacrylamide gel by staining DNA with SYBR green I dye.

Surprisingly, neither a band representing the covalent M.Hae III-DNA adduct nor a band for the unmodified DNA was detected. Additional experiments revealed that linear DNA seemed to be unstable in *in vitro* transcription-translation mix. DNA fragments (1500 bp) were incubated for either 15, 30 or 60 minutes in 1:20 diluted transcription-translation mixture. Incubations were stopped by heating (15 min at 70°C) to inactivate nucleases eventually present in the *E.coli* lysate. Samples were then loaded on a polyacryamide gel and DNA was stained using SYBR green I dye. The experiment clearly showed a time dependent degradation of DNA in the transcription-translation mixture. Degradation could be reduced to some degree by the addition of an excess of irrelevant DNA (Figure 3.8.), probably by transient saturation of nuclease-active sites.

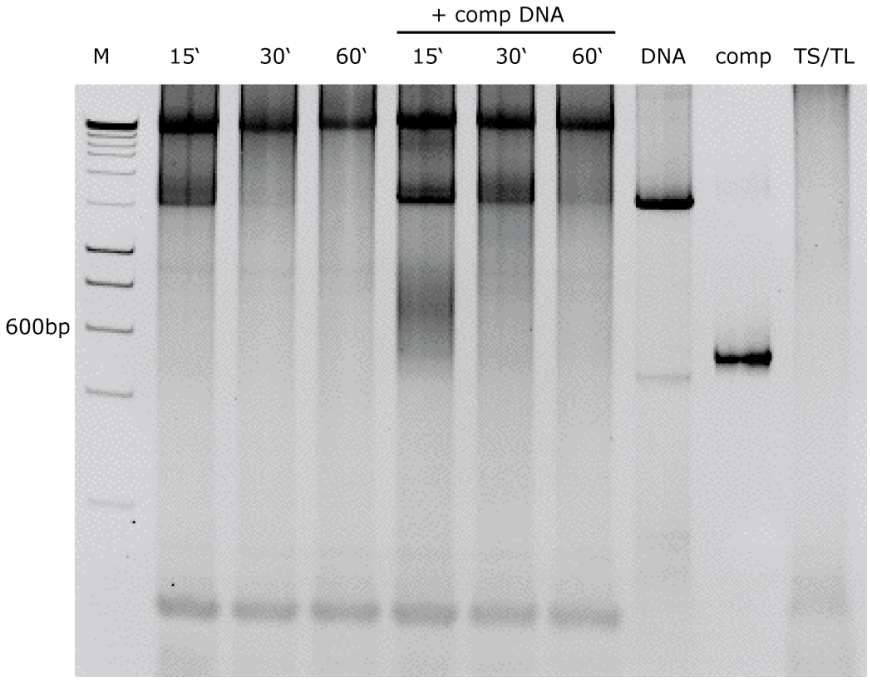

Figure 3.8. Linear DNA (1500 bp) was incubated in *in vitro* transcription-translation mixture at 37°C either in presence or absence of competing irrelevant DNA (~500 bp). Incubation was stopped after different time periods by heating for 15 min to 70°C. Samples were loaded on a native polyacrylamide gel for analysis. Incubation times are indicated in minutes above the gel. DNA: only 1500 bp DNA fragment; comp: only competing DNA; TS/TL: transcription-translation mix only. DNA was stained with SYBR green I dye.

Degradation of DNA in *in vitro* transcription-translation mixture is very undesirable because the carrier of genetic information would be lost in the selection procedure shown in Figure 2.22. The fact that competing DNA could

prolong DNA survival, implies the existence of exonucleases digesting the DNA molecules in the transcription-translation mixture.

3.4.1. DNA stability is temperature dependent

In further experiments, degradation of DNA in transcription-translation mixture was investigated at different temperatures. Surprisingly, DNA was much more stable if incubated in 1:10 diluted transcription-translation mixture at 30°C instead of 37°C. After incubation, the samples were heated for 15 min to 70°C to stop nucleolytic activity in the transcription-translation mixture and then loaded on a native polyacrylamide gel (Figure 3.9.).

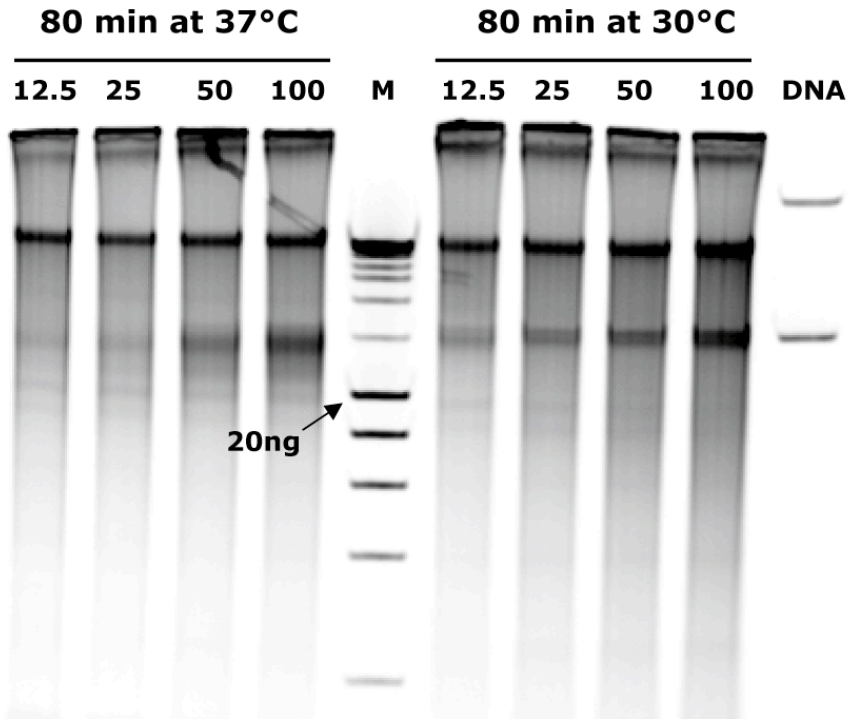

Figure 3.9. Different amounts (12.5-100 ng, indicated above the gel) of linear DNA (1500 bp) were incubated in 1:10 diluted transcription-translation mixture for 80 min at either 37°C or 30°C. Before loading the samples on the native polyacrylamide gel, they were heated to 70°C during 15 min. DNA was stained with SYBR green I dye. M: DNA marker; DNA: 25 ng of DNA incubated in water instead of transcription-translation mix.

For the samples incubated at 37°C, bands were not only weaker but also more diffuse compared to incubation at 30°C, thus indicating again that exonucleases present in the transcription-translation mixture might be responsible for DNA degradation.

3.4.2. Formation of covalent DNA-protein adducts in transcription-translation mixture

The mechanism based inhibitor 5-fluorodeoxycytidine is introduced in DNA templates during PCR by using a primer containing the modified methylation sequence 5'-GGFC-3'. Therefore, the mechanism based inhibitor, situated at the extremity of the DNA template, is vulnerable towards exonucleases degrading the ends of the DNA template. In order to check whether covalent cross-linking in transcription-translation is hampered by the existence of exonucleases, two DNA templates were created. In the first DNA molecule, 5-fluorodeoxycytidine was placed at the extremity, and in the second DNA molecule, the inhibitor was located at an internal site about 300 bp away from the end of the DNA (Figure 3.10.). The DNA templates with the internal 5-fluorodeoxycytidine were created by ligation of two DNA fragments one of which carried the mechanism based inhibitor for M.Hae III.

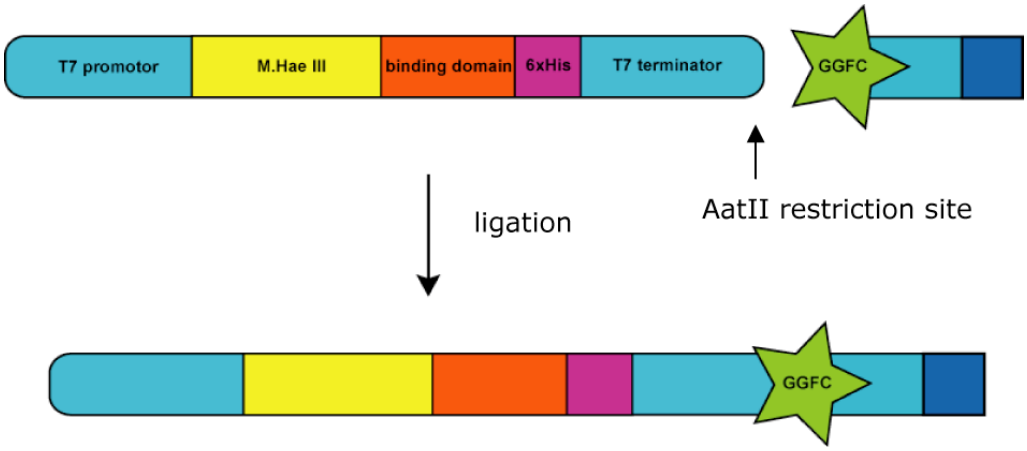

Figure 3.10. Construction of a DNA template with an internal cross-linking site 5'-GGFC-3'. Two DNA fragments amplified by PCR were ligated after digestion with restriction nuclease AatII.

Both DNA templates were incubated in 1:10 diluted transcription-translation mixture in the presence of cofactor SAM and 1 μM recombinant M.Hae III protein. In control experiments, DNA was incubated with enzyme in methylase reaction buffer with or without SAM. After 3 h at 30°C, all samples were heated to 70°C for 15 min and loaded on a native polyacrylamide gel to visualise the outcome of the experiment (Figure 3.11.).

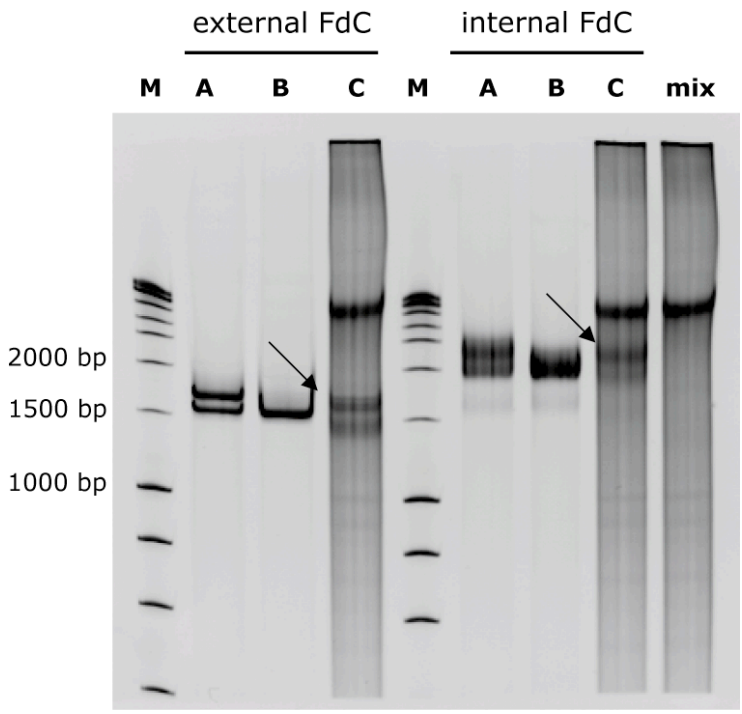

Figure 3.11. DNA templates (80 ng/sample) containing 5-fluorodeoxycytidine (FdC) either at the extremity (external FdC, 1500 bp) or at an internal site (internal FdC, 1850 bp) were used for cross-linking experiments. Samples were incubated for 3 h at 30°C, heated for 15 min to 70°C and loaded on a native polyacrylamide gel. DNA was visualised by SYBR green I staining. Arrows indicate the DNA band representing the covalent DNA-protein adduct. A: DNA incubated with cofactor SAM (80 μM) and recombinant M.Hae III (1 μM); B: As *A*, but without cofactor; C: as *A*, but in transcription-translation mixture diluted 1:10; mix: only transcription-translation mixture treated as the other samples; M: DNA marker.

Despite being blurred in the samples with transcription-translation mix, the two bands representing covalently cross-linked (arrows) and unmodified DNA are clearly visible on the gel (Figure 3.11., C). This finding indicates that DNA templates incubated at 30°C are stable and possess an intact mechanism based inhibitor site at their extremity.

One could now object, that the stoichiometry between transcription-translation mix and DNA molecules was not realistic. In the experiment above, 80 ng of DNA were dissolved in 10 μl of 1:10 diluted transcription translation mixture. The dilution of transcription-translation mixture is necessary because such solutions cannot be applied to gel electrophoresis undiluted, probably due to high concentration of salts and proteins. If one wanted to repeat the experiment at the same stoichiometric conditions in 50 μl of a 100% transcription-translation mixture, the addition of 4 μg DNA (2.5×10^{12} molecules) would be necessary. This

amount of DNA, however, is far more than would be ever used for a real selection experiment because the number of DNA molecules ($2.5x10^{12}$ molecules) would exceed the number of compartments in the water-in-oil emulsion ($\sim 9x10^{10}$) (see Table 2.3).

If one wanted to perform the experiment with less DNA (e.g. 100 ng $\approx 6x10^{10}$ molecules), one would not be able to visualise the DNA on the polyacrylamide gel due to the background staining caused by the transcription-translation mixture.

3.5 Affinity capture of DNA-protein adducts

In order to show that DNA molecules could be isolated based on the protein they encode, capture experiments using DNA templates in combination with various capture agents immobilised on two different types of magnetic beads were performed. Binding DNA molecules were amplified from the beads by PCR without prior elution and the experiments were interpreted either by analysis of the PCR products on agarose gels or by real-time PCR.

3.5.1. Capture on magnetic beads coated with Ni^{2+}

It should be possible to capture DNA templates coding for M.Hae III-6xHis tag by using magnetic beads coated with Ni^{2+}. In order to test this hypothesis, DNA fragments either encoding M.Hae III-6xHis tag or –Flag tag were incubated in transcription-translation mixture for expression and cross-linking. Magnetic beads coated with nickel were used to capture DNA-protein adducts. Unbound DNA was removed by washing the beads with imidazole containing wash buffer with 0.1% Tween 20 and PBS, and remaining DNA was amplified by performing PCR directly on the beads with suitable primers (Figure 3.12.). The outcome of the experiment was visualised by loading the PCR products on an agarose gel containing ethidium bromide.

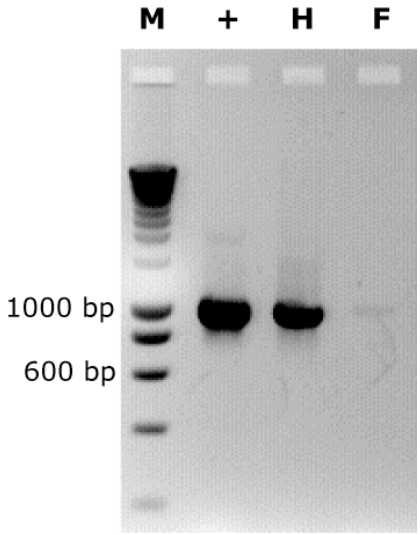

Figure 3.12. Affinity capture of DNA-M.Hae III-6xHis adducts. DNA templates coding for M.Hae III-6xHis tag or –Flag tag were incubated in transcription-translation mixture at 30°C for protein expression and cross-linking of DNA and protein. DNA-M.Hae III fusions were purified on magnetic nickel beads, and after washing, remaining DNA was detected by PCR. Products of the PCR reactions were loaded on an agarose gel and DNA was stained with ethidium bromide. M: DNA marker, +: DNA incubated with recombinant, purified M.Hae III-6xHis tag in methylase reaction buffer, H: DNA coding for M.Hae III-6xHis tag as input, F: DNA coding for M.Hae III-Flag tag as input.

DNA molecules were only efficiently recovered, if they were associated with M.Hae III-6xHis tag proteins. In contrast, DNA coding for M.Hae III-Flag tag was not selected. Thus, it was shown that DNA molecules can be selected via the protein they encode.

As shown in Figure 3.4., DNA and M.Hae III proteins can be bound to each other in two ways: first, they can form a covalent adduct via the mechanism based inhibitor 5-fluorodeoxycytidine or secondly, they can form multivalent, non-covalent complexes. Under native conditions, both types of interactions will contribute to the selection of DNA molecules coding for target binding M.Hae III fusion proteins. On the contrary, affinity selection under denaturing conditions should allow the isolation of only those DNA-protein adducts, which are bound to each other in a covalent manner. Therefore, affinity selection experiments were performed, where DNA-protein adducts were captured in a buffer containing 6 M guanidinium hydrochloride (GuHCl), and the nickel-coated beads were washed with buffer comprising 8 M urea. As above, remaining DNA molecules were amplified by PCR on the magnetic beads, and the outcome of the experiment was analysed on agarose gel.

Concretely, 100 ng of DNA and purified M.Hae III-6xHis tag enzyme (1 μM) were incubated in methylase reaction buffer either in the presence or absence of cofactor SAM. Cross-linking was allowed to proceed for 2.5 h at 30°C. Subsequently, an excess of Herring sperm DNA (1mg/ml) followed by denaturing buffer (6 M GuHCl) and magnetic nickel-beads were added. This mixture was incubated for 1 h at room temperature whilst shaking at 200 rpm. The beads were washed extensively with buffer containing 8 M urea and Herring sperm DNA. In order to remove denaturing agents before PCR, the beads were further washed with PBS Tween 0.1%. The amount of DNA recovered was visualised on agarose gel (Figure 3.13.).

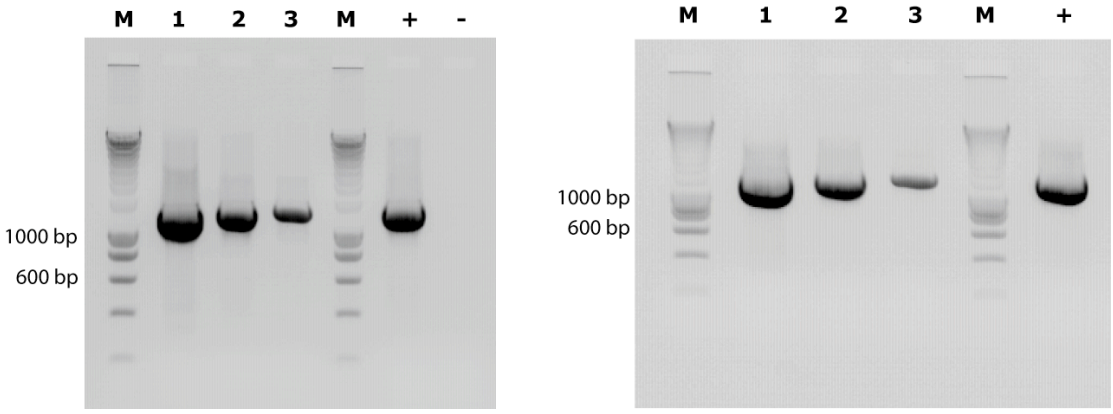

Figure 3.13. DNA templates containing the mechanism based inhibitor 5-fluorodeoxycytidine were incubated with M.Hae III-6xHis tag enzyme for 2.5 h at 30°C in the presence (1) or absence (2) of cofactor SAM. In sample (3), no enzyme was added. After incubation, M.Hae-DNA fusions were captured in 6 M GuHCl and Herring sperm DNA containing buffer on nickel-coated magnetic beads, which were then washed with buffer comprising 8 M urea with Herring sperm DNA to remove unbound DNA. Denaturing agents were removed by washing with PBS Tween 0.1% and remaining DNA molecules were amplified by PCR. For analysis, PCR products were loaded on agarose gel and stained with ethidium bromide. Affinity capture was most efficient, if M.Hae III-6xHis tag could bind covalently to DNA (1). The experiment was performed twice on different days. M: DNA marker, +/-: positive and negative control of PCR reaction.

Figure 3.13. shows an advantage in affinity selection for DNA molecules connected by a covalent bond to M.Hae III-6xHis protein. However, also in sample 3 (without enzyme) of Figure 3.13., a considerable amount of DNA was still present on the beads after washing, indicating a relatively strong unspecific interaction between DNA and beads under denaturing conditions. Interestingly, DNA recovery was higher than background, even if M.Hae III-6xHis was non-covalently associated with DNA (Figure 3.13., lanes 2). There are two possible

explanations for this result: first, not all M.Hae III proteins were denatured, or alternatively, the nickel-coated beads became stickier in the presence of denatured M.Hae III DNA-methyltransferase on their surface, and thus, absorbed unspecifically more DNA. The latter explanation is more likely because background recovery of DNA was markedly higher if the capture step was performed in presence of bovine serum albumin (BSA).

Due to the limitations of the experiments shown in section 3.4.2., we envisaged to investigate whether M.Hae III-6xHis tag DNA-methyltransferase expressed *in situ* from a DNA template containing the cross-linking site could form a covalent bond with its encoding DNA, which would allow the selective capture of covalent protein-DNA fusions. For this purpose, DNA templates coding for M.Hae III-6xHis tag protein, either with (lanes 1 & 2) or without (lanes 3 & 4) 5-fluorodeoxycytidine, were produced. 60 ng of each of these two DNA templates were incubated in presence of SAM with (Figure 3.14., lanes 1 & 3) or without (Figure 3.14., lanes 2 & 4) additional recombinant M.Hae III-6xHis tag protein. This set of samples was either incubated in methylase reaction buffer or in *in vitro* transcription-translation mixture (Figure 3.14.). DNA-protein complexes were, as before, captured on nickel-coated magnetic beads and washed under denaturing conditions as described above. After PCR amplification of affinity-purified DNA molecules, the experiment was analysed on an agarose gel stained with ethidium bromide (Figure 3.14.).

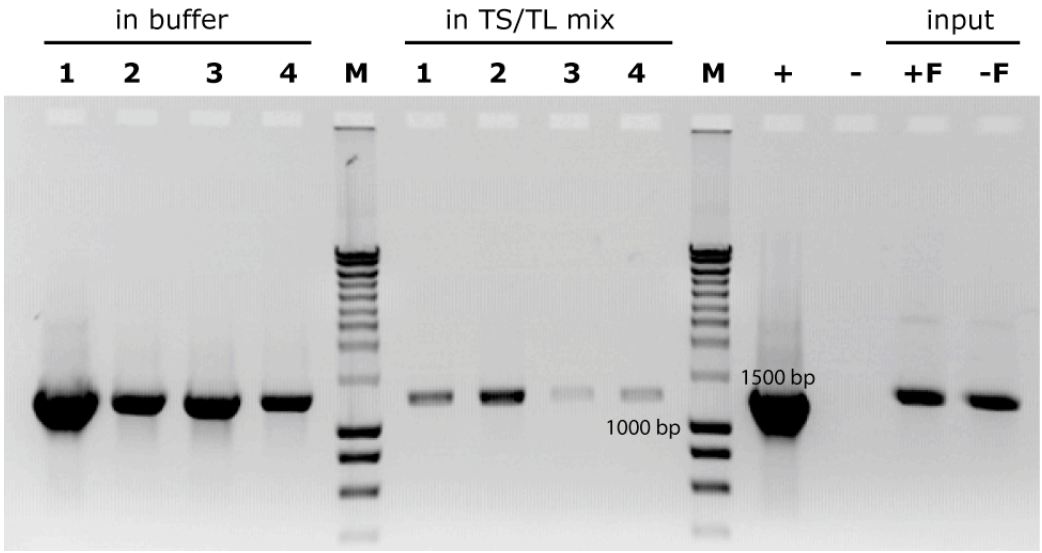

Figure 3.14. *In situ* formation of covalent DNA-protein fusions after expression from DNA templates coding for M.Hae III-6xHis tag. DNA templates (60 ng) with (lanes 1 & 2) or without (lanes 3 & 4) the inhibitor 5-fluorodeoxycytidine were incubated with SAM in the

presence (lanes 1 & 3) or absence (lanes 2 & 4) of additional recombinant M.Hae III-6xHis tag protein. This set of samples was either incubated in methylase reaction buffer or in *in vitro* transcription-translation mixture. Using nickel-coated magnetic beads, DNA-protein complexes were captured and washed under denaturing conditions as described above. After PCR amplification of affinity purified DNA molecules, the experiment was analysed on an agarose gel stained with ethidium bromide. The covalent association of M.Hae III-6xHis DNA-methyltransferase with DNA led to more efficient affinity capture of DNA-protein complexes. M: DNA marker, +/-; positive and negative controls for PCR reaction, F+/F-: input of DNA fragments with or without mechanism based inhibitor.

The relative intensities of the bands representing the samples which were incubated in methylase reaction buffer are in good correlation with the results shown in Figures 3.13. DNA molecules covalently bound to M.Hae III-6xHis tag were captured more efficiently than "naked" DNA or DNA, which did not include 5-fluorodeoxycytidine. If the samples were incubated in transcription-translation mixture, DNA fragments containing the cross-linking sequence 5'-GGFC-3' (lanes 1 & 2) were recovered with higher yields than compared to DNA without inhibitor (lanes 3 & 4). However, these DNA bands were weak and the difference in intensities small, so the experiment was repeated.

As above, DNA templates coding for M.Hae III-6xHis tag with (Figure 3.15., 1-4) or without (5 & 6) the inhibitor 5-fluorodeoxycytidine were incubated in transcription-translation mixture with SAM in the presence (1, 3 & 5) or absence (lanes 2, 4 & 6) of additional recombinant M.Hae III-6xHis tag protein. In control experiments, DNA comprising the cross-linking sequence was mixed with SAM and methylase reaction buffer, either with (1) or without (2) M.Hae III-6xHis DNA-methyltransferase (Figure 3.15.).

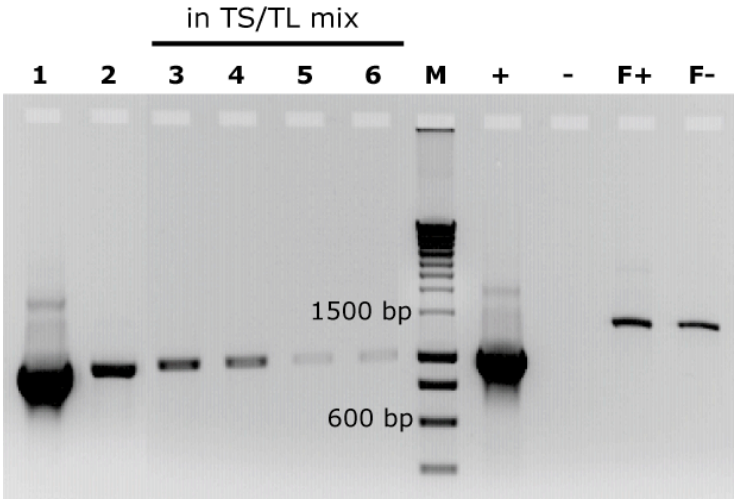

Figure 3.15. Selections under denaturing conditions were carried out as described above. In methylase reaction buffer, DNA containing the mechanism based inhibitor 5-fluorodeoxycytidine was efficiently captured on nickel-coated magnetic beads in the presence of recombinant M.Hae III-6xHis tag protein (1), whereas most DNA was washed away if there was no His-tagged DNA-methyltransferase (2). If incubation was carried out in transcription-translation mixture, DNA molecules were captured with more or less the same efficiency, not depending on the addition of M.Hae III-6xHis tag enzyme (3 & 4). However, DNA templates lacking 5-fluorodexycytidine were poorly recovered, even if purified His-tagged DNA-methyltransferase was present (5 & 6). M: DNA marker; +/-: positive and negative control of PCR; F+/F-: amount of input DNA with or without 5-fluorodeoxycytidine.

Figures 3.14. and 3.15. indicate that DNA templates coding for M.Hae III-6xHis tag and carrying the cross-linking sequence 5'-GGFC-3' can be selected by affinity purification due to *in situ* expression of the enzyme and its covalent cross-linking with the encoding DNA molecule. However, band intensities were weak and it would be preferable to increase the recovery of DNA containing 5-fluorodeoxycytidine.

3.5.2. Capture on magnetic beads coated with specific ligands

So far, nickel-coated magnetic beads were used for the affinity selection of DNA molecules associated with M.Hae III-6xHis tag enzyme. It would be very interesting to achieve the isolation of DNA coding for a certain binding phenotype by use of a solid phase, on which different selection targets can be easily immobilised. These requirements for a solid phases are fulfilled by magnetic

beads coated with streptavidin. Streptavidin-coated magnetic beads allow the facile immobilisation of any ligand, which can be biotinylated.

In order to test the performance of streptavidin beads coated with biotinylated antibodies for the capture of DNA molecules attached to a specific M.Hae III fusion protein, streptavidin beads were coated with either anti-His or anti-Flag tag antibodies. 50 ng of DNA fragments encoding M.Hae III-6xHis or –Flag tag were transcribed and translated *in vitro* and mixed with nickel-coated magnetic beads or streptavidin beads coated with either anti-His or anti-Flag tag antibodies. The beads were washed with PBS 0.1% Tween and PBS, followed by PCR amplification and analysis on agarose gel (Figure 3.16.). In lanes designated with 1, DNA coding for M.Hae III-6xHis tag was used for transcription and translation, whereas DNA encoding M.Hae III-Flag tag was taken for *in vitro* expression in lanes labeled with 2. Affinity capture on Ni^{2+}-coated beads was selective, whereas the background of DNA connected to non-binding protein was very high in the experiments with streptavidin beads (Figure 3.16.).

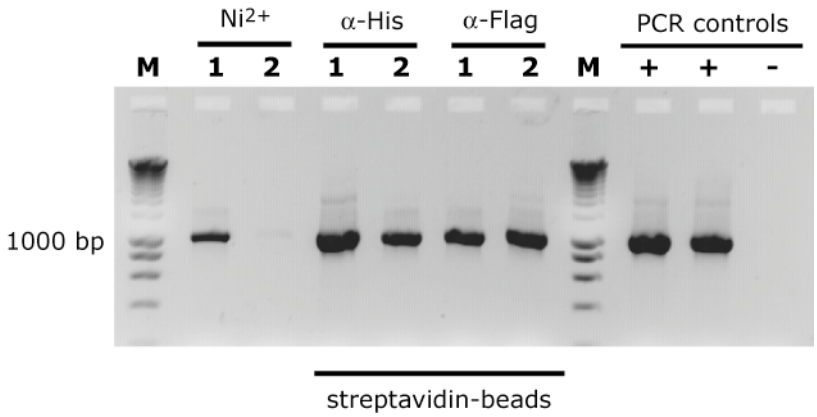

Figure 3.16. 50 ng of DNA templates coding for M.Hae III-6xHis tag (1) or –Flag tag (2) was used for *in vitro* expression. Resulting DNA-protein fusions were captured on nickel-coated beads (Ni^{2+}) or streptavidin beads coated with either anti-His antibody (α-His) or anti-Flag antibody (α-Flag). In the experiment where nickel-beads were applied, a selective capture of M.Hae III-6xHis tag DNA was observed. Unspecific binding of DNA to streptavidin beads was very strong, thus giving high background levels. M: DNA marker; +/-: positive and negative controls of PCR.

Several attempts were undertaken to lower the background of DNA binding non-specifically to streptavidin beads. Excess of irrelevant DNA was added to compete for unspecific binding to the beads, which did not help to improve the specificity of the affinity capture. Interestingly, if short, double-stranded DNA fragments

biotinylated at one end were used for the blocking of streptavidin beads, background was lowered dramatically (see below).

The next step was to investigate, whether affinity selection experiments would also work, if protein expression was performed within the water compartments of a water-in-oil emulsion. So far, all expression steps had been performed in transcription-translation mix under standard conditions in free solution.
Therefore, amounts of 100 ng of DNA molecules coding either for M.Hae III-6xHis tag or M.Hae III-Flag tag protein were emulsified separately. After protein expression at 30 °C for 2.5 h, the two emulsions were centrifuged for 10 min at 7’000 r.p.m.. The oil supernatants were removed and the water phases, containing the protein–DNA adducts, were extracted from the pellets formed by the still intact water compartments at the bottom of the tubes by addition of diethyl ether. Subsequently, the extracted water phases were dried to remove residual traces of diethyl ether and divided into two halves each. The first two halves of the water phases containing DNA-M.Hae III-6xHis or DNA-M. Hae III-Flag adducts were used for affinity selections with biotinylated anti-His and the second two halves were taken for pannings with biotinylated anti-Flag antibody immobilised on magnetic streptavidin beads. The beads had been blocked with short, biotinylated double-stranded DNA fragments, and the capture of DNA-protein fusions on the streptavidin beads was performed in the presence of these competing biotinylated DNA fragments.
DNA molecules still bound to the beads after washing were amplified by PCR (amplicon length = 1020 bp) without prior elution from the magnetic beads. The outcome of the selections was analysed on an agarose gel stained with ethidium bromide (Figure 3.17.). As expected, only DNA coding for M.Hae III-6xHis was selected if the affinity capture was performed with anti-His antibody (1), whereas DNA coding for M.Hae III-Flag tag could not be detected (2). However, if anti-Flag antibody was used for affinity selection, only DNA coding for M.Hae III-Flag was recovered and detected on the agarose gel after PCR (3). Therefore, we concluded that it is possible to select DNA molecules via the binding specificity of the encoded protein after its expression within the water compartments of a water-in-oil emulsion.

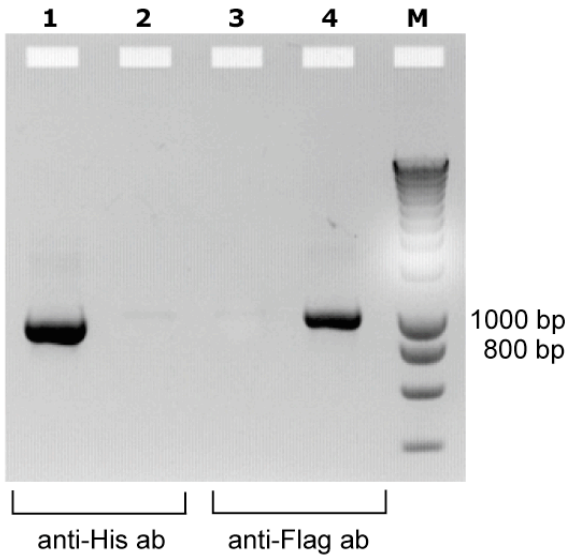

Figure 3.17. Model selections: 100 ng linear DNA templates ($6.6x10^{10}$ molecules) coding for either M.Hae III-6xHis tag or M.Hae III-Flag tag were emulsified with 50 μl of *in vitro* transcription/translation mixture. Covalent protein–DNA adducts were extracted from the emulsions after expression for 2.5 h at 30°C. Affinity panning experiments were done with magnetic streptavidin beads coated either with biotinylated anti-His or anti-Flag antibody. In order to amplify the selected DNA molecules, PCR was performed directly on the beads after washing. PCR products (1020 bp long) were loaded on an agarose gel. 1 & 3: DNA coding for M.Hae III-6xHis tag used for selection; 2 & 4: DNA coding for M.Hae III-Flag tag used for selection. Affinity panning with anti-His or anti-Flag antibody as indicated below the agarose gel. M: DNA marker.

It should be noted that the selection experiments, independent of whether expression was done in emulsion or in free solution, worked equally well and reliably (data not shown).

3.5.3. Quantification of DNA molecules in capture experiments by real-time PCR

For the measurement of the efficiency with which DNA–protein adducts were selected by virtue of their binding affinity to the immobilised antibody on the magnetic beads, model selection experiments were performed as described above using additional M.Hae III fusion constructs, and the number of recovered DNA molecules recovered was determined by real-time PCR.

Real-time PCR measurements were carried out by using a TaqMan® probe specific for the gene of interest (e.g., M.Hae III coding sequence). A TaqMan® probe is added to the PCR reagent master mix and is designed to anneal to a specific sequence of template between the forward and reverse primers used for the PCR reaction. TaqMan® probes are designed with a fluorescent dye termed a reporter at the 5'-end, and another fluorophore termed a quencher at the 3'-end. When this probe is intact and excited by a light source, the reporter dye's emission is suppressed by the quencher dye as a result of the close proximity of the dyes (Figure 3.18, A). The probe sits in the path of the enzyme as it starts to copy DNA (B). When the enzyme reaches the annealed probe the 5'-exonuclease activity of the enzyme cleaves the probe, and thus, the distance between the reporter and the quencher increases causing the transfer of energy to stop (C). The fluorescent emissions of the reporter increase and the quencher decrease. The increase in reporter signal is captured by a detector and displayed by the software. The amount of reporter signal increase is proportional to the amount of product being produced for a given DNA sequence.

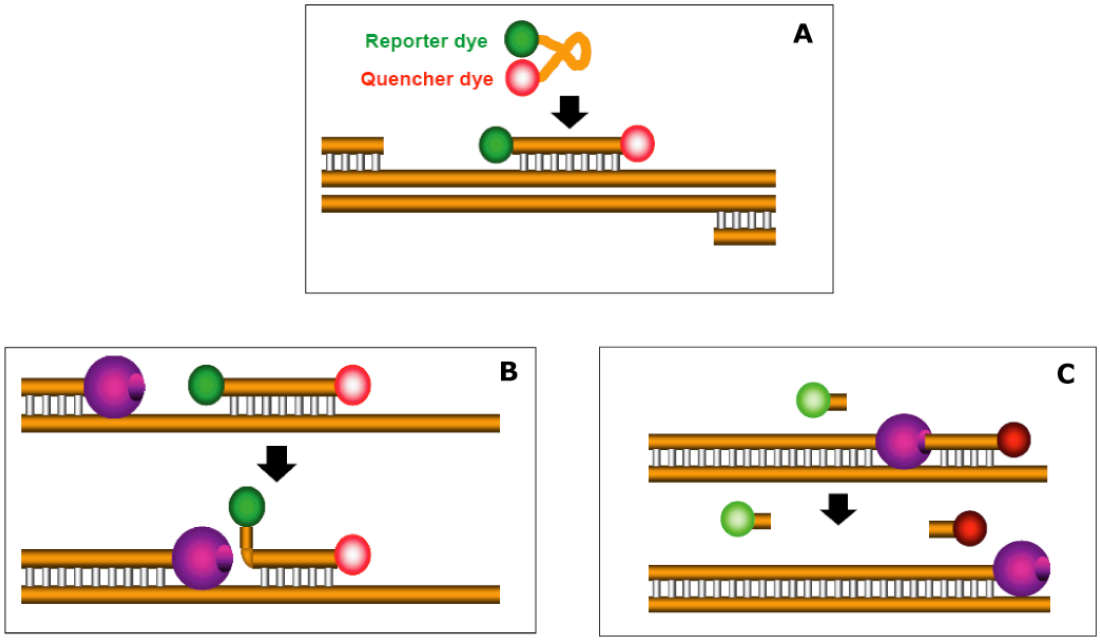

Figure 3.18. Schematic representation of real-time PCR chemistry using TaqMan® probes. A TaqMan® probe specific for the gene sequence of interest is designed together with two suitable primers, which anneal up- and downstream of the probe. The TaqMan® probe is labeled with two different fluorescent dyes: a reporter dye at the 5'-end and a quencher dye at the 3'-end. If the probe is not annealed to its sequence, the dyes are in close proximity, and therefore, the fluorescence of the reporter dye is quenched (*A*). However, when aligned to the target sequence (*B*), the probe sits in the way of the DNA polymerase

and is degraded by the 5'-exonuclease activity of the enzyme. Thus, the reporter and quencher fluorophores a separated from each other and the fluorescence of the reporter dye increases. The more target DNA sequence is present in the PCR solution, the more fluorescence is created in a PCR cycle, thus correlating fluorescence intensity with the number of DNA molecules present in the sample (adapted from Applied Biosystems).

During amplification of DNA in PCR the number of DNA copies is growing exponentially. Depending on the starting number of DNA molecules, a certain number of PCR cycles is needed to reach a certain threshold (set by the investigator) of reporter dye mediated fluorescence intensity. The cycle number, at which this threshold is reached is designated as cycle threshold number. The log of the number of DNA molecules correlates linearly with the cycle threshold number, therefore giving a straight line if plotted in a graph (Figure 3.19.).

For our purposes, a standard curve was created by running several real-time PCR measurements in the presence of a known number of DNA molecules coding for M.Hae III-6xHis tag. The probe used for these experiments aligns on the gene of M.Hae III DNA-methyltransferase. The log of the number of DNA molecules was then plotted against the cycle threshold number (Figure 3.19.), resulting in a straight line with a good correlation coefficient.

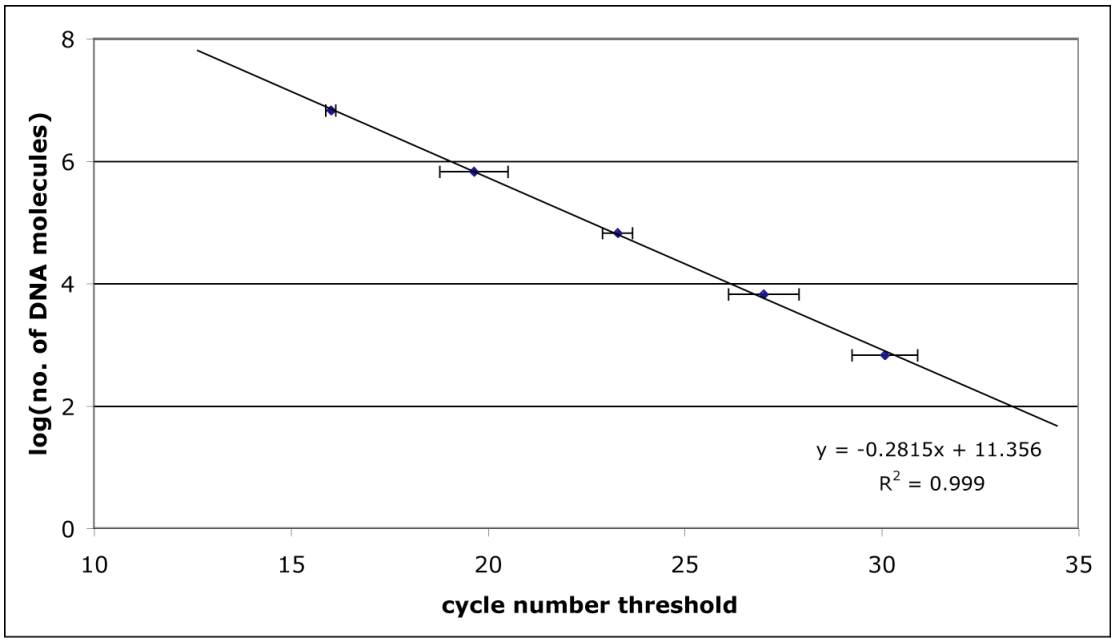

Figure 3.19. Standard curve with TaqMan® probe specific for M.Hae III gene. Between 10^7 and 10^3 DNA molecules coding for M.Hae III were used for real-time PCR measurements. Cycle number thresholds obtained by the measurements were plotted against the log(number of DNA molecules), and a straight line was fitted to the data points

by linear regression. Each sample was measured in triplicate. Standard errors are indicated by black lines.

If a sample containing an unknown number of DNA molecules is measured by real-time PCR, the cycle number threshold determined experimentally can be inserted into the equation of the standard curve (*x*, Figure 3.19.), which allows the calculation of the number of DNA molecules present in said sample.

Affinity selection experiments were performed essentially as described above (Figure 3.17.). In a first set of experiments (Table 3.1., upper half), two emulsions were prepared, one of which contained DNA coding for M.Hae III-6xHis and the other DNA encoding M.Hae III-Flag. After protein expression and extraction of the water phases, affinity selections using magnetic streptavidin beads coated with biotinylated anti-Flag antibody were performed. In Table I, the number of DNA molecules used as input for the selections and the corresponding outputs are given. The data show that DNA-M.Hae III-Flag adducts were selected 1000 times more efficiently than an irrelevant DNA–protein adduct (in this case M.Hae III-6xHis).

In a second set of experiments, various M.Hae III constructs were separately expressed in emulsion and, after extraction from the emulsion, affinity selections were performed using magnetic streptavidin beads coated with specific or irrelevant antibody. If the streptavidin beads were coated with specific antibody/peptide, DNA molecules were recovered much more efficiently than compared to the control selection, in which an irrelevant antibody was immobilised on the beads (Table 3.1, lower half). For the capture of DNA-M.Hae III-ED-B fusions, biotinylated anti-ED-B antibody called L19 small immunoprotein (Li *et al.*, 1997) was used. DNA-M.Hae III-CaM adducts were recovered by coating the beads with biotinylated calmodulin-binding peptide (Montigiani *et al.*, 1996).

Table 3.1. Affinity selections with different M.Hae III fusion constructs

M.Hae III construct	input	capture agent	output [a] (±stdev [b])	ratio [c]
M.Hae III-Flag tag	$6x10^{10}$	anti-Flag ab	$2.3±0.2x10^{6}$	
M.Hae III-6xHis	$6x10^{10}$		$7.6±2.0x10^{2}$	**3026**
M.Hae III-Flag tag	$6x10^{10}$	anti-Flag ab	$4.0±0.5x10^{5}$	
M.Hae III-6xHis	$6x10^{10}$		$5.0±0.8x10^{2}$	**800**
M.Hae III-Flag tag	$6x10^{10}$	anti-Flag ab	$1.7±0.1x10^{6}$	
M.Hae III-6xHis	$6x10^{10}$		$5.5±1.4x10^{2}$	**3091**
M.Hae III-Flag tag	$6x10^{10}$	anti-Flag ab	$1.9±0.2x10^{6}$	
M.Hae III-6xHis	$6x10^{10}$		$5.2±1.6x10^{2}$	**3654**
M.Hae III-6xHis	$3x10^{10}$	anti-His ab	$4.4±0.3x10^{5}$	
	$3x10^{10}$	anti-Flag ab	$7.9±3.6x10^{2}$	**557**
M.Hae III-6xHis	$3x10^{10}$	anti-His ab	$2.0±0.1x10^{6}$	
	$3x10^{10}$	anti-Flag ab	$2.9±0.9x10^{2}$	**6897**
M.Hae III-EDB	$5x10^{10}$	anti-EDB ab (L19)	$1.3±0.1x10^{6}$	
	$5x10^{10}$	anti-Flag ab	$4.0±0.2x10^{2}$	**3250**
M.Hae III-CaM	$5x10^{10}$	CaM-pep	$2.9±0.2x10^{5}$	
	$5x10^{10}$	anti-Flag ab	$1.7±0.6x10^{2}$	**1076**
M.Hae III-CaM	$5x10^{10}$	CaM-pep	$7.8±1.1x10^{5}$	
	$5x10^{10}$	anti-Flag	$6.9±1.4x10^{2}$	**1130**

[a]Output values were measured by real-time PCR using 0.1% of the beads after washing
[b]Standard deviation: all measurements were performed in triplicate
[c]Ratio of the two corresponding output values

The water phase extracted from the emulsion had to be dried well to remove residual traces of diethyl ether, which was used for the extraction process. If diethyl ether was still present in the samples as streptavidin beads were added, unspecific binding of DNA to the beads led to high background levels.

For real-time PCR measurements, only 0.1 μl of the 100 μl beads obtained after selection was taken for analysis, because fluorescence was quenched in the presence of too many beads (data not shown). Therefore, the values given in Table 3.1. in the column *"output"* must be multiplied by 10^3 to get the real absolute number of DNA molecules recovered from affinity selection.

The results shown in Table 3.1. illustrate the efficiency and reproducibility of the methodology in these model affinity selection experiments.

3.5.3. Capture efficiency of DNA-M.Hae III-6xHis tag fusions

The amount of DNA used as input for the experiments shown in Table 3.1. varied between 3-6x10^{10} molecules. From the information given in Table 3.1., the efficiency of recovery can be readily calculated (=*input/output*) (Table 3.2.).

Table 3.2. Recovery efficiency of DNA-M.Hae III fusions in capture experiments

M.Hae construct/capture agent	**input**	**output**	**[a]recovery (in%)**
-Flag/anti-Flag antibody	6x10^{10}	2.3x10^{9}	3.8
-Flag/anti-Flag antibody	6x10^{10}	4.0x10^{8}	0.7
-Flag/anti-Flag antibody	6x10^{10}	1.7x10^{9}	2.8
-Flag/anti-Flag antibody	6x10^{10}	1.9x10^{6}	3.2
-6xHis/anti-His antibody	3x10^{10}	4.4x10^{8}	1.5
-6xHis/anti-His antibody	3x10^{10}	2.0x10^{9}	6.7
-EDB/L19 antibody	5x10^{10}	1.3x10^{9}	2.6
-CaM/Cam-binding peptide	5x10^{10}	2.9x10^{8}	0.6
-CaM/CaM-binding peptide	5x10^{10}	7.8x10^{8}	1.6

[a]Recovery calculated from data given in Table 3.1.

The percentages of DNA molecules recovered by affinity capture ranged between 0.6% and 6.7% (Table 3.2.). In order to investigate the efficiency of recovery more systematically, DNA molecules coding for M.Hae III-6xHis tag protein were used for affinity selection experiments. About 5x10^{10}, 5x10^{9} and 5x10^{8} DNA molecules were emulsified together with *in vitro* transcription-translation mixture and cofactor SAM for expression of protein and formation of DNA-M.Hae III-6xHis fusions. Affinity selection was performed with streptavidin beads coated with anti-His or anti-Flag antibody (as negative control), and the number of bound DNA molecules was determined by real-time PCR in triplicate after washing. The number of DNA molecules taken as input for the selection was determined as well. Each experiment was performed twice, each time on a different day. The results are summarised in Table 3.3.

Table 3.3. Recovery efficiency of DNA-M.Hae III-6xHis fusions in dependence of the number of input DNA molecules.

input	output α-His ab	output α-Flag ab	recovery (α-His ab)
$5.0x10^{10}$	$1.5x10^{8}$	$1.6x10^{5}$	0.3%
$6.5x10^{10}$	$2.2x10^{8}$	$4.8x10^{5}$	0.3%
$5.0x10^{9}$	$9.0x10^{7}$	$2.6x10^{4}$	1.8%
$6.5x10^{9}$	$5.1x10^{7}$	$3.5x10^{4}$	0.8%
$5.0x10^{8}$	$1.6x10^{7}$	n.d.	3.2%
$6.5x10^{8}$	$1.6x10^{7}$	n.d.	2.5%

The relative amount of recovered DNA molecules (expressed in percentage of input) clearly increased with decreasing input of DNA, thus indicating that the number of binding sites on the beads might be the limiting factor for the capture of DNA-protein fusions. The reproducibility of the experiments was remarkable since even the absolute number of recovered DNA molecules was almost equal in experiments with the same input.

However, with an input of about $5x10^{10}$, the efficiency of recovery was worse by almost a factor ten compared to the data presented in Table 3.1., which might be explained by an impaired binding activity of the anti-His antibody. Background recovery with irrelevant anti-Flag antibody (negative control) was similar to the outcome of earlier experiments shown in Table 3.1.

The binding capacity of streptavidin beads (Dynabeads) as communicated by the manufacturer is about 10 μg of biotinylated antibody per mg of beads (concentration of the beads: 10 mg/ml). Since we have used 50 μl (= 0.5 mg) of beads per sample for our selection experiments, we should be able to immobilise about 5 μg antibody. Assuming a molecular weight of 150 kDa for an antibody ($1.5x10^{11}$ μg/mol), 5 μg of antibody correspond to $3.3x10^{-11}$ mol, which equals $2x10^{12}$ antibody molecules. If $5x10^{10}$ DNA molecules are used as input for protein expression, about $5x10^{12}$ M.Hae III-6xHis tag proteins will be produced (see section 3.3.4.: $2x10^{11}$ DNA molecules produced an M.Hae III-6xHis concentration of about 1500 nM, which corresponds to $4.5x10^{13}$ molecules in 50 μl of transcription-translation mixture). In conclusion, by using $5x10^{10}$ DNA molecules as input, we came probably close to the binding capacity of the streptavidin beads in the volume chosen, but did not surpass it significantly.

3.6. Model selections with DNA mixtures

It has been shown in chapter 3.5. that the affinity selection of DNA-M.Hae III fusions is efficient and reliable for a set of polypeptides fused to the DNA-methyltransferase. However, only one species of DNA molecules was expressed per emulsion, which does not reflect the real situation when working with protein libraries. For the isolation of protein candidates from a library of mutants, a mixture of DNA coding for a repertoire of different phenotypes must be used for selection experiments.

We tried to mimic this situation by preparing a mixture of two different DNA templates, of which one encoded M.Hae III-CaM-6xHis tag and the other M.Hae III-EDB-6xHis tag protein. We chose these two small globular proteins for our experiments because the covalent DNA display methodology will be preferentially used for the selection of globular proteins of desired binding specificities. In the ideal case, such proteins are devoid of disulfide bonds, are stably folded and are well expressed in bacteria. Calmodulin (Neri *et al.*, 1995; Montigiani *et al.*, 1996) and a number of fibronectin domains (Zardi *et al.*, 1987; Xu *et al.*, 2002) fulfil these requirements.

Unfortunately, if DNA templates coding for M.Hae III-CaM or M.Hae-EDB were mixed in a stoichiometric ratio of 1:1, affinity selection with calmodulin-binding peptide did not result in an enrichment of DNA molecules coding for M.Hae III-CaM fusion proteins. This "cross-talking" activity, which led to the incorrect assignment of genotype to phenotype, could have the following reasons: (*i*) several DNA molecules (instead of only one) could be present per compartment, (*ii*) protein expression could have started before the water compartments reached their final size, (*iii*) fusion of water compartments during expression of the proteins, and (*iv*) exchange of non-covalently bound proteins between different DNA molecules after extraction of the water phase from the emulsion.

3.6.1. Reasons for the incorrect assignment of protein to DNA

In order to find out where in the selection cycle the unwanted exchange of proteins and/or DNA molecules can take place, we performed various experiments. For the analysis of these experiments, TaqMan probes and primers specific for the coding sequence of either calmodulin or EDB were designed and

standard curves were determined for each coding sequence (Figure 3.20 and 3.21).

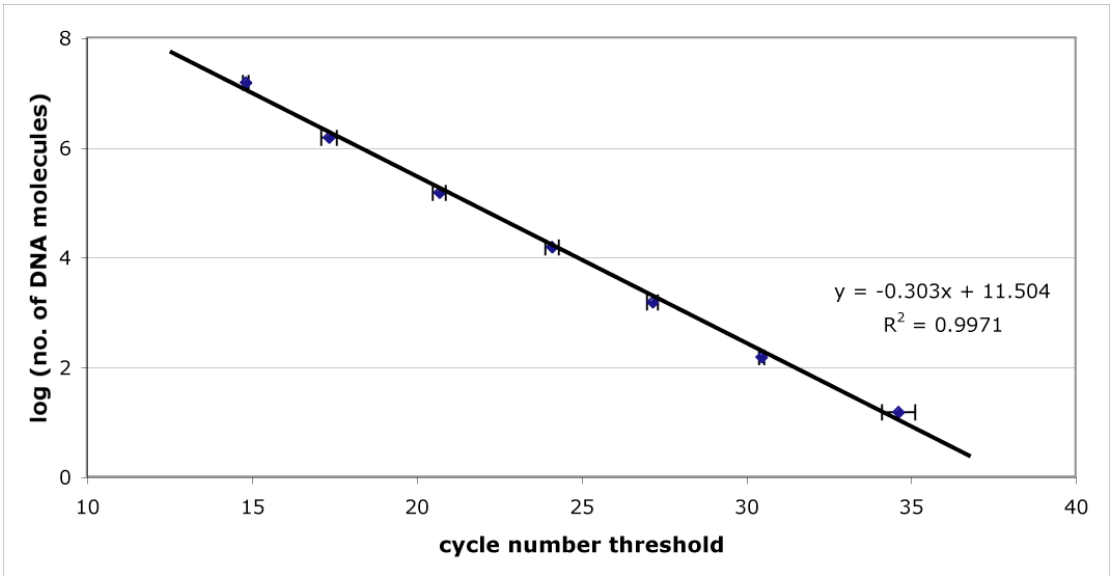

Figure 3.20. Standard curve with TaqMan® probe specific for calmodulin (CaM) gene. Between 10^7 and 10^1 DNA molecules coding for CaM were used for real-time PCR measurements. Cycle number thresholds obtained by the measurements were plotted against the log(number of DNA molecules), and a straight line was fitted to the data points by linear regression. Each sample was measured in triplicate. Standard errors are indicated by black lines.

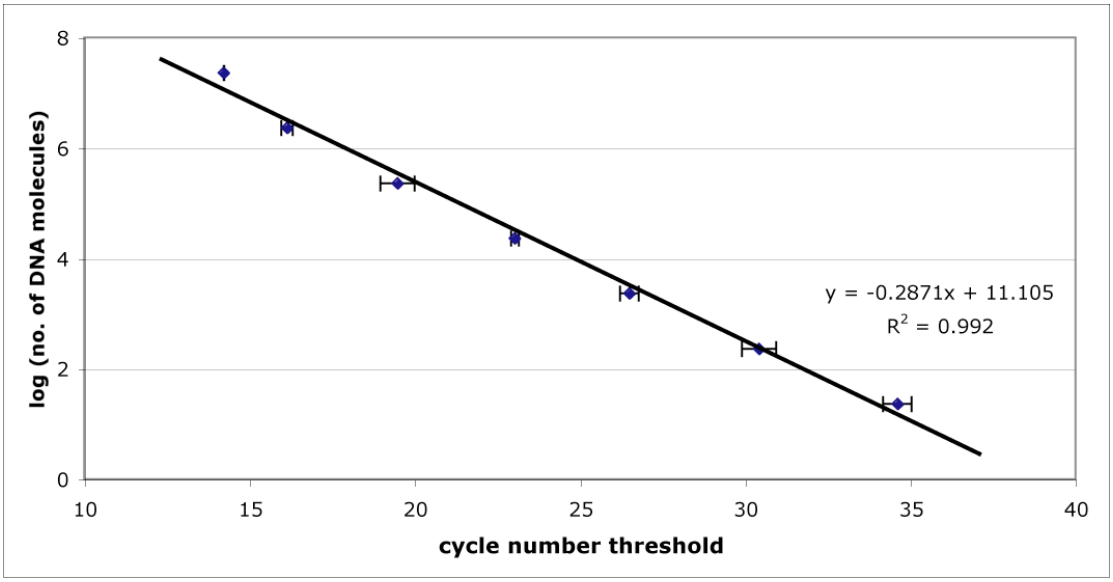

Figure 3.21. Standard curve with TaqMan® probe specific for the extra-domain B of fibronectin (EDB) gene. Between 10^7 and 10^1 DNA molecules coding for EDB were used for real-time PCR measurements. Cycle number thresholds obtained by the measurements

were plotted against the log(number of DNA molecules), and a straight line was fitted to the data points by linear regression. Each sample was measured in triplicate. Standard errors are indicated by black lines.

Standard curves were fitted to the data points obtained by real-time PCR, resulting in linear curves with high correlation coefficients.

For the subsequent selection experiments, DNA molecules coding for either M.Hae III-CaM or M.Hae III-EDB fusion proteins were incubated separately in transcription-translation mixture for protein expression in solution (not emulsified). After expression, TBS containing 1 mM $CaCl_2$ (TBSC) was added to both samples because CaM can only bind to its target peptide in the presence of Ca^{2+} ions (Montigiani *et al.*, 1996). Short double-stranded biotinylated DNA fragments (5 μM) were added to the samples to avoid unspecific binding of DNA to streptavidin beads. Using these two solutions, six different samples were prepared and used in affinity selections with CaM-binding peptide or anti-Flag antibody (neg. control) on streptavidin beads.

The first sample only contained M.Hae III-CaM, the second only M.Hae III-EDB, the third a mixture of M.Hae III-EDB/-CaM solutions with a ratio of 100/1, and the fourth a mixture of M.Hae III-EDB/-CaM solutions with a ratio of 10'000/1. The fifth and the sixth sample were prepared as samples three and four, respectively, with the exception that a short double-stranded DNA fragment (81 bp) containing five times the methlyation target sequence 5'-GGCC-3' (5xGGCC DNA) was added (final concentration of 1 μM) to capture free M.Hae III fusion proteins. The hypothesis was, that 5xGGCC DNA could serve as a sink for redundant M.Hae III proteins, and thus, hamper the exchange of DNA-methyltransferase fusion proteins between different DNA templates.

The six samples were then subjected to affinity selection with streptavidin beads coated either with CaM-binding peptide or anti-Flag antibody. After washing, the beads of samples one (only CaM) and two (only EDB) (Table 3.4.) were directly used for real-time PCR. From mixed samples, a small amount of the beads was taken for PCR amplification of the isolated DNA molecules. The PCR products were used for real-time PCR measurements in order to quantify the absolute number of DNA molecules coding either for M.Hae III-EDB or –CaM in the output of the selection experiment (Table 3.4.).

Table 3.4. Selection experiments with DNA mixtures prepared after *in vitro* expression of the proteins in solution.

	[c]only CaM	[c]only EDB	[d]E/C=10^2	[d]E/C=10^4	[d,e]E/C=10^2	[d,e]E/C=10^4
α-Flag[a]	$8x10^3$	$8x10^4$	24	[g]n.d.	n.d.	n.d.
CaM-pep[b]	$9x10^6$	$9x10^3$	2	11	0.25	13
Enrichment[f]	-	-	*50*	*909*	*400*	*769*

[a]Anti-Flag antibody used for affinity selection
[b]Calmodulin-binding peptide used for affinity selection,
[c]Samples containing only DNA coding for M.Hae III-CaM or EDB.
[d]Mixtures of M.Hae III-EDB/-CaM solutions in ratios as indicated.
[e]With 5xGGCC DNA as sink for redundant M.Hae III proteins.
[f]Putative enrichment factor based on the ratio how the starting mixtures were prepared.
[g]Not detected: no M.Hae III-CaM encoding DNA molecules detectable.

The samples with only one species of DNA molecules served as positive and negative control experiments. Affinity selections using anti-Flag antibody were done to investigate whether the ratio of the two different DNA molecules could change independently of the bait molecule used during the selection process, e.g., due to uneven amplification of the two genes during PCR.

In the output of the anti-Flag antibody-selection of the sample with a starting mixture of genes M.Hae III-EDB/-CaM around 100/1, 24 times more DNA molecules encoding M.Hae III-EDB than DNA molecules coding for M.Hae III-CaM were found (if 5xGGCC DNA was absent). In the other outputs fom anti-Flag selections with starting mixtures of DNA molecules, a ratio of EDB genes to CaM genes could not be calculated because a signal for CaM genes could not detected in real-time PCR. Probably, the DNA molecules encoding M.Hae III-CaM were diluted out during washing of the beads. In theory, the ratio of EBD/CaM genes after selection with anti-Flag antibody (sample E/C = 10^2, without 5xGGCC DNA) should be 100 and not 24. Reasons for this lower observed selection efficieny might be the uneven amplification of M.Hae III-EDB and M.Hae III-CaM templates during PCR or inaccurate handling of the samples.

If CaM-binding peptide was used as bait molecule for selections, EDB encoding genes surpassed those coding for M.Hae III-CaM by only a factor 2-13. In the sample, where M.Hae III-EDB and –CaM encoding molecules had been initially mixed in a ratio of 100/1 and 5xGGCC DNA was present during selection, even 4 times more CaM genes than EDB encoding genes were found (Table 3.4.). Therefore, the selection of DNA molecules coding for M.Hae III-CaM proteins was successful in this experimental set-up. In the samples with very few CaM genes

(M.Hae III-EDB/-CaM = 10^4), the enrichment factors were higher than in those with 100 times more M.Hae III-CaM DNA molecules, because there were less M.Hae III-CaM fusion proteins which could bind to DNA templates encoding M.Hae III-EDB. Therefore, fewer false positives were selected. The addition of 5xGGCC DNA capturing free DNA-methyltransferase fusion proteins helped to decrease the number of false positives, leading to higher enrichment factors.

In summary, the results of these experiments imply that, if protein expression and cross-linking can be performed separately (here in two different tubes), the negative influence of free M.Hae III proteins is limited or even not measurable if the abundance of DNA molecules encoding binding proteins is low enough.

Additional experiments with mixtures of DNA molecules coding either for M.Hae III-CaM or –EDB were carried out.

In the first experiment, DNA encoding M.Hae III-CaM or –EDB was expressed separately in solution. After expression, a small amount of the sample with M.Hae III-CaM encoding DNA was added to the solution with M.Hae III-EDB genes, so that the ratio of EDB/CaM DNA molecules should be 1000.

In the second experiment, two separate emulsions were prepared containing DNA templates encoding either M.Hae III-EDB or –CaM. After expression, the water phases were extracted and mixed so that the ratio of M.Hae III-EDB/-CaM DNA molecules should be 1000.

In the third experiment, two separate emulsions were prepared containing DNA templates encoding either M.Hae III-EDB or –CaM proteins. Just after emulsification of the DNA molecules and transcription-translation mix, 1 µl of the emulsion containing the M.Hae III-CaM genes was added to the emulsion comprising the M.Hae III-EDB genes (1 ml), thus creating a ratio of M.Hae III-EDB/-CaM encoding DNA molecules of 1000.

In the fourth – and last – experiment, DNA molecules were directly mixed in a ratio of M.Hae III-EDB/-CaM = 1000. This DNA mixture was then emulsified and used for selection.

Therefore, mixtures of DNA molecules were prepared in four different ways to obtain ratios of M.Hae III-EDB/M.Hae III-CaM = 1000: (*i*) a mixture of solutions,

(*ii*) a mixture of extracts, (*iii*) a mixture of emulsions, and (*iv*) a mixture of DNA molecules. In all four cases, selections were performed with streptavidin beads coated with biotinylated CaM-binding peptides in order to test if it was possible to enrich DNA molecules coding for M.Hae III-CaM fusion proteins. Affinity selections with anti-Flag antibodies were performed as negative control. The DNA mixtures, obtained from PCR amplification of recovered DNA molecules on the magnetic beads, were used for real-time PCR analysis.

This set of experiments was carried out twice, and the results are shown in Table 3.5. In all selections with the anti-Flag antibody, except the experiment where emulsions were mixed, the ratio of M.Hae III-EDB/CaM could not be determined because CaM DNA sequences were not detected. The M.Hae III-EDB/CaM ratios obtained from the mix of emulsions-experiments (*iii*) corresponded about to the expected ratio of 1000 (429 and 364, respectively, Table 3.5.).

If CaM-binding peptide was used as bait in the affinity capture, there was a clear enrichment of CaM DNA sequences in all experiments. However, the selections for M.Hae III-CaM encoding DNA molecules were not equally efficient in every experiment.

Table 3.5. Ratio of M.Hae III-EDB/-CaM encoding DNA molecules after selection starting from DNA mixtures prepared in four different ways.

	mixture of solutions[c]	**mixture of extracts[d]**	**mixture of emulsions[e]**	**mixture of DNA molecules[f]**
α-Flag[a]	[g]n.d.	n.d.	429	n.d.
α-Flag	n.d.	-	364	n.d.
CaM-pep[b]	16	15	6	49
CaM-pep	1	-	1	88

[a]Affinity capture with anti-Flag antibody.

[b]Affinity capture with calmodulin(CaM)-binding peptides.

[c]Separate expression in solution, then mixed M.Hae III-EDB/CaM = 1000.

[d]Separate expression in emulsion; extracts mixed that M.Hae III-EDB/CaM = 1000.

[e]Separate emulsification, then mixed that M.Hae III-EDB/CaM = 1000.

[f]Mixture of DNA molecules; ratio of M.Hae III-EDB/CaM = 1000.

[g]Not detected: no M.Hae III-CaM encoding DNA molecules detectable.

The input ratio of the two DNA species was also measured in the starting mixture of DNA molecules used for the experiments shown in the last column of Table

3.5. In addition, this DNA mixture was diluted and amplified by PCR, after which the ratio of M.Hae III-EDB/-CaM encoding DNA molecules was determined once more. This was done in order to find out, whether the two types of DNA are amplified with the same efficiency during PCR. The ratio of M.Hae III-EDB/-CaM DNA molecules in the starting DNA mixture was 4619 and 4444, respectively. After dilution and amplification by PCR, the ratios were determined to be 1919 and 1909, respectively, implying that M.Hae III-CaM molecules were amplified in PCR slightly more efficiently than DNA molecules encoding M.Hae III-EDB fusion proteins.

The efficiency with which DNA templates coding for M.Hae III-CaM were selected seemed to be worst in the experimental set-up where the DNA molecules were mixed before being emulsified. However, the data are not comparable in every respect because the input ratio of DNA molecules was not measured in the mixtures of solutions, extracts and emulsions. Therefore, direct comparison of enrichment factors is difficult. Our experimental data indicated that there is not much difference between the experiments where proteins were first separately expressed in solution and then mixed, compared to those experiments where DNA molecules were compartmentalised, but mixed just after emulsification. This observations indicates, that the compartments of the water-in-oil emulsions were stable and not exchanging their contents during expression.

The weaker performance of the selection experiments in which DNA molecules were mixed from the beginning might be due to the following two reasons: (*i*) the water compartments contained several DNA molecules, and (*ii*) protein expression could have started before the water compartments reached their final size, thus exchanging proteins and/or mRNA molecules.

If one or both of these hypotheses apply to our observations, the selection process should become more efficient if fewer DNA molecules are used as input. Therefore, two separate emulsions containing DNA molecules encoding M.Hae III-EDB or –CaM were prepared and mixed just after completed emulsification such as the ratio of M.Hae III-EDB/-CaM molecules should be 1000. Additionally, a DNA mixture (M.Hae III-EDB/-CaM = 1000) was prepared and directly used for expression in emulsion. Both experiments were done with 10^{10} (as in the experiments above) and 10^{9} DNA molecules as input for the selections. Again, affinity capture of the DNA-protein fusions was performed with CaM-binding peptide and anti-Flag antibody (as negative control). Recovered DNA molecules

were amplified by PCR and the ratio of the two different DNA molecules in the resulting solutions was determined by real-time PCR. The starting DNA mixture again was diluted and amplified by PCR. The ratio of DNA molecules encoding M.Hae III-EDB/-CaM fusion proteins in the input-mixture was measured before and after PCR.

Table 3.6. Selections with different amounts of mixed DNA molecules.

experiment	input	input ratio[a]	input PCR[b]	α-Flag[c]	CaM-peptide[c]
mixture of em.	10^{10}	-	-	2870	15
mixture of em.	10^{9}	-	-	n.d.[d]	15
mixture of DNA	10^{10}	4167	2727	980	100
mixture of DNA	10^{9}	4200	2047	n.d.[d]	28

[a]Input ratio of M.Hae III-EDB/-CaM DNA molecules.

[b]Ratio after PCR of input DNA mixture.

[c]Ratio after selection agent with capture agent (anti-Flag antibody / CaM-binding peptide).

[d]Not detected: no M.Hae III-CaM encoding DNA molecules detectable.

Selections worked better if only 10^9 instead of 10^{10} DNA molecules were used as input. Again, DNA molecules coding for M.Hae III-CaM fusion proteins were amplified more efficiently in PCR than those encoding M.Hae III-EDB.

3.6.2 Model selections using an improved protocol

In order to test the robustness of the methodology, a number of model selections using different amounts of DNA as input were performed. As above, mixtures of DNA templates coding either for M.Hae III-CaM or M.Hae III-EDB were prepared in such a way that the ratio of M.Hae III-EDB/-CaM DNA molecules was ~1000 and that ~10^9–10^{10} DNA molecules were used per selection experiment. Biotinylated CaM-binding peptide was used to coat the magnetic streptavidin beads in order to enrich DNA coding for M.Hae III-CaM from these DNA mixtures, whereas biotinylated anti-Flag antibody served as a negative control because neither of the two proteins contained a Flag-tag, and therefore, the ratio of M.Hae III-EDB/-CaM DNA molecules should not change as compared with the input ratio. DNA molecules recovered from the selection process were amplified by PCR. Selection experiments were analysed by real-time PCR using TaqMan probes specific for either M.Hae III-EDB or -CaM DNA. The ratio of M.Hae III-EDB/-CaM

DNA molecules was determined before selection (input) and after selection with CaM-binding peptide or anti-Flag antibody (Table 3.7.). In the last column of Table 3.7., enrichment factors were calculated by dividing the value of the input ratio by the value of the output ratio (using CaM-binding peptide as capture agent). The DNA mixtures, obtained from PCR amplification of recovered DNA molecules on the magnetic beads, were diluted by a factor $1:10^4$–10^5 for real-time PCR analysis. Using 10^{10} DNA molecules as input for the selection procedure, enrichment factors ranged between 42 and 94 (Table 3.7., selection experiments 1–7). If only 10^9 DNA molecules were used, enrichment factors increased to 81–153 (Table 3.7., selection experiments 8–10). In some control selections with the irrelevant anti-Flag antibody, DNA coding for M.Hae III-CaM was not detectable by real-time PCR, having been diluted out during the washing steps.

Table 3.7. Overview: model selections with mixtures of M.Hae III-EDB and –CaM DNA molecules

exp.	*no. of DNA molecules* [a]	*input EDB*	*input CaM*	*input ratio EBD/CaM* [b]	*output EDB anti-Flag*	*output CaM anti-Flag*	*ratio anti-Flag* [c]	*output EDB CaM-pep*	*output CaM CaM-pep*	*ratio CaM-pep* [d]	*enrichment factor* [e]
1	10^{10}	$9.7\pm0.7\times10^{5}$	$2.1\pm0.2\times10^{2}$	*4619*	$9.7\pm1.1\times10^{6}$	n.d. [f]	*n.d.*	$3.6\pm0.2\times10^{6}$	$7.4\pm1.6\times10^{4}$	*49*	**94**
2	10^{10}	$1.2\pm0.2\times10^{6}$	$2.7\pm0.4\times10^{2}$	*4444*	$3.4\pm0.3\times10^{6}$	n.d.	*n.d.*	$2.3\pm0.1\times10^{6}$	$2.6\pm0.3\times10^{4}$	*88*	**50**
3	10^{10}	$1.0\pm0.2\times10^{6}$	$2.4\pm0.2\times10^{2}$	*4167*	$4.9\pm0.8\times10^{5}$	$5.0\pm0.2\times10^{2}$	*980*	$5.5\pm0.8\times10^{5}$	$5.5\pm0.5\times10^{3}$	*100*	**42**
4	10^{10}	$2.5\pm0.3\times10^{6}$	$1.5\pm0.3\times10^{3}$	*1667*	$7.2\pm0.4\times10^{5}$	$1.9\pm0.6\times10^{2}$	*3789*	$3.5\pm0.2\times10^{5}$	$9.1\pm1.5\times10^{3}$	*38*	**43**
5	10^{10}	$2.3\pm0.3\times10^{6}$	$1.3\pm0.4\times10^{3}$	*1769*	$1.7\pm0.1\times10^{6}$	$6.7\pm2.1\times10^{2}$	*2537*	$7.0\pm0.1\times10^{5}$	$1.8\pm0.5\times10^{4}$	*39*	**45**
6	10^{10}	$1.5\pm0.2\times10^{6}$	$6.8\pm1.4\times10^{2}$	*2206*	$6.1\pm1.2\times10^{5}$	$7.4\pm0.9\times10^{2}$	*824*	$8.4\pm0.4\times10^{5}$	$1.9\pm0.2\times10^{4}$	*44*	**50**
7	10^{10}	$1.5\pm0.2\times10^{6}$	$6.8\pm1.4\times10^{2}$	*2206*	$4.5\pm0.9\times10^{5}$	n.d.	*n.d.*	$1.3\pm0.5\times10^{6}$	$2.8\pm0.5\times10^{4}$	*46*	**48**
8	10^{9}	$8.4\pm0.6\times10^{5}$	$2.0\pm0.1\times10^{2}$	*4200*	$1.5\pm0.1\times10^{6}$	n.d.	*n.d.*	$1.4\pm0.2\times10^{6}$	$5.1\pm0.7\times10^{4}$	*27*	**153**
9	10^{9}	$9.5\pm1.6\times10^{5}$	$5.1\pm1.1\times10^{2}$	*1863*	$7.9\pm1.1\times10^{4}$	n.d.	*n.d.*	$1.5\pm0.1\times10^{5}$	$6.5\pm0.1\times10^{3}$	*23*	**81**
10	10^{9}	$1.3\pm0.4\times10^{6}$	$4.2\pm0.3\times10^{2}$	*3095*	$6.6\pm0.9\times10^{4}$	n.d.	*n.d.*	$2.4\pm0.3\times10^{5}$	$9.2\pm1.9\times10^{3}$	*26*	**119**

[a]Total number of DNA molecules used for selections.
[b]Ratio of M.Hae III-EDB/M.Hae III-CaM DNA molecules used as input; PCR with diluted input mixtures decreased input ratio by a factor 2-3.
[c]Ratio of M.Hae III-EDB/M.Hae III-CaM DNA molecules after selection with anti-Flag antibody as capture reagent.
[d]Ratio of M.Hae III-EDB/M.Hae III-CaM DNA molecules after selection with CaM-binding peptide as capture reagent.
[e]Enrichment = [input ratio]/[output ratio (obtained with CaM-binding peptide)].
[f]n.d.: no DNA molecules were detectable in these samples.

3.7. Amplification strategies

An important part of an *in vitro* selection methodology is the amplification of the genetic information obtained from affinity purification. In practice, the amplification of DNA/RNA molecules by standard PCR using only one pair of primers can be difficult because, after three to five rounds of selection followed each by PCR, the total number of PCR cycles could easily exceed 100. This high number of PCR cycles can lead to a significant amount of unspecific side-products and non-functional DNA templates. Therefore, some amplification strategies have been proposed in the literature for the specific amplification of DNA *in vitro*.

In CIS display, the N-terminal region of the recovered DNA molecules was amplified by using a nested primer for each round (e.g. primer A in round 1, primer B in round 2, and so forth up to primer E). The PCR products were reattached to the C-terminal *RepA-CIS-ori* sequence by restriction ligation and further amplification to produce input DNA for the next round of selection (Odegrip *et al.*, 2004). In another study (Bernath *et al.*, 2005), the surviving DNA from each round of selection was cloned into a vector containing the regulatory sequences needed for *in vitro* expression. Linear DNA templates were then amplified from this vector with suitable primers for the next round of selection. In contrast, Yonezawa *et al.* only needed one pair of primers for the amplification of DNA over five rounds of selection (Yonezawa *et al.*, 2003), which shows that specific amplification of DNA in more than 125 cycles can be achieved.

In our experiments, DNA templates obtained after two rounds of selection and a total of 70 PCR cycles were not expressing any detectable amounts of proteins (data not shown). Therefore, an amplification strategy comprising nested PCR with two pairs of primers followed by assembly was devised.

3.7.1. Nested PCR and PCR assembly

DNA fragments isolated from affinity selection should first be amplified (20 PCR cycles) with a primer pair aligning at the extremities of the DNA template. This first PCR would be followed by a second amplification (20 PCR cycles) using a different primer pair aligning about 100 bp away from the DNA ends. By using PCR assembly with a third primer pair, sequences needed for efficient expression *in vitro* could be reattached (Figure 3.22.).

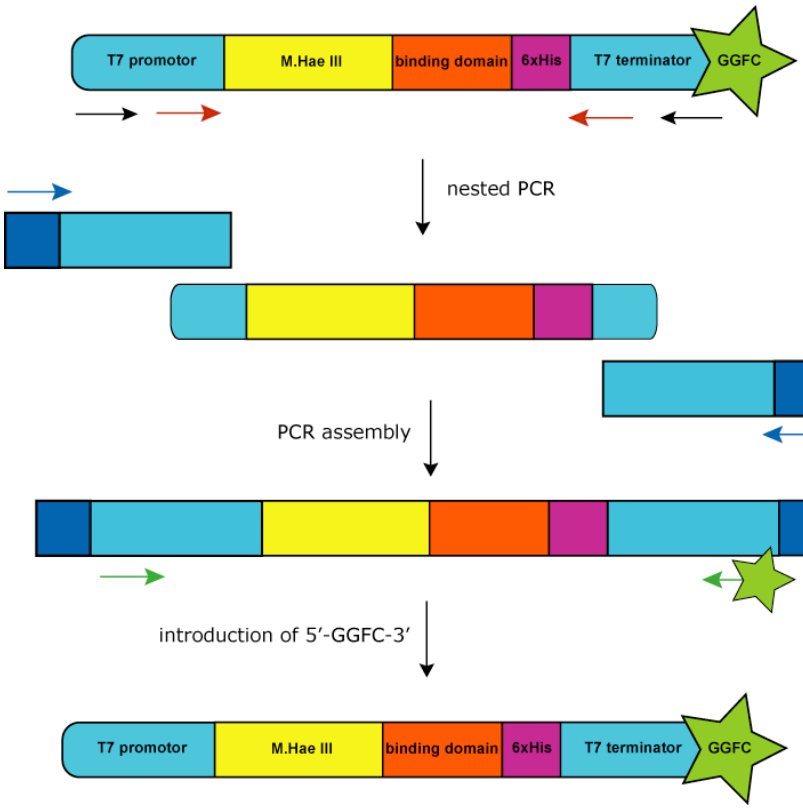

Figure 3.22. Strategy for the amplification of DNA molecules recovered from affinity selection by nested PCR. After a first PCR with primers aligning at the extremities of the DNA template (black arrows), a second PCR is performed with different primers (red arrows) to increase the specificity of the amplification. Full-length template is produced by PCR assembly (with primers indicated as blue arrows), which is followed by a last PCR step for the incorporation of 5-fluorodeoxycytidine (green arrows).

The most difficult step of this amplification strategy is probably the PCR assembly. A sufficient number of DNA molecules must be assembled into full-length templates to ensure that diversity is not lost during amplification.

DNA templates containing a stretch of nine fully randomised amino acids between the sequences coding for M.Hae III and 6xHis tag was used for one round of selection against anti-Flag antibody immobilised on streptavidin beads. After washing, the recovered DNA molecules were amplified by nested PCR and purified on an agarose gel. The small DNA fragments used for PCR assembly had been previously produced by PCR using plasmid DNA (designed for *in vitro* transcription from T7 promotor and translation by the manufacturer of transcription-translation mixture).

The middle DNA fragment and the two outer short fragments (3-fold molar excess) were mixed for PCR assembly using 7 cycles of PCR. External primers (Figure 3.22, blue arrows) were included to promote efficient amplification of full-length DNA molecules (Figure 3.23.).

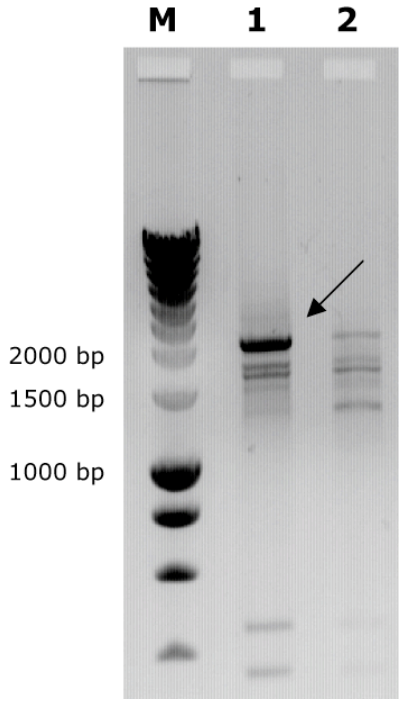

Figure 3.23. PCR assembly of the middle (long) with the two outer (short) DNA fragments. The outer DNA fragments (351 and 451 bp, respectively) were used in a 3-fold molar excess compared to the middle DNA fragment (1301 bp). In lane one (1), external primers were included in PCR assembly whereas in lane two (2) they were omitted. Full-length assembled DNA (2061 bp) (arrow) as well as only partially assembled DNA fragments (1630 bp and 1732 bp, respectively) were produced. In the sample without outer primers, the band representing full-length DNA is very weak. Note that in lane one (1), very few of the middle DNA fragments were not used for assembly. M: DNA marker.

Not all DNA molecules used as input for the PCR assembly were converted to full-length products. Some DNA fragments only assembled with one of the two outer fragments, yielding DNA molecules of 1630 bp and 1732 bp, respectively. If external primers were omitted, some of the middle DNA fragment was not employed in any assembly (Figure 3.23., 2).

For a next round of selection, full-length DNA molecules would be excised from the agarose gel, purified and used for a last PCR to introduce the mechanism based inhibitor target sequence 5'-GGFC-3'. Assembly only worked well, if the

middle DNA fragments were previously well purified on agarose gel (data not shown).

Restriction digestion followed by ligation represents an alternative approach to obtain full-length DNA templates after nested PCR (Figure 3.24.).

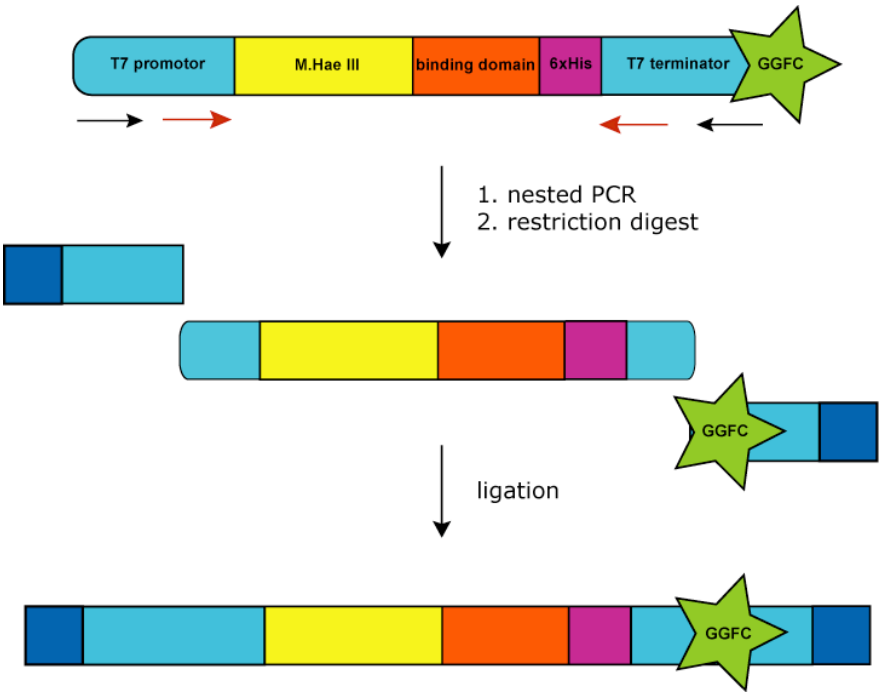

Figure 3.24. Alternative amplification strategy, where nested PCR with two primer pairs (black and red arrows) is followed by restriction digest and subsequent ligation.

At the moment, we have not yet proven that the proposed amplification strategies will allow the specific and efficient production of DNA over several rounds of selection and amplification. Further experiments will hopefully show, which of the possible strategies will be most suitable.

4. Discussion

In this thesis, we describe an *in vitro* selection system in which phenotype (protein) and genotype (DNA) are physically linked by a covalent bond. This physical linkage between a protein and its encoding DNA molecule was created by exploiting the ability of Hae III DNA-methyltransferase from *Haemophilus aegypticus* to form a covalent bond with DNA fragments containing the sequence 5'-GGFC-3' (F = 5-fluorodeoxycytidine) (Chen *et al*., 1991). DNA molecules coding for M.Hae III-CaM proteins were successfully and reproducibly selected in model selection experiments from a more than 1000-fold excess of DNA molecules coding for M.Hae III-EDB. However, only future experiments can show whether this methodology can be routinely used for the selection of binding proteins with desired biophysical properties.

4.1. Strengths and weaknesses of covalent DNA display

M.Hae III DNA-methyltransferase was covalently cross-linked to DNA containing the mechanism based inhibitor target sequence 5'-GGFC-3'. The reaction occurred with suitable kinetics (50% cross-linking in about 2 h) at very low concentrations of DNA (2 nM) and M.Hae III (38 nM), which corresponds to one DNA and 16 M.Hae III molecules within a compartment of 1 μm in diameter. However, the maximum yield of cross-linked DNA-protein fusions was limited to 50%, even after 4 h of incubation and at higher concentrations of both DNA and enzyme. At first, we thought that 5-fluorodeoxycytidine could be damaged during PCR used for the incorporation of the modified methylase target sequence 5'-GGFC-3' in DNA templates. However, cross-linking experiments with the 5-fluorodeoxycytidine comprising primer hybridised to a complementary oligonucleotide showed the same yield of DNA-protein adducts, indicating that not all of the oligonucleotides contained a functional cross-linking site.

M.Hae III DNA-methyltransferase not only bound to DNA molecules in a covalent manner, but was shown to form multivalent complexes with DNA as well. M.Hae III has a dissociation constant for fully methylated and nonspecific DNA of ~12.5 nM (Chen *et al.*, 1993), and thus, many molecules of M.Hae III can bind to a DNA template of about 1500 bp. After incubation of a 268 bp-DNA fragment (containing 4 GGCC and 1 GGFC site) with M.Hae III DNA-methyltransferase followed by detection of DNA-protein complexes on a native polyacrylamide gel,

DNA molecules associated with different numbers of M.Hae III proteins formed a ladder-like pattern, of which the highest step corresponded to DNA-$(M.Hae\ III)_7$ complexes. This result implies that as many as 35 M.Hae III proteins could bind to DNA molecules of ~1500 bp, used as templates for *in vitro* transcription-translation.

This multivalent binding of M.Hae III to DNA might be advantageous and disadvantageous for the selection methodology at the same time. Disadvantageous, because, after extraction of the water phase from the emulsion, free and non-covalently bound M.Hae III fusion proteins could bind many different DNA molecules, thus leading to the incorrect assignment of DNA to protein. However, it should be possible to minimise this cross-contamination by the addition of irrelevant DNA molecules in large excess.
Multivalent DNA display could be advantageous, because it might allow the isolation of weakly binding proteins due to avidity effects during affinity capture.

In affinity maturation experiments, where many clones are expected to bind the target with moderate affinity, the number of M.Hae III fusion proteins bound to the DNA molecules might be reduced in two different ways.
First, by using biotinylated DNA templates, multivalent DNA-protein complexes can be immobilised on magnetic streptavidin beads. The strength of the streptavidin-biotin interaction could allow stringent washes (e.g. high salt concentrations, denaturing agents) to remove M.Hae III proteins non-covalently bound to DNA. Monovalent DNA-protein fusions can be quantitatively eluted in nonionic aqueous solutions by short incubation (few seconds) at 70°C (Holmberg *et al.*, 2005). This monovalent population of DNA-protein fusions could then be used for selection experiments. In addition, such a procedure could yield binding proteins, which can either reversibly refold or are resistant to the harsh conditions used for the washing of the streptavidin beads.
Secondly, a methodology has been reported in a recent publication, which allows the delivery of metal ions to the water compartments of water-in-oil emulsions by using a nano-droplets delivery system (Bernath *et al.*, 2005). Such a system could be applied for the delivery of chemicals (e.g. antibiotics), which inhibit protein synthesis, but not the cross-linking activity of M.Hae III DNA-methyltransferase. Thus, the production of M.Hae III proteins could be stopped at any time point, and at the same moment, compartments could be preserved until cross-linking is expected to have succeeded.

One problem to keep in mind is that denaturing agents or conditions could lead to DNA-protein fusions comprising an unfolded M.Hae III protein, which could be sticky and thus harm the efficiency of the selection experiment by nonspecifically binding to the solid support or the selection target.

The *in vitro* expression of M.Hae III fused to two peptide tags and two small globular domains (EDB and calmodulin) was shown to be quite efficient. It should be noted that not only absolute expression levels are important, but also the integrity of the fusion protein is fundamental, because M.Hae III proteins lacking a fused binding domain could outcompete fusion proteins during cross-linking. Therefore, a linkage of phenotype to genotype would not occur.
In our expression studies, proteolytic fragments were not detectable on Western blots when using antibodies binding to C-terminal epitope tags of the fusion proteins. These results, however, do not prove that only partially translated M.Hae III fusion proteins without C-terminal epitope tags were not present in the transcription-translation mixture. Whether this fraction of partially translated M.Hae III proteins is large or not might be tested by expressing the methylase fusion proteins with an N-terminal epitope tag, such as 6xHis.

In our experiments, we observed the rapid degradation of DNA templates in *in vitro* transcription-translation mixture at 37°C, most probably by exonucleases. Fortunately, the stability of DNA molecules was dramatically improved when the temperature was lowered to 30°C because even relatively low amounts of DNA (12.5 ng) incubated for 80 min in 1:10 diluted transcription-translation mixture were detectable on polyacrylamide gels stained with SYBR green I. Despite the apparent integrity of the DNA molecules, we could not yet prove whether the extremities of the DNA templates were intact. In order to clarify this point, DNA coding for M.Hae III-6xHis tag proteins were produced by PCR, either with or without 5-fluorodeoxycytidine. After *in vitro* transcription and translation, affinity selections using nickel-coated magnetic beads and strongly denaturing conditions (6 M GuHCl and 8 M urea) were performed, and DNA molecules comprising the mechanism based inhibitor 5-fluorodeoxycytidine were recovered more efficiently than DNA molecules without the modified nucleotide. However, the recovery seemed to be rather low and the size of the fraction of DNA molecules with a covalently bound M.Hae III-6xHis tag could not be determined.
The efficiency of cross-linking in *in vitro* transcription-translation mixture might be investigated by applying the procedure described above for the isolation of monovalent DNA-protein fusions (Holmberg *et al.*, 2005). In this way, biotin-

DNA-M.Hae III fusions could be purified from *in vitro* transcription-translation mixture, and after elution, analysed on a gel (*in vitro* mixture cannot be applied undiluted to gel electrophoresis due to high concentrations of salts and proteins).

Affinity capture experiments worked efficiently and reliably both with nickel-coated beads and streptavidin beads as solid phase. In first experiments using streptavidin beads, the background recovery of DNA encoding non-binding proteins was very high. Interestingly, this background could be lowered by a factor ~1000 by adding short, double-stranded biotinylated DNA molecules. This observation might be explained by the following hypothesis: biotinylated DNA fragments probably bound to free binding sites on the streptavidin beads which had been previously coated with biotinylated capture agent, thereby bringing negative charges to the surface of the beads. This negative charge might repel other DNA molecules binding unspecifically to the beads.

Model selection experiments with DNA mixtures containing two types of templates were performed. As shown in Table 3.1., the efficiency of DNA recovery from affinity selections varied by more than three orders of magnitude, depending on the binding specificity of the polypeptide used as capture agent. These results imply that DNA molecules coding for a protein with appropriate binding characteristics could be enriched from a mixture of different DNA molecules by a factor of ~1000. However, enrichment factors obtained from model selection experiments starting from mixtures of two different DNA templates were lower than expected by a factor of 10–20 (Table 3.7.). These findings may be explained by different reasons.

First, after extraction of the water phase from the emulsion, free M.Hae III fusion proteins, which are not linked to their encoding DNA molecule, might bind to other DNA molecules in a non-covalent manner and, therefore, lead to the selection of DNA molecules coding for non-binding proteins. In order to abolish cross-contamination, short double-stranded DNA fragments (81 bp) containing five times the methylation target sequence 5'-GGCC-3' (5xGGCC DNA) were added to capture free M.Hae III fusion proteins. The hypothesis was, that 5xGGCC DNA could serve as a sink for redundant M.Hae III proteins, and thus, hamper the exchange of DNA-methyltransferase fusion proteins between different DNA templates (M.Hae III-EDB and –CaM).
If the ratio of DNA templates encoding non-binding (M.Hae III-EDB) to binding (M.Hae III-CaM) proteins was 100, the addition of 5xGGCC DNA significantly

increased the enrichment factor from 50 to 400 (Table 3.4.). However, if the abundance of DNA templates encoding binding proteins was low (ratio M.Hae III-EDB/-CaM = 10'000), 5xGGCC DNA did not improve the outcome of the selection. These data imply that cross-contamination after extraction of the water phase is only important if the fraction of binding proteins is relatively high (e.g., after several rounds of selection or in affinity maturation). It should be possible to eliminate cross-contamination after water phase extraction by (*i*) addition of competing DNA in excess, and (*ii*) lowering the number of DNA molecules used as input. After several rounds of selection or in affinity maturation, a population of 10^8-10^9 DNA molecules should be enough to cover the diversity needed.

Second, more than one DNA molecule might be present in one water compartment of the water-in-oil emulsion, which would lead to the incorrect assignment of genotype and phenotype and, therefore, to the selection of DNA molecules coding for non-binding proteins. This could explain why enrichment factors increased by almost a factor of three if 10^9 instead of 10^{10} DNA molecules were used as input for the selection experiments (Table 3.7.). With the mean diameter of a water compartment of the emulsion being ~1 μm and given that 50 μl of water phase are encapsulated in 950 μl of oil phase, there should be roughly 10^{11} distinct water compartments in the resulting emulsion. However, this calculation contradicts the results obtained from selections with 10^{10} and 10^9 DNA molecules. If 10^{11} water compartments were formed in 1 ml of emulsion and 10^{10} DNA molecules were used as input, there would be only 0.1 DNA molecule on average per water compartment. Therefore, one would not expect the enrichment factors to improve if the starting amount of DNA molecules is lowered by a factor of 10. However, enrichment factors improved by decreasing the amount of DNA molecules in the selection and, therefore, the size distributions might be shifted to larger mean diameters as described previously (Tawfik and Griffiths, 1998). It should be noted that emulsions had to be diluted in mineral oil before size distributions could be measured, which could influence the compartment size.

The stability of the emulsions seemed to be reliable, as expected from the measurements of the size distributions. When DNA molecules coding for either M.Hae III-EDB or –CaM proteins were encapsulated separately and then immediately mixed, the enrichment of M.Hae-CaM genes was significantly more efficient than if the two types of templates were mixed before compartmentalisation.

When only 10^9 instead of 10^{10} DNA molecules were used as input in selection experiments, a higher enrichment factor for of M.Hae III-CaM genes was observed. However, if emulsions were mixed right after their preparation, the amount of input DNA had no influence on the outcome of the selection (Table 3.6.). Therefore, in the selections starting from DNA mixtures, there were either several DNA molecules per compartment or transcription and translation already started during stirring of the emulsion, which could have led to a spread of different types of mRNAs and/or proteins to several compartments.

A suitable strategy for the amplification of DNA recovered from affinity selection is crucial to the success of the methodology. Nested PCR, followed by either restriction and ligation or PCR assembly, was discussed in the results section. Another attractive approach would include the preparation of extended DNA templates, which would allow the use of a new primer pair for every amplification step. Such a procedure would be advantageous compared to restriction followed by ligation or PCR assembly because less work and time would be necessary for a complete selection cycle.

4.2. Future developments

There are some immediate experiments, which could help to improve the performance of the proposed selection technology.

- The quality of the oligonucleotide comprising 5-fluorodeoxycytidine might be improved in order to get a higher percentage of cross-linked DNA molecules.

- A methodology allowing the control of the valency of display would be extremely useful. Therefore, the feasibility of the proposed pre-selection on streptavidin beads followed by a quantitative release of the DNA-protein fusions from the beads should be investigated in detail.

- Using DNA mixtures as input, model selection experiments over several rounds should be performed in order to further test the efficacy of the selection technology and to evaluate the different amplification strategies.

- Finally, selections with libraries of proteins or peptides will show the robustness and efficiency of this selection technology for the *in vitro* evolution of proteins.

Besides the already mentioned nano-droplet delivery system, which can be used to change the composition of the water phase within the compartments of the emulsion (Bernath *et al*, 2005), a methodology for the re-emulsification of water-in-oil emulsions giving water-in-oil-in-water (w/o/w) emulsions was recently proposed (Bernath *et al.*, 2004) to allow the sorting of these w/o/w emulsions by flow cytometry. In model experiments, genes embedded in water compartments together with a fluorescent marker were isolated and enriched form an excess of genes embedded in water droplets without a fluorescent marker. These findings might expand the field of applications of water-in-oil emulsions to evolve enzymes converting substrate to fluorescent products.

With the approach presented here, a combination of a very stable physical linkage and high expression levels of different fusion proteins was achieved. In addition, the number of M.Hae III fusion proteins covalently linked per DNA molecules might be varied by simply adjusting the number of M.Hae III substrate sequences containing the modified base 5-fluorodeoxycytidine in the DNA molecule. Covalent complexes of a DNA gene with a defined, small number of copies of the encoded protein could also reduce the bias in selection due to different expression levels.

The tight linkage of DNA and protein could allow the isolation of very stable binding proteins, which are highly soluble, and thus, might have a reduced immunogenic potential and eventually suited for routes of administration different than intravenous injection. Coupling proteins covalently to DNA is potentially not only useful for selection of binding activity, but DNA–protein fusions could also be re-compartmentalised in emulsions and selected for catalysis (Griffiths and Tawfik, 2003; Bernath *et al.*, 2004). In addition, more complex selection protocols for the engineering of allosteric proteins could also be applied more easily owing to the covalent nature of the linkage between protein and DNA.

5. Material and Methods

5.1. Reagents

Reagents and instruments used for the experiments are described in the text below. Water was deionised, filtered using a Milli-Q RG System (Millipore, USA) and autoclaved if used for *in vitro* transcription-translation. If not stated otherwise, centrifugation steps were performed in an Eppendorf 5415D (Eppendorf, Germany) table-top centrifuge. All PCR reactions were done in the presence of 200 μM dNTP nucleotides (Roche Diagnostics, Switzerland) and 400 nM primers if not stated otherwise. If DNA was used for further experiments (e.g. selections, cloning etc.), HiFi DNA polymerase was used (Roche Applied Sciences, Switzerland). If PCR was performed for the analysis of an experiment, and therefore, PCR products were discarded, Taq DNA polymerase was used (Qiagen, Germany). Oligonucleotides were ordered at Operon (Germany, former subsidiary of Qiagen) or Microsynth (Switzerland). PCR primers designated as *ba* are forward, and those designated as *fo* are reverse primers.

5.2. Preparation of water-in-oil emulsions

Water-in-oil emulsions were essentially prepared as described by Tawfik and Griffiths (1998). A 10-fold concentrated solution of oil phase was prepared by mixing 50% (v/v) mineral oil (Sigma-Aldrich) with 45% (v/v) Span 80 (Fluka, Buchs, Switzerland) and 5% (v/v) Tween 80 (Fluka). For the preparation of the emulsions, this concentrated stock solution was diluted 10-fold in mineral oil and 950 μl were used for the encapsulation of 50 μl of water phase. The mixing of the water and the organic phases was done by the incremental addition of 5x10 μl of ice-cooled water phase to ice-cooled oil phase in a glass vial (Forma Vitrum, Switzerland, 40x18.75 mm). Stirring was done at 2200 r.p.m. using a magnetic stirrer (MR 1000, Heidolph, Germany) and a stirring bar (3 mm) with a pivot ring. After addition of water phase during 2 min, stirring was continued for further 5 min to obtain the desired size distribution of the water compartments.

5.3. Laser diffraction measurements

The size distributions of the water compartments were measured by laser diffraction using a Mastersizer X (Malvern Instruments, UK). A 30 μl volume of an emulsion was added to 10 ml of mineral oil and mixed in the presentation cell of the Mastersizer (30–40% stirring power). Particle size distributions were calculated with the instrument's built-in Fraunhofer optical model (Figure 3.1.) or Mie optical model (Figure 3.2.) using refractive indices of 1.470 for mineral oil and 1.330 for the aqueous phase.

5.4. Formation of covalent DNA-protein adducts

A 268 bp DNA fragment (Figure 3.3.) and full-length DNA templates for *in vitro* expression containing the suicide inhibitor sequence 5'-GGFC-3' for M.Hae III were used for the cross-linking experiments. The modified nucleotide 5-fluorodeoxcytidine was introduced by PCR amplification (T3 Thermocycler, Biometra) of DNA supplied in vial 3 of the RTS *Escherichia coli* linear template generation set (268 bp fragment) with the oligonucleotides Hae sub fo (5'-C GTC AT**G GFC** TAT GCG GGC GAC CAC ACC CGT CCT GTG GAT-3', modified substrate sequence in bold letters) and Hae sub ba 1 (5'-TAT TGC TAT GTA CCT AGC GGG GGG GGT TCT CAT CAT CAT-3') (both oligonucleotides from Microsynth, Switzerland). Full-length DNA templates were assembled by PCR employing the same template generation set (Roche Applied Sciences) with the synthetic oligonucleotides ext ba 2 (5'-GAT GCC GGC CAC GAT GCG TCC GGC GTA GAG-3', Qiagen) and ext 2 fo (5'-GCT AAT TA**G GFC** ACC ACA CCC GTC CT-3', modified substrate sequence in bold letters; Microsynth, Switzerland).

A 2 nM amount of DNA was incubated with 38 nM M.Hae III (New England Biolabs, USA) and 80 μM S-adenosylmethionine (New England Biolabs) in DNA-methyltransferase reaction buffer (50 mM NaCl, 50mM Tris–HCl (pH 8.5), 10 mM dithiothreitol; New England Biolabs). Cross-linking reactions were stopped by heat inactivation (15 min at 70 °C). Samples were loaded on a denaturing 10% TBE urea or a native 6% TBE polyacrylamide gel (Invitrogen) and DNA was visualised with SYBRgreen I or II (MolecularProbes, USA). Gels were run in TBE buffer (10.8 g/l Tris-Cl, 5.5 g/l boric acid, 0.4 mM EDTA, pH = 8.0) at 200V, 15 mA for about 1 h. 10bp DNA Ladder (Invitrogen) or Smart Ladder (Eurogentech) was used as DNA marker. If the samples were analysed on a native gel, DNA gel

loading buffer (5x concentrated: 40% (w/v) saccharose, 0.25% (w/v) bromophenol blue (Sigma, USA)) was added before the heating step to 70°C.

5.5. Cloning in *Escherichia coli*

A strain designated as TG1 was used for all cloning procedures (K12, D(*lac-pro*), *supE, thi, hsdD5/F'traD36, proA*$^{+}$*B*$^{+}$*, laqI*^{q}*, lacZDM15)*. Bacteria were grown in liquid media (2xYT (Bio101 Incorporation, USA): 16 g/l bacto-tryptone, 10 g/l bacto-yeast extract, 5 g/l NaCl pH 7.4), which had been previously sterilised by autoclaving, or on agar plates (2xYT as liquid medium plus 15 g/l agar, autoclaved). For the preparation of agar plates, 100 μg/ml ampicillin (Applichem, Germany) and 1% glucose were added to the agar when a temperature of 50°C was reached after autoclavation. Warm agar was poured into petri dishes with a diameter of 10 cm. Bacteria were grown in liquid media at 37°C in an orbital shaker at 200 r.p.m. and the optical density (OD) was measured at a wavelength of 600 nm with a spectrophotometer (Cary300, Varian, USA). Glycerol stocks were made with cultures at an OD(600 nm) of about 1.

Electrocompetent TG1 cells were prepared by inoculating 1 l of 2xYT medium with 1:100 volume of an overnight culture of bacteria. The bacteria were then grown to an OD(600) = 0.5-0.7, spun at 4000x g (Sorvall RC5Cplus centrifuge) for 15 min in pre-chilled bottles and a pre-chilled rotor (4°C) (SLA 3000, Sorvall), and resuspended in the original volume of sterile water containing 5% glycerol (Hänseler, Switzerland). After this, cells were centrifuged as before, but were resuspended only in half of the original volume of ice-cold sterile water/5% glycerol. Then, the cells were centrifuged again and resuspended in 40 ml sterile water with 10% glycerol. After a final centrifugation and resuspension in 2 ml sterile water/10% glycerol, the bacteria were frozen in liquid nitrogen and stored at –80°C in 50 μl aliquots.

M.Hae III-Ala-Gly-6xHis was cloned into the vector pQE-12 (Qiagen, Germany) using the restriction sites EcoRI and HindIII. Plasmid DNA was isolated with a QIAfilter Plasmid Midi or Maxi kit from Qiagen (Germany) and the yield was determined by loading different volumes of isolated DNA on an agarose gel. The gene coding for M.Hae III was amplified from *Haemophilus aeqypticus* (ATCC strain 11116) using the primers M.Hae Eco ba (5'-CAC ACA GAA TTC ATT AAA GAG GAG AAA TTA ACT ATG AAT TTA ATT AGT CTT TTT TCA GGT-3') and M.Hae

Hind fo (5'-AGA TCT AAG CTT TTA CCC TTA GTG ATG GTG ATG GTG ATG TCC TGC ATT ACC TTT ACA AAT TTC CAA TGC-3') using the PCR cycling program 94°C(10')-[94°C(1')-57°C(1')-72°C(2'30'')]$_{30}$-72°C(5'). PCR products were purified by gel electrophoresis on agarose gel (1.2%).

Both vector pQE-12 and the DNA fragments to be cloned were digested with the restriction enzymes EcoRI and HindIII (Promega, USA) during 4 h at 37°C. Digested plasmid was purified by gel electrophoresis, cut from the gel, and elution from the gel was performed with QIAquick gel extraction kit (Qiagen, Germany). Digested insert DNA was purified by using QIAquick PCR purification kit (Qiagen, Germany). Vector and insert were ligated with T4 DNA ligase (QBiogen, Switzerland) for 2 h at room temperature and purified using the QIAquick PCR purification kit.

4 μl of the purified ligation was electroporated into 50 μl electrocompetent TG1. Bacteria and ligation were added to the bottom of a pre-cooled electroporation cuvette (BioRad, Germany). The settings of the electroporator (BTX electro cell manipulator 600) were 2.5 kV, 25 mF and 200 Ω. Just after the electric pulse for electroporation, 50 μl of ice-cold 2xYT was added to the cuvette and the cells were plated on agar plates containing the antibiotic ampicillin (plates were incubated at 37°C overnight). The next day, bacterial colonies were screened by PCR and positive clones were used for expression screening (see chapter 5.5.). The best expressing clone, pJB18/1 was used for the production of M.Hae III-6xHis protein.

5.5 Protein production in *E.coli*

Freshly electroporated TG1 cells (with plasmid pJB18/1) were used for the expression of M.Hae III-6xHis tag DNA-methyltransferase. A pre-culture (10 ml 2xYT with 100 μg/ml ampicillin and 2% glucose) of *E.coli* transformed with pJB18/1 was grown overnight at 37°C in a rotary shaker at 200 r.p.m.. On the following day, 2x100 ml of 2xYT containing ampicillin were each inoculated with 100 μl of the pre-culture. Bacteria were grown at 37°C until the optical density at 600 nm was 0.6, which was followed by induction of protein expression by addition of 1 mM isopropyl-β-thiogalactoside (IPTG, Applichem, Germany). Protein expression was performed overnight at 30°C, shaking at 200 r.p.m. The day after, bacteria were pelleted by centrifugation and resuspended in 10 ml lysis

buffer (50 mM NaH_2PO_4, 300 nM NaCl, 10 mM imidazole, pH = 8.0). 1 mg/ml lysozyme was added and the cells were incubated on ice for 30 min. After cell lysis by sonication, the lysates were centrifuged for 25 min at 11'000 r.p.m.. 2 ml of Ni^{2+}-NTA slurry (Qiagen) was added to the cleared lysate to capture M.Hae III-6xHis tag proteins (incubation on ice for 1 h whilst shaking on a rotary shaker). The resin was washed 2 times with 10 ml wash buffer (50 mM NaH_2PO_4, 300 nM NaCl, 20 mM imidazole, pH = 8.0) and the protein was eluted with 4x 1 ml elution buffer (50 mM NaH_2PO_4, 300 nM NaCl, 250 mM imidazole, pH = 8.0). The protein was not dialysed but directly stored after the addition of glycerol (50% (v/v) and 0.1 M DTT (Fluka, Switzerland)). The concentration of the purified protein was determined by SDS-PAGE using known amounts of a standard protein (Ovalbumin, Sigma, USA).

5.6. M.Hae III activity assay

100 ng DNA molecules containing a NotI restriction site were incubated for 1 h at 37°C in methylase reaction buffer (see chapter 5.4.) with 80 μM S-adenosylmethionine (SAM) and M.Hae III proteins (either purified or in 1:10 diluted transcription-translation mixture after expression). After this, the samples were heated to 70°C for 15 min to inactivate M.Hae III. $MgCl_2$ was added to a final concentration of 10 mM followed by NotI (10 units) restriction enzyme (Qbiogen, Switzerland) and the samples were incubated for 2 h at 37°C. The outcome of the experiment was analysed on a 1.2% agarose gel stained with ethidium bromide. Smart Ladder DNA marker (Eurogentech) was loaded as standard.

5.7. Preparation of templates for *in vitro* expression

The gene coding for Hae III DNA-methyltransferase was ordered at ATCC (strain 11116) and amplified by PCR using the primers Hae vitro ba (5'-CT TTA AGA AGG AGA TAT ACC ATG AAT TTA ATT AGT CTT TTT TCA GGT-3') and Hae vitro fo (5'-TG ATG ATG AGA ACC CCC CCC ATT ACC TTT ACA AAT TTC CAA TGC-3') (PCR program: 94°C(10')–[94°C(1')–54°C(1')–72°C(2'30'')]$_{25}$–72°C(5')).

The DNA templates for the expression of the various M.Hae III constructs were assembled following the instructions of the RTS *E.coli* Linear Template Generation

Set (Roche Applied Sciences) with the exception that the oligonucleotides ext ba 2 (5'-GAT GCC GGC CAC GAT GCG TCC GGC GTA GAG-3', Qiagen) and ext 2 fo (5'-GCT AAT TA**G GFC** ACC ACA CCC GTC CT-3', modified substrate sequence in bold letters; Microsynth, Switzerland) were used for the final PCR assembly instead of the primers supplied with the linear template generation set. DNA was purified using a QIAquick gel extraction kit or QIAquick PCR purification kit (Qiagen). If DNA was used directly for *in vitro* transcription/translation reactions, DNA samples never were purified over an agarose gel because agarose inhibits *in vitro* protein expression (Manual RTS *E.coli* HY kit, Roche Applied Sciences).

Further details concerning the assembly of the different genetic M.Hae III fusions are described below.

PCR cycling for assembly: 94°C(3')-[94°C(1')-58°C(1')-72°C(2'15'')]$_{25}$-72°C(4')

M.Hae III-Flag tag: the amino acid residues GGSG-DYKDDDK followed by two stop codons were appended to the DNA-methyltransferase gene (PCR with the primer Hae vitro Flag fo: 5'-TG ATG ATG AGA ACC CCC CCC TTA AGC TTA TTT GTC ATC GTC GTC CTT GTA GTC TCC TGA GCC ACC ATT ACC TTT ACA AAT TTC CAA TGC-3') in order to avoid the expression of the 6xHis tag, which is included in the linear template generation set.

For the construction of M.Hae III-EDB/CaM fusion proteins, M.Hae III was amplified with a special reverse primer (M.Hae lin fo: 5'-GCT TGC ACC TGA TCC GCC ATT ACC TTT ACA AAT TTC CAA TGC-3') to introduce the overlap-sequence needed for the assembly of M.Hae III with the genes encoding EDB and CaM.
The sequences coding for calmodulin (CaM) (Demartis *et al.*, 1999) and the extra-domain B of fibronectin (EDB) (Fattorusso *et al.,* 1999) were available in our laboratory. The amino acid residues GGSGAS were used as linker for the C-terminal fusion of EDB/CaM to M.Hae III. EDB was amplified using the primers Lin EDB ba (5'-GGT AAT GGC GGA TCA GGT GCA AGC GAG GTG CCC CAA CTC ACT GAC CTA-3') and EDB vitro fo (5'-TG ATG ATG AGA ACC CCC CCC CGT TTG TTG TGT CAG TGT AGT AGG-3'), whereas CaM was amplified using the primers Lin Cam ba (5'-GGT AAT GGC GGA TCA GGT GCA AGC GCT GAC CAA CTG ACA GAA GAG CAG-3') and CaM vitro fo (5'-TG ATG ATG AGA ACC CCC CCC CTT TGC TGT CAT CAT TTG TAC-3'). M.Hae III and EDB/CaM coding sequences were assembled by PCR using the following cycling program (94°C(3')-[94°C(1')-55°C(1')-72°C(2'30'')]$_{30}$-72°C(5').

For the construction of GST-M.Hae III-6xHis tag DNA template, the M.Hae III gene was cloned as a C-terminal fusion to glutathione-S-transferase (GST) in the vector pGEX-4T-2 (Amersham Biosciences) using the restriction sites BamHI and NotI (both enzymes from Promega). Both genes were amplified together from the resulting vector (primers: GST vitro ba: 5'-CT TTA AGA AGG AGA TAT ACC ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT-3' and Hae vitro fo: 5'-TG ATG ATG AGA ACC CCC CCC ATT ACC TTT ACA AAT TTC CAA TGC-3') and the PCR product was taken for PCR assembly using the linear template generation set.

5.8. Protein expression *in vitro* and western blot

All *in vitro* transcription/translation reactions were performed using the RTS 100 *E.coli* HY kit (Roche Applied Sciences). In order to allow the formation of covalent DNA–protein adducts, 80 μM S-adenosylmethionine was added to the transcription-translation mix. For western blot analysis, 2 μl of the transcription-translation mixtures (non-emulsified expression) were loaded on a 10% Bis-Tris SDS–polyacrylamide gel [Invitrogen, protein marker RPN 800 (Amersham Biosciences)] followed by blotting on a PVDF membrane (0.45 μm pores, Immobilon Transfer Membrane, Millipore), which was previously activated by soaking for 15 sec in 100% methanol followed by 2 min in water. After blotting, the PVDF membrane was quickly rinsed with water, soaked for 2x15 sec in methanol and then air-dried on a filter-paper. M.Hae III fusion proteins bearing a 6xHis tag were detected using an anti-His-HRP immunoconjugate (Sigma-Aldrich). M.Hae III-Flag tag protein was detected using anti-Flag tag (M2) as primary antibody (Sigma-Aldrich) and anti-mouse-HRP immunoconjugate as secondary antibody (Sigma-Aldrich). The PVDF membrane was incubated with the antibodies for 45 min at room temperature in 3% milk PBS (shaking at 65 r.p.m.). Subsequently, the membrane was washed four times with PBS. HRP activity was detected with the ECL kit (Amersham Biosciences) as indicated by the manufacturer.

If expression was performed in emulsions, 50 μl of transcription-translation mixture were emulsified and after expression, the water phase was extracted in a volume of 100 μl. 10 μl of this extract was used for western blot analysis as described above.

5.9. Native gel electrophoresis with M.Hae III-DNA complexes

Samples from cross-linking reactions either in methylase reaction buffer (New England Biolabs, USA), or in diluted/undiluted transcription-translation mixture, were mixed with DNA loading buffer (5x concentrated: 40% (w/v) saccharose, 0.25% (w/v) bromophenol blue (Sigma, USA)). Undiluted transcription-translation mixtures could not be loaded on native gels because the samples hardly entered the gel and subsequent staining revealed smears covering large areas on the gel. Therefore, undiluted transcription-translation mixture was diluted 1:10 in water, followed by the addition of DNA loading buffer (5x concentrated: 40% (w/v) saccharose, 0.25% (w/v) bromophenol blue (Sigma, USA)).
If only covalently cross-linked DNA-protein fusions should be visualised on the gel, samples were heated to 70°C for 15 min. Samples were not heated, if non-covalent DNA-protein complexes should be detected on the gels, too.

6% TBE polyacrylamide gels (Invitrogen) were run in TBE buffer (10.8 g/l Tris-Cl, 5.5 g/l boric acid, 0.4 mM EDTA, pH = 8.0) at 200 V, 15 mA for about 1 h. DNA was stained for 30 min in TBE buffer containing SYBRgreen I (Molecular Probes, USA) as recommended by the manufacturer. Pictures of the gels were taken under UV light using the DIANA III imaging system (raytest, Germany).

5.10. Model selection experiments

In vitro transcription-translation mixture (TS/TL-mix) was prepared on ice according to the instructions of the manufacturer of the kit (RTS 100 E.coli HY kit, Roche Applied Sciences). A 40 μl volume of the TS/TL-mix, 5 μl of S-adenosylmethionine (stock, 800 μM; final concentration, 80 μM), DNA and 5 μl of water were mixed in order to obtain a final volume of 50 μl. DNA was added just before preparing the emulsion. A 50 μl volume of ice-cooled water phase (*in vitro* TS/TL-mix containing the DNA template) was added to 950 μl of oil phase (ice-cold) in a stepwise manner (5x10 μl during 2 min) whilst stirring at 2200 r.p.m. in a glass vial. After stirring, the emulsion was transferred to a 1.5 ml microcentrifuge tube. Proteins were expressed in emulsion by incubation for 150 min in a heat block at 30°C. After protein expression and formation of the covalent DNA–protein adducts, emulsions were broken by centrifugation for 10 min at 7000 r.p.m. in a table-top centrifuge (Eppendorf). Water droplets formed

a pellet and the organic supernatant was removed. A 150 μl volume of room temperature breaking buffer was added [PBS (or TBS with 1 mM $CaCl_2$ for selections with calmodulin (TBSC)], pH 7.4, 5 μM biotinylated ds DNA fragments [5'-biotin-GGA GCT TCT GCA TTC TGT GTG CTG-3' (Qiagen)], 1 μM competing ds DNA fragments (only in selections with mixtures of different DNA molecules) [5'-ATC TAA **GGC C**AA TGT ACT AGA **CGG C**CA TTC CAG ATG CA**G GCC** AAG CGT ACA TAC **GGC C**TA GCT AAA TCA A**GG CC**G TAT CGT-3', substrate recognition sites of M.Hae III in bold letters, (Qiagen)] followed by 1 ml of room temperature diethyl ether (water saturated). The tube was vortexed at full speed for 2x10 sec. After separation of the water phase and the organic phase, the water phase was removed from the bottom of the tube and transferred to a 24-well plate (NUNCLON Surface, Nunc). The extracted water phase was air-dried at room temperature for 10 min until residual Et_2O was completely evaporated.

Volumes of 25–50 μl of magnetic streptavidin-coated Dynabeads (Dynal, Norway) were incubated for 15 min with the capture agent at room temperature {biotinylated anti-penta-His antibody (4 μl or 800 ng/50 μl beads) (Qiagen), biotinylated anti-Flag antibody (M2, 2 μl or 2 μg/50 μl beads) (Sigma- Aldrich), biotinylated anti-EDB antibody (5 μl/50 μl beads, 200 nM) [L19 small immunoprotein (Li et al., 1997), which was kindly provided by P.Alessi, Institute of Pharmaceutical Sciences, ETH Zurich, Switzerland], biotinylated calmodulin binding peptide [biotin-CAAARWKKAFIAVSAANRFKKIS (Montigiani et al., 1996), which was kindly provided by L. Lozzi, Dipartimento di Biologia Molecolare, Universita' di Siena, Italy] (4 μl/50 μl beads, 800 nM)}. The beads were washed once with PBS 0.1% Tween 20 (PBST) or TBSC 0.1% Tween 20 (TBSCT) (if CaM-binding peptide was used as capture agent) and blocked with 5 μM biotinylated ds DNA fragments [5'-biotin-GGA GCT TCT GCA TTC TGT GTG CTG-3' (Qiagen)] for 15 min at room temperature. The beads were then incubated with the extracted water phase for 45 min at room temperature. The samples were shaken gently from time to time or placed on a rotary shaker at 100-150 r.p.m. The magnetic beads were washed six times with 100 μl of PBST or TBSCT, followed by one washing step with 100 μl of PBS or TBSC using a magnetic separator (Dynal, Norway). The beads were resuspended in 100 μl of water and 1 μl (in 25 μl PCR volume) was used directly for PCR amplification of the DNA molecules remaining on the beads using the primers Ampl ba (5'-CCC GCG AAA TTA ATA CGA CTC A-3', Qiagen) and Ampl fo (5'-AAA ACC CCT CAA GAC CCG TT-3', Qiagen) (PCR cycles: 94°C(3')-[94°C(1')-51°C(1')-72°C(1'40'')]$_{30}$-72°C(4')). In selections with

mixtures of DNA, 5 µl of the beads were taken for amplification (in 25 µl PCR volume) and the products of the reaction were analysed by real-time PCR. As an exception, PCRs for Figures 3.12./3.16./3.17. were done using the primers M.Hae Nco ba (5'-GGA GAT ATA ACC ATG GGC AAT TTA ATT AGT CTT TTT TCA GGT-3', Qiagen) and M.Hae Xho His fo (5'-TAT AGC TCG CTC GAG ATT ACC TTT ACA AAT TTC CAA TGC AGA-3', Qiagen) (PCR cycles: 94°C(3')-[94°C(1')-55°C(1')-72°C(1'30'')]$_{25}$-72°C(4')). Smart Ladder (Eurogentech) was used as DNA marker on the agarose gels.

In capture experiments using Ni^{2+}-beads, proteins were not expressed in emulsions. After transcription and translation (100 ng DNA template in 25 µl volume) of M.Hae III proteins in free solution at 30°C for 2 h, 50 µl lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH = 8.0) and 20 µl nickel coated beads were added to the samples. After incubation for 45 min at room temperature, the beads were washed 6 times in 100 µl washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.1% Tween 20, pH = 8.0). After the last washing step, beads were resuspended in 100 µl PBS and directly used for PCR (1 µl beads in 25 µl PCR volume) with the primers M.Hae Nco ba and M.Hae Xho His fo (as above).

5.11. Selections under denaturing conditions

Selections under denaturing conditions were performed with DNA-protein fusions from cross-linking reactions with purified M.Hae III-6xHis in methylase reaction buffer or from transcription and translation reactions in a volume of 15 or 25 µl using 60 ng of DNA as template. After expression and cross-linking for 2.5 h at 30°C, 65 µl denaturing buffer (100 mM NaH_2PO_4, 10 mM Tris-Cl, 6 M GuHCl, 0.1% Tween 20, pH = 8.0) and Herring Sperm DNA (to a final concentration of 1 mg/ml) were added. 20 µl Ni^{2+}-beads were mixed with the samples and incubated for 60 min at room-temperature on a rotary shaker at 200 r.p.m.. Beads were washed 6x in 100 µl washing buffer (100 mM NaH_2PO_4, 10 mM Tris-Cl, 01% Tween 20, 8 M urea, pH = 8.0) followed by 3 washes with 100 µl PBS 01.% Tween 20. Finally, the beads were resuspended in 100 µl PBS and 1 µl of them was taken for PCR analysis (25 µl volume) using the primers Ampl ba and Ampl fo (Figure 3.14.) or M.Hae Nco ba and M.Hae Xho His fo (Figure 3.15.).

5.12. Real-time PCR

An ABI PRISM 7000 instrument (Applied Biosystems), ABI PRISM 96-well Optical Reaction Plates with MiniAmp Optical Caps (Applied Biosystems) and TaqMan Universal PCR Master Mix (Applied Biosystems) were used for all real-time PCR measurements. All reactions were done in triplicate for each sample. Primers and TaqMan® probes were designed with Primer Express 1.5 software (Applied Biosystems) (Table 5.1.). Primers were used at a concentration of 200 nM and the TaqMan® probe at 100 nM. DNA samples were diluted in such a way that the expected number of molecules detected by real-time PCR was between 10^2 and 10^6.

In selection experiments with only one DNA template, 0.1 μl of the streptavidin beads were taken for real-time PCR (25 μl volume). In selection experiments starting from a mixture of DNA templates, DNA bound to the beads was first amplified by PCR using the primers Ampl ba and Ampl fo (PCR cycles see above). PCR products were not purified but directly used for real-time PCR (diluted $1:10^3$-10^5).

Table 5.1. TaqMan® probes and primers used for real-time PCR.

Target gene	Probe/primers	Sequence
M.Hae III	TaqMan Hae	5′-GAACAGCCTTATTATGACGCTGAGCCATCA-3′
	RT Hae ba	5′-TTTCTTGCCGAAAATGTCAAAG-3′
	RT Hae fo	5′-CAGCATTATCAAATTCTTGGATAATTCT-3′
CaM	TaqMan CaM	5′-CGGTCGCTTGGACAAAACCCAACA-3′
	RT CaM ba	5′-CACAAAGGAACTTGGCACTGTTAT-3′
	RT CaM fo	5′-TGATCATATCCTGCAATTCTGCTT-3′
EDB	TaqMan EDB	5′-TCCACCATTATTGGGTACCGCATCACA-3′
	RT EDB ba	5′-GGTGGACCCCGCTAAACTC-3′
	RT EDB fo	5′-ACCTTCTCCTGCCGCAACTA-3′

PCR cycling program used for all real-time PCR measurements:

$$50°C(2') - 95°C(10') - [95°C(15'') - 60°C(1')]_{40\ cycles}$$

5.13. Nested PCR and PCR assembly

Nested PCR was performed using two different primer pairs. The first PCR was carried out using primers, which align at the extremities of the DNA templates (vial 1 and vial 2 of Roche Linear Template Generation Set). 10 µl of totally 100 µl beads were used in a volume of 50 µl per PCR reaction, and DNA molecules were amplified using the following temperature cycles: 94°C(3')-[94°C(1')-58°C(1')-72°C(2'15'')]$_{25}$-72°C(4'). The PCR products were purified with QUAquick PCR purification kit and 10% of the resulting DNA solution was used as template for the second PCR with the primers Ampl ba and Ampl fo (primer sequences see above) (PCR program: 94°C(3')-[94°C(1')-51°C(1')-72°C(1'40'')]$_{25}$-72°C(4')). Products from this second PCR were loaded on an 1.5% agarose gel and the bands were excised and purified (QIAquick gel extraction kit).

The two outer DNA fragments were produced by amplification of two stretches of DNA of the vector pIVEX2.3d (Roche Applied Sciences, Switzerland). The DNA fragment upstream of M.Hae III was amplified with the primers ext 5' ba and Ampl x fo, whereas the fragment downstream of the expression cassette was amplified with the primers Ampl y ba and ext 3' fo (PCR cycling: 94°C(3')-[94°C(45'')-51°C(1')-72°C(1')]$_{25}$-72°C(3')). The DNA fragments were purified from an agarose gel using the QIAquick gel extraction kit.

The three DNA fragments were assembled in PCR by mixing 100 ng of each of the fragments in a total volume of 30 µl. External primers were added (ext 5' ba and ext 3' fo) and assembly was performed using the following PCR program: 94°C(3')-[94°C(1')-51°C(1')-72°C(2'30'')]$_{7}$-72°C(4'). The products of PCR assembly were loaded on an agarose gel and the band of with correct size was excised in order to purify the DNA.

6. References

Arnold, F.H. (1998) *Nat. Biotechnol.*, **16**, 617-618.

Atwell, S. and Wells, J.A. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 9497-9502.

Baggio, R., Burgstaller, P., Hale, S.P., Putney, A.R., Lane, M., Lipovsek, D., Wright, M.C., Roberts, R.W., Liu, R., Szostak, J.W. and Wagner R.W. (2002) *J. Mol. Recognit.*, **15**, 126-134.

Bagshawe, K.D. (1987) *Br. J. Cancer*, **56**, 531-532.

Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A.D., Magdassi, S. and Tawfik, D.S. (2004) *Anal. Biochem.*, **325**, 151-157.

Bernath, K., Magdassi, S. and Tawfik, D.S. (2005) *J. Mol. Biol.*, **345**, 1015-1026.

Bertschinger, J. and Heinis, C. (2003) *International patent application,* WO 2005/030957 A1.

Bertschinger, J. and Neri, D. (2004) *Protein Eng. Des. Sel.*, **17**, 699-707.

Bessette, P.H., Rice, J.J. and Daugherty, P.S. (2004) *Protein Eng. Des. Sel.*, **17**, 731-739.

Beste, G., Schmidt, F.S., Stibora, T. and Skerra, A. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 1898-1903.

van den Beucken, T., van Neer, N., Sablon, E., Desmet, J., Celis, L., Hoogenboom, H.R. and Hufton, S.E. (2001) *J. Mol. Biol.*, **310**, 591-601.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G. and Plückthun, A. (2004) *Nat. Biotechnol.*, **22**, 575-582.

Boder, E.T. and Wittrup, K.D. (1997) *Nat. Biotechnol.*, **15**, 553-557.

Boder, E.T. and Wittrup, K.D. (2000) *Methods Enzymol.*, **328**, 430-444.

Boder, E.T., Midelfort, K.S. and Wittrup, K.D. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 10701-10705.

Bork, P. (1993) *Proteins: Sruct. Funct. Genet.*, **17**, 363-374.

Chen, G., Hayhurst, A., Thomas, J.G., Harvey, B.R., Iverson, B.L. and Georgiou G. (2001) *Nat. Biotechnol.*, **19**, 537-542.

Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nipkay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry,* **30**, 11018-11025.

Chen, L., MacMillan, A.M., Verdine, G.L. (1993) *J. Am. Chem. Soc.*, **115**, 5318-5319.

Chester, K.A., Begetn, R.H., Robson, L., Keep, P., Pedley, R.B., Boden, J.A., Boxer, G., Green, A., Winter, G., Cochet, O. *et al.* (1994) *Lancet*, **343**, 455-456.

Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. (2000) *J. Mol. Biol.*, **297**, 309-319.

Cohen, H.M., Tawfik, D.S. and Griffiths, A.D. (2004) *Protein Eng. Des. Sel.*, **17**, 3-11.

Colby, D.W., Garg, P., Holden, T., Chao, G., Webster, J.M., Messer, A., Ingram, V.M. and Wittrup, K.D. (2004a) *J. Mol. Biol.*, **342**, 901-912.

Colby, D.W., Chu, Y., Cassady, J.P., Duennwald, M., Zazulak, H., Webster, J.M., Messer, A., Lindquist, S., Ingram, V.M. and Wittrup, K.D. (2004b) *Proc. Natl. Acad. Sci. USA*, **101**, 17616-17621.

Cortez-Retamozo, V., Backmann, N., Senter, P.D., Wernery, U., De Baetselier, P., Muyldermans, S. and Revets, H. (2004) *Cancer Res.*, **64**, 2853-2857.

Cowan, S.W., Newcomer, M.E., Jones, T.A. (1990) *Proteins*, **8**, 44-61.

Davies, J. and Riechmann, L. (1994) *FEBS Lett.*, **339**, 285-290.

Davies, J. and Riechmann, L. (1995) *Biotechnology*, **13**, 475-479.

Deamer, D.W. (1997) *Microbiol. Mol. Biol. Rev.*, **61**, 763-797.

Decanniere, K., Muyldermans, S. and Wyns, L. (2000) *J. Mol. Biol.*, **300**, 83-91.

Demartis, S., Huber, A., Viti, F., Lozzi, L., Giovannoni, L., Neri P., Winter, G. and Neri, D. (1999) *J. Mol. Biol.*, **286**, 617-633.

Desmyter, A., Transue, T.R., Ghahroudi, M.A., Thi, M.H., Poortmans, F., Hamers, R., Muyldermans, S. and Wyns, L. (1996) *Nat. Struct. Biol.*, **3**, 803-811.

Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S. and Cambillau C (2002) *J. Biol. Chem.*, **277**, 23645-23650.

Doi, N. and Yanagawa, H. (1999) *FEBS Lett.,* **457**, 227-230.

Doi, N., Kumadaki, S., Oishi, Y., Matsamura, N. and Yanagawa, H. (2004) *Nucleic Acids Res.*, **32**, e95.

Dreher, M.L., Gherardi, E., Skerra, A. and Milstein, C. (1991) *J. Immunol. Methods*, **139**, 197-205.

Dumoulin, M., Conrath, K., Van Meirhaeghe, A., Meersman, F., Heremans, K., Frenken, L.G., Muyldermans, S., Wyns, L. and Matagne A (2002) *Protein Sci.*, **11**, 500-515.

Ewert, S., Cambillau, C., Conrath, K. and Pluckthun, A. (2002) *Biochemistry*, **41**, 3628-3636.

Ewert, S., Huber, T., Honegger, A. and Pluckthun, A. (2003) *J. Mol. Biol.*, **325**, 531-553.

Fattorusso, R., Pellecchia, M., Viti, F., Neri, P. and Neri, D. (1999) *Structure Fold Des.*, **7**, 381-390.

Feldhaus, M.J., Siegel, R.W., Opresko, L.K., Coleman, J.R., Feldhaus, J.M.W., Yeung, Y.A., Cochran, J.R., Heinzelman, P., Colby, D., Swers, J., Graff, C., Wiley, H.S. and Wittrup, K.D. (2003) *Nat. Biotechnol.*, **21**, 163-170.

Feldhaus, M.J. and Siegel, R.W. (2004) *J. Immunol. Methods*, **290**, 69-80.

Flower, D.R. (1996) *Biochem. J.*, **318**, 1-14.

Forrer, P., Stumpp, M.T., Binz, H.K. and Pluckthun, A. (2003) *FEBS Lett.*, **539**, 2-6.

Francisco, J.A., Campbell, R., Iverson, B.L. and Georgiou, G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 10444-10448.

Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L. and Curitss III, R. (1997) *Nat. Biotechnol.*, **15**, 29-34.

Ghadessy, F.J., Ong, J.L., Holliger, P. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 4552-4557.

Ghadessy, F.J. and Holliger, P. (2004) *Protein Eng. Des. Sel.*, **17**, 201-204.

Ghadessy, F.J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R. and Holliger P. (2004) *Nat. Biotechnol.*, **22**, 755-759.

Gherardi, E., Pannell, R. and Milstein, C. (1990) *J. Immunol. Methods*, **126**, 61-68.

Giovannoni, L., Viti, F., Zardi, L. and Neri D. (2001) *Nucleic Acids Res.*, **29**, e27.

Graff, C.P., Chester, K., Begent, R. and Wittrup K.D. (2004) *Protein Eng. Des. Sel.*, **4**, 293-304.

Griffiths, A.D. and Tawfik, D.S. (2002) *US patent* 6,489,103 B1.

Griffiths, A.D. and Tawfik, D.S. (2003) *The EMBO Journal*, **22**, 24-35.

Hanes, J. and Pluckthun, A. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 4937-4942.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R. and Pluckthun, A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 14130-14135.

Hanes, J, Jermutus, L., Schaffitzel, C. and Pluckthun, A. (1999) *FEBS Lett.*, **450**, 105-110.

Hanes, J., Schaffitzel, C., Knappik, A. and Pluckthun, A. (2000) *Nat. Biotechnol.*, **18**, 1287-1292.

Harvey, B.R., Georgiou, G., Hayhurst, A., Jeong, K.J., Iverson, B.L. and Rogers, G.K. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 9193-9198.

Heinis, C., Huber, A., Demartis, S., Bertschinger, J., Melkko, S., Lozzi, L., Neri, P. and Neri, D. (2001) *Protein Eng.*, **14**, 1043-1052.

Heinis, C., Melkko, S., Demartis, S. and Neri, D. (2002) *Chem. Biol.*, **9**, 383-390.

Hogbom, M., Eklund, M., Nygren, P.A. and Nordlund, P. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 3191-3196.

Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J. and Uhlen, M. (2005) *Electrophoresis*, **26**, 501-510.

Holt, L.J., Herring, C., Jespers, L.S., Woolven, B.P. and Tomlinson, I.M. (2003) *Trends in Biotechnol.*, **21**, 484-490.

Huber, A.H. and Weis, W.I. (2001) *Cell*, **105**, 391-402.

Ibrahim, S.F. and van den Engh, G. (2003) *Curr. Opin. Biotechnol.*, **14**, 5-12.

Jespers, L., Schon, O., Famm, K. and Winter, G. (2004) *Nat. Biotechnol.*, **22**, 1161-1165.

Jestin, J.-L., Volioti, G. and Winter G. (2001) *Res. Microbiol.*, **152**, 187-191.

Joyce, G.F. (1989) *Nature*, **338**, 217-224.

Keefe, A.D. and Szostak, J.W. (2001) *Nature*, **410**, 715-718.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wllnhofer, G., Hoess, A., Wolle, J., Pluckthun, A. and Virnekas, B. (2000) *J. Mol. Biol.*, **296**, 57-86.

Kobe, B. and Deisenhofer, J. (1995) *Nature*, **374**, 183-186.

Koide, A., Bailey, C.W., Huang, X. and Koide, S. (1998) *J. Mol. Biol.*, **204**, 1141-1151.

Korndorfer, I.P., Beste, G. and Skerra, A. (2003) *Proteins*, **53**, 121-129.

Kurtzman, A.L., Govindarajan, S., Vahle, K., Jones, T.J., Heinrichs, V. and Patten, P.A. (2001) *Curr. Opin. Biotechnol.*, **12**, 361-370.

Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.*, **206**, 305-312.

Lee, Y.F., Tawfik, D.S. and Griffiths, A.D. (2002) *Nucleic Acids Res.*, **30**, 4937-4944.

Li, E., Pedraza, A., Bestagno, M., Mancardi, S., Sachez, R. and Burrone, O. (1997) *Protein Eng.*, **10**, 731-736.

Lipovsek, D. and Pluckthun, A. (2004) *J. Immunol. Methods*, **290**, 51-67.

Liu, R., Barrick, J.E., Szostak, J.W. and Roberts, R.W. (2000) *Methods Enzymol.*, **318**, 268-93.

Lu, C.-F., Montijn, R.C., Brown, J.L., Klis, F., Kurjan, J., Bussey, H. and Lipke, P.N. (1995) *J. Cell. Biol.*, **128**, 333-340.

Main, A.L., Harvey, T.S., Baron, M., Boyd, J. and Campbell, I.D. (1992) *Cell*, **71**, 671-643.

Malek, S., Huxford, T. and Gosh, G. (1998) *J. Biol. Chem.*, **273**, 25427-25435.

Marvin, D.A. (1998) *Curr. Opin. Biotechnol.*, **8**, 150-158.

Masai, H., Kaziro, Y., and Arai, K. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6814-6818.

Masai, H. and Arai, K. (1988) *Nucleic Acids Res.*, **16**, 6493-6514.

Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) *Proc. Natl Acad. Sci. USA,* **91**, 9022-9026.

Mercader, J.V. and Skerra, A. (2002) *Anal. Biochem.*, **308**, 269-277.

Michaely, P., Tomchick, D.R., Machius, M. and Anderson, R.G.W. (2002) *EMBO J.*, **21**, 6387-6396.

Montigiani, S., Neri, G., Neri, P. and Neri, D. (1996) *J Mol Biol.,* **258**, 6-13.

Moore, P. (2003) *Nature*, **426**, 726-731.

Nametkin, S.N., Kolosov, M.I., Ovodov, S.Y., Alexandrov, A.N., Levashov, A.V., Alakhov, V.Y. and Kabanov A.V. (1992) *FEBS,* **309**, 330-332.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) *FEBS Lett.*, **414**, 405-408.

Neri, D., de Lalla, C., Petrul, H., Neri, P., Winter, G. (1995) *Biotechnology*, **13**, 373-377.

Nikoletti, S., Bird, P., Praszkier, J. and Pittard, J. (1988) *J. Bacteriol.,* **170**, 1311-1318.

Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A. and Uhlen, M. (1987) *Protein Eng.*, **1**, 107-113.

Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. and Nygren P.A. (1995) *Protein Eng.*, **8**, 601-608.

Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M. and Nygren, P.A. (1997) *Nat. Biotechnol.*, **15**, 772-777.

Nord, K., Nord, O., Uhlen, M., Kelley, B., Ljungqvist, C. and Nygren, P.A. (2001) *Eur. J. Biochem.*, **268**, 4269-4277.

Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlmann, P.A., Ullman, C., FitzGerald, K. and McGregor, D. (2004) *Proc. Natl Acad. Sci. USA*, **101**, 2806-2810.

Oldfield, C. and Freedman, R.B. (1989) *Eur. J. Biochem.*, **183**, 347-355.

Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G. and Iverson, B.L. (2000) *Nat. Biotechnol.*, **18**, 1071-1074.

Orr, B.A., Carr, L.M., Wittrup, K.D., Roy, E.J. and Kranz, D.M. (2003) *Biotechnol. Prog.*, **19**, 631-638.

Pavlou, A.K. and Reichert, J.M. (2004) *Nat. Biotechnol.*, **22**, 1513-1519.

Plaxco, K.W., Spitzfaden, C., Campbell, I.D. and Dobson, C.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 10703-10706.

Posfai, J., Bhagwat,, A.S., Posfai, G. and Roberts R.J. (1989) *Nucleic Acids Res.*, **17**, 2421-2435.

Praszkier, J. and Pittard, A.J. (1999) *J. Bacteriol.*, **181**, 2765-2772.

Praszkier, J., Murthy, S. and Pittard, A.J. (2000) *J. Bacteriol.*, **182**, 3972-3980.

Rauchenberger, R., Borges, E., Thomassen-Wolf, E., Rom, E., Adar, R., Yaniv, Y., Malka, M., Chumakov, I., Kotzer, S., Resnitzky, D., Knappik, A., Reiffert, S., Prassler, J., Jury, K., Waldherr, D., Bauer, S., Kretzschmar, T., Yayon, A. and Rothe, C. (2003) *J. Biol. Chem.*, **278**, 38194-38205.

Razin, A., Cedar, H. and Riggs, A.D., Editors (1984) *DNA Methylation: Biochemistry and Biological Significance,* Springer Verlag, Berlin.

Reichert, J.M. and Pavlou, A.K. (2004) *Nat. Rev. Drug Discovery*, **3**, 383-384.

Reiersen, H., Lobersli, I., Loset, G.A., Hvattum, E., Simonsen, B., Stacy, S.E., McGregor, D., FitzGerald, K., Welschof, M., Brekke, O.H. and Marvik, O.J. (2005) *Nucleic Acids Res.*, **33**, e10.

Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) *Cell*, **76**, 357-369.

Reiter, Y., Schuck, B., Boyd, L.F. and Plaksin, D. (1999) *J. Mol. Biol.*, **290**, 685-698.

Riechmann, L. (1996) *J. Mol. Biol.*, **259**, 957-969.

Roberts, R.W. and Szostak, J.W. (1997) *Proc. Natl Acad. Sci. USA,* **94**, 12297-12302.

Robinson, K. (2002) *Biopharm Int.* **15**, 20-24.

Rondot, S., Koch, J,. Breitling, F. and Dubel, S. (2001) *Nat. Biotechnol.*, **19**, 75-78.

Rönnmark, J., Gronlund, H., Uhlen, M. and Nygren, P.A. (2002) *Eur. J. Biochem.*, **269**, 2647-2655.

Sandström, K., Xu, Z., Forsberg, G. and Nygren, P.A. (2003) *Protein Eng.*, **16**, 691-697.

Schaffitzel, C., Hanes, J., Jermutus, L. and Plückthun, A. (1999) *J. Immunol. Methods,* **231**, 119-135.

Schlehuber, S., Beste, G. and Skerra, A. (2000) *J. Mol. Biol.*, **297**, 1105-1120.

Schlehuber, S. and Skerra, A. (2005) *Drug Discov. Today*, **10**, 23-33.

Schreuder, M.P., Mooren, A.T., Toschka, H.Y., Verrips, C.T. and Klis, F.M. (1996) *Trends in Biotechnol.*, **14**, 115-120.

Schubert, W.-D., Urbanke, C., Ziehm, T., Beier, V., Machner, M.P., Domann, E., Wehland, J., Chakraborty, T. and Heinz, D.W. (2002) *Cell*, **111**, 825-836.

Sedgewick, S.G. and Smerdon, S.J. (1999) *Trends Biochem. Sci.*, **24**, 311-316.

Sepp, A., Tawfik, D.S. and Griffiths A.D. (2002) *FEBS Lett.,* **532**, 455-458.

Silacci, M., Brack, S., Schirru, G., Marlind, J., Ettorre, A., Merlo, A., Viti, F. and Neri, D. (2005) *Proteomics*, **5**, in press.

Skerra, A., Dreher, M.L. and Winter G. (1991) *Anal. Biochem.*, **196**, 151-155.

Skerra, A. (2000) *Biochim. Biophys. Acta*, **1482**, 337-350.

Smith, G.P. (1985) *Science*, **228**, 1315-1317.

Southern, J.A., Young, D.F., Heaney, F., Baumgartner, W.K. and Randall, R.E. (1991) *J. Gen. Virol.*, **72**, 1551-1557.

Spinelli, S., Frenken, L.G., Burgeois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C. and Tegoni, M. (1996) *Nat. Struct. Biol.*, **3**, 752-757.

Spinelli, S., Frenken, L.G., Hermans, P., Verrips, T., Brown, K., Tegoni, M. and Cambillau C. (2000) *Biochemistry*, **39**, 1217-1222.

Stanfield, R.L., Dooley, H., Flajnik, M.F. and Wilson, I.A. (2004) *Science*, **305**, 1770-1773.

Stemmer, W.P. (1994) *Nature*, **370**, 389-391.

Tabuchi, I., Soramoto, S., Nemoto, N. and Husimi, Y. (2001) *FEBS Lett*., **508**, 309-312.

Takahashi, T.T., Austin, R.J. and Robers, R.W. (2003) *Trends in Biochem. Sci.*, **28**, 159-165.

Tawfik, D.S. and Griffiths, A.D. (1998) *Nat. Biotechnol.,* **16**, 652-656.

Taylor, P., Dornan, J., Carrello, A., Minchin, R.F., Ratajczak, T. and Walkinshaw, M.D. (2001) *StructureI, **9**,* 431-438.

Tomlinson, I.M. (2004) *Nat Biotechnol*., **22**, 521-522.

Uhlen, M., Forsberg, G., Moks, T., Hartmanis, M. and Nilsson, B. (1992) *Curr. Opin. Biotechnol.*, **3**, 363-369.

Verdine, G.L. and Norman, D.P.G. (2003) *Annu Rev. Biochem.*, **72**, 337-366.

Vetter, I.R., Arndt, A., Kutay, U., Görlich, D. and Wittinghofer, A. (1999) *Cell*, **97**, 635-646.

Viti, F., Nilsson, F., Demartis, S., Huber, A. and Neri, D. (2000) *Methods Enzymol.*, **326**, 480-505.

Walsh, G. (2000) *Nat. Biotechnol.*, **18**, 831-833.

Walsh, G. (2003) *Nat. Biotechnol.*, **21**, 865-870.

Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T. and Winter, G. (1989) *Nature*, **341**, 544-546.

Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C. and Kuimelis, R.G. (2002) *Proteomics*, **2**, 48-57.

Wikman, M., Steffen, A.-C., Gunneriusson, E., Tolmachev, V., Adams, G.P., Carlsson, J. and Stahl, S. (2004) *Prot. Eng. Des. Sel.*, **5**, 455-462.

Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 3750-3755.

Winter, G., Griffiths, A.D., Hawkins, R.E. and Hoogenboom, H.R. (1994) *Annu. Rev. Immunol.*, **12**, 433-455.

Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., Weng, S., Wagner, R.W., and Lipovsek, D. (2002) *Chemistry & Biology*, **9**, 933-942.

Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa T. and Yanagawa, H. (2003) *Nucleic Acids Res.*, **31**, e118.

Zardi, L., Carnemolla, B., Siri, A., Petersen, T.E., Paolella, G., Sebastio, G., and Baralle, F.E. (1987) *EMBO J.*, **6**, 2337-2342.

7. Curriculum vitae of Julian Bertschinger

Personal details

Name	Julian Bertschinger
Date of birth	September 15, 1976
Place of birth	Zurich, Switzerland
Nationality	Swiss
Marital status	unmarried
Private address	Winkelstrasse 7 8046 Zürich Switzerland Phone: +41 43 811 31 10 Mobile: +41 76 384 36 53 julian.bertschinger@pharma.ethz.ch

Education

2001-2005	Ph.D. thesis in the group of Professor Dr. Dario Neri, Institute of Pharmaceutical Sciences, ETH Zürich.
1996 - 2001	Studies in Biochemistry at the ETH Zürich, Switzerland final examinations: Biotechnology, Immunology, Biochemistry, Biological Chemistry and Molecular Biology
2000	Diploma thesis: „Functionalisation of the recombinant tumour targeting antibody fragment L19: a versatile coupling site for therapeutic agents" (supervisor: Prof. Dr. D. Neri, Institute of Pharmaceutical Sciences, ETH Zurich). average mark: 5.5 out of 6

1995	Cambridge First Certificate in English; certificate grade A.
1989 – 1996	Gymnasium (Kantonsschule Zürcher Oberland Wetzikon). Matura Typus B (comprising Latin)

Awards and scholarships

2005	Award from ETH Zürich Winner of the contest "ETH Visionen 2030"
2004-2005	Scholarship from Novartis

Languages

German	native speaker
English	fluent
French	good knowledge
Italian	basic knowledge

Scientific publications

Bertschinger, J. and Neri, D. (2004) Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Prot. Eng. Des. Sel.*, **17**, 699-707.

Bertschinger, J., Heinis, C. and Neri, D. Selection for catalysts. *Phage Display in Biotechnology and Drug Discovery*, (ed. Sachdev Sidhu), invited chapter, in press.

Heinis, C., **Bertschinger, J.** and Neri, D. (2003) Calmodulin-tagged phage and two-filter sandwich assays for the identification of enzymatic activities. *Methods Mol Biol.,* **230**, 313-28.

Heinis, C., Huber, A., Demartis, S., **Bertschinger, J.**, Melkko, S., Lozzi, L., Neri, P., Neri D. (2001) Selection of catalytically active biotin ligase and trypsin mutants by phage display. *Protein Eng.*, **14**, 1043-52.

Patents

Bertschinger, J., Heinis, C. (2004) Method for *in vitro* evolution of polypeptides. *International patent application*, **WO 2005/030957 A1**.

Hobbies

Violin
President Akademisches Orchester Zürich (since 2003)
Paragliding
Biking

8. Acknowledgements

Above all, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Dario Neri, who offered me an inspiring and exciting scientific environment for my thesis. He gave the chance to me to work on my own project, which has led to the results presented in this thesis. Despite the academic freedom I enjoyed, Dario never left me alone with my scientific questions and always found some spare time in his extremely busy schedule to discuss my experiments with me and to bring in his ideas and critical questions. I also thank Prof. Dr. Seeberger for having accepted to be my co-examiner, and Barbara Gemperle for her valuable help in the laboratory.

I am most grateful to Christian Heinis, Samu Melkko, Simon Brack and Michela Silacci for discussions, advices and critical input. I enjoyed very much the agreeable atmosphere created by all the members of our research group during my Ph.D. thesis, and I would like to thank them all.

In addition, I would like to thank the Bühler Familienstiftung, which has supported all my education up to this point, and Novartis, which offered me a scholarship for the last year of my Ph.D.

Special thanks go to Murielle Ehrler for her support and care. And last but not least, I am most grateful to my parents who helped me all my life to reach my goals.