Universität für Bodenkultur Wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Paving the way for a binding site selection of Wnt-transcription factors: expression of truncated Lef1 and Lef-β-catenin fusion proteins and the establishment of a method to multimerize the selected binding sites“

Verfasser

Stefan Gabriel, Bakk. techn.

angestrebter akademischer Grad

Diplom Ingenieur

Wien, 2014

Studienkennzahl lt. Studienblatt:	H 066 418
Studienrichtung lt. Studienblatt:	Biotechnologie
Betreuerin / Betreuer:	Wilson Iain B.H., Ao.Univ.Prof. Dr.phil. Czerny Thomas, FH-Prof. Dr.rer.nat

Für meine Eltern

Danksagung

Eine Diplomarbeit schreibt sich nicht allein, weshalb ich mich auf dieser Seite bei allen Unterstützern bedanken möchte. Für die mentale Unterstützung und den fortwährenden Antrieb möchte ich mich bei meiner Schwester Birgit, Ihrem Freund Mathieu und bei meinen Freunden Judith und Lukas bedanken. Besonderer Dank gilt meiner Freundin Cordula, die mir während dieser Zeit konsequent Rückhalt und Kraft gegeben hat.

Diese Arbeit wäre sicherlich nicht in diesen Umfang entstanden, hätte ich mich in dem Labor nicht so wohl gefühlt. Deshalb möchte ich mich ganz herzlich bei Gerlinde, Viktoria und Lisi für die nicht immer nur wissenschaftlichen Ratschläge und das freundliche Arbeitsklima bedanken. Ich bedanke mich auch bei meinem Betreuer Thomas Czerny der mich auf meinem Weg in die Wissenschaft unterstützt hat.

Nicht zuletzt danke ich meinen Eltern, Ernst und Inge, für die Unterstützung und den Rückhalt während des gesamten Studiums, ohne den all dies nicht möglich gewesen wäre.

Abstract

The Wnt signaling pathway is important in development. Misregulation of this pathway leads to various diseases and cancer. To understand the mechanisms regulated by the Wnt pathway, it is crucial to know all of the Wnt target genes. Transcription of these target genes is mainly activated by the interaction of β-catenin with the architectural transcription factors Tcf/Lef. This interaction occurs between the armadillo repeats of β-catenin with the N-terminal β-catenin binding domain of Tcf/Lef. An additional interaction of β-catenin's carboxy-terminal domain with the HMG DNA-binding domain of Tcf/Lef was discovered in 2005. Based on this knowledge, it was hypothesized that this interaction changes the DNA-binding sequence of Tcf/Lef and therefore alters the set of Wnt target genes. To test this hypothesis a PCR based binding site selection experiment was chosen. In this type of in-vitro binding site selection experiments a pool of random oligonucleotides is offered to -previously expressed and purified- proteins. Oligonucleotides containing the specific Tcf/Lef binding site sequence bind to the proteins. Unspecific oligonucleotides are removed and the bound, specific oligonucleotides are subsequently amplified by PCR and are used for further rounds of selection. After several rounds of selection the oligonucleotides are cloned into plasmids and sequenced. In the present thesis the different preparation steps for the binding site selection experiment are described. Two different truncated Lef1 proteins (LefHMG and Lef293) with and without fusion to truncated ß-catenin (LefHMGβcat, Lef293βcat) were expressed in bacteria and cell culture. The bacterial proteins were purified under denaturing or under native conditions. The proteins expressed in mammalian cell culture were extracted from the cell nuclei. The functionality of these proteins was tested by an electrophoretic mobility shift assay (EMSA). The EMSAs showed that only the Lef293 and Lef293βcat proteins, whether expressed in bacterial or cell culture, were able to bind to and shift DNA, whereas LefHMG and its fusion protein did not. This thesis further describes a newly established method for the multimerization of the binding sites obtained from the binding site selection experiment. The concatemers produced by the multimerization method increases the yield of sequenced binding sites approximately tenfold compared to literature.

Zusammenfassung

Der Wnt Signalweg ist ein wichtiger Signaltransduktionsweg in der embryonalen Entwicklung. Fehlregulierungen des Wnt Signalweges führen zu verschiedenen Krankheiten wie Diabetes und Krebs. Um zu verstehen, welche Mechanismen durch den Wnt Signalweg geregelt werden, ist es notwendig alle Zielgene des Signalweges zu kennen. Die Aktivierung dieser Zielgene erfolgt durch die Interaktion von β-catenin mit den Transkriptionsfaktoren Tcf/Lef. Diese Interaktion findet zwischen der Armadillo Domäne von β-catenin und der amino-terminalen Domäne von Tcf/Lef statt. Im Jahr 2005 wurde eine zusätzliche Interaktion zwischen der carboxy-terminalen Domäne von β-catenin und der HMG DNA-Bindungsdomäne von Tcf/Lef entdeckt. Basierend auf dieser Entdeckung wurde die Hypothese formuliert, dass die Interaktion von β-catenin mit der HMG Domäne von Tcf/Lef dessen DNA-Bindungseigenschaften verändert und somit das Setup an Wnt Zielgenen. Um diese Hypothese zu überprüfen wurde ein auf PCR basierendes Binding Site Selection Experiment ausgewählt. Bei diesem in-vitro Binding Site Selection Experiment werden zufällige Oligonukleotide, zuvor exprimierten und gereinigten, Proteinen angeboten. Oligonukleotide die eine spezifische Tcf/Lef Bindungsstelle enthalten binden an die Proteine. Die unspezifischen Oligonukleotide werden entfernt, die gebundenen, spezifischen Oligonukleotide werden anschließend mittels PCR vervielfältigt und für weitere Selektionsrunden eingesetzt. Nach mehreren Selektionsrunden werden die erhaltenen Oligonukleotide in Plasmide kloniert und sequenziert. In der vorliegenden Arbeit werden die verschiedenen Vorbereitungsschritte für das Binding Site Selection Experiment beschrieben. Zwei unterschiedliche, verkürzte Lef1 Proteine (LefHMG und Lef293) mit und ohne Bindung an ein verkürztes β-catenin (LefHMGβcat, Lef293βcat) wurden in bakterieller und Säugetier-Zellkultur exprimiert. Die Proteine aus Bakterien wurden unter natürlichen und denaturierenden Bedingungen aufgereinigt, wobei die in Säugetierzellen exprimierten Proteine aus deren Zellkern extrahiert wurden. Die Funktionalität der Proteine wurde mit einem elektrophoretischen Mobilitäts Shift Assays (EMSA) überprüft. Die EMSAs zeigten, dass nur Lef293 und Lef293βcat in der Lage waren, DNA zu binden und diese zu verlagern. Im Gegensatz dazu konnten LefHMG und sein Fusionsprotein dies nicht. Desweitern beschreibt die vorliegende Arbeit eine neu etablierte Methode um die von dem Binding Site Selection Experiment erhaltenen Bindungsstellen zu multimerisieren. Die durch diese Methode produzierten Concatemere erhöhen die Ausbeute an sequenzierten Bindungsstellen um das zehnfache verglichen mit der Literatur.

Index

List of figures

List of tables

1 Introduction

To understand life, its development and disease, it is necessary to understand the function of the genome. This requires not only the knowledge about the function of individual genes, but also the understanding of the network regulating them. A critical point in the characterization of such regulatory networks is the identification of transcription factors, their binding sites and target genes.

1.1 Regulatory networks / signalling pathways

To react to their surrounding environment, cells established a way to transport signals from the outside of the membrane to the site of action. These signals range from small molecules, for example, changing osmotic conditions[1], to complex signalling peptides, called hormones, emitted by other cells in a multicellular organism. Most of these signals cannot cross the cellular membrane. They bind to receptor proteins on the cell-surface, which in turn generates an intracellular response.

1.1.1 Wnt-signalling pathway

The Wnt signalling pathway is highly conserved throughout the animal kingdom[2]. It has a crucial role in embryonic development and disease. During development Wnt signals control processes like proliferation, cell fate specification, cell polarity and cell migration[3].It is also involved in stem cell renewal, it affects apoptosis and is involved in axon guidance[4,5,6,7,8]. Aberrations of the Wnt signalling pathway contribute to numerous human cancers[9,10,11], type II diabetes[12,13] and osteoporosis[14]. The Wnt signalling pathway is activated by interaction of a Wnt ligand with its membrane-bound receptor. These Wnt ligands are secreted lipid-modified glycoproteins[4]. 19 Wnt ligands can be found in humans and are further divided into 12 subfamilies. Eleven of these Wnt subfamilies are found in Cnidaria, a sea anemone, showing that they are highly conserved through evolution[15,16,17]. This family of Wnt ligands consists of 350-400 amino acid proteins with many conserved cysteine residues, glycosylation sites and several highly charged amino acids.[18,19,20]. The lipid modification of Wnt ligands is important for signalling and secretion. One of these lipid modifications is a palmitoleic acid attached to a conserved serine[21,22,4]. This mono-saturated fatty acid takes part in the interaction of the Wnt ligand and the Frizzled receptor's cysteine rich domain[23]. In humans ten different Frizzled receptors can be found[3]. Frizzleds are seven transmembrane domain receptors with an extracellular cysteine rich domain sufficient to bind Wnts.[23] To activate the Wnt signalling pathway Frizzled needs co-receptors like the low density lipoprotein receptor-related protein 5 and 6 (LRP5/6, arrow in Drosophila) or the tyrosine protein kinases ROR and RYK[24,25]. Whereas LRP5/6, single-membrane-spanning receptor proteins, are the most important co-receptors for the activation of the canonical / β-catenin dependent signalling[26,27]. There are several other transmembrane proteins that act as specific Wnt receptors, which enhance or repress Wnt signalling. For instance the activation of the receptors Zinc/RING finger protein 3 (ZNRF3) and RING finger protein 43 (RNF43), both single pass membrane spanning E3 ubiquitin-protein ligases, lead to endocytosis/degradation and therefore to a turnover of Frizzled[28,29].

Different downstream signalling pathways can be activated by the interaction of the diverse family of Wnt ligands with different heterodimer receptor complexes.

There are at least three Wnt pathways:

- The Wnt/β-catenin pathway (canonical Wnt pathway)

The defining event of the canonical Wnt pathway is the cytoplasmic accumulation of β-catenin, its translocation into the nucleus, and activation of Wnt target genes.

- The non-canonical planar cell polarity (PCP) signalling pathway

Planar cell polarity is needed for the orientation of cells in epithelia according to the body axis. This process takes place in early embryonic development[30]. Frizzled and Dishevelled (Dvl), key proteins in the canonical Wnt pathway are also required in the PCP pathway[31,32]. Other proteins of the canonical signalling pathway downstream of Dvl are not involved in

PCP. In PCP signalling Dvl activates the small GTPase Rho, which in turn activates the c-Jun N-terminal kinase (JNK) pathway and other proteins involved in cytoskeleton rearrangement[33,34]. The distinction between the Frizzled mediated planar cell polarity signalling and the canonical, Wnt/β-catenin signalling is thought to be mediated by receptor context, where LRP5/6 is needed for Wnt/β-catenin signalling but not for the planar cell polarity signalling[27].

- The non-canonical Wnt/Ca^{2+} pathway

Ca^{2+} waves in the Xenopus embryo promote ventral and antagonize dorsal cell fate, in regulating gastrulation movements and regulating the heart development[35,36,37]. In this pathway, Wnt proteins act via cell surface receptors to stimulate an increase in intracellular Ca2+ and the subsequent activation of protein kinase C (PKC). This pathway shares a number of components of the PCP pathway, like Dishevelled and Frizzled. The Frizzled receptors can distinguish between different Wnt ligands, this selectivity seems to be important for the determination of activation of the different downstream pathways[38]. Additionally to this selectivity, ROR2 or Knypek is needed alongside Frizzled as co-receptors to induce the downstream Wnt/Ca^{2+} pathway[39]. Once activated the signal is transduced via heterotrimeric GTP-binding proteins and leads to an intracellular Ca^{2+} release from the endoplasmatic reticulum[40,41,42,43].

The canonical Wnt signalling pathway is the best understood of the above mentioned. Under unstimulated conditions a cytoplasmic β-catenin destruction complex consisting of axin, adenomatous polyposis coli (APC) protein, glycogen synthase kinase 3β (GSK3β) and casein kinase 1 (CK1) is formed. This destruction complex leads to phosphorylation of β-catenin by casein kinase 1α (CK1α) and Glycogen synthase kinase 3 (GSK3)[44,45]. Phosphorylated β-catenin is then ubiquitinated by the E3 ubiquitin ligase βTrCP, and thereby marked for proteasomal degradation[46].

When the canonical Wnt signalling pathway is activated through Wnt ligand binding to Frizzled and its co-receptor LRP5/6, the cytoplasmic tail of LRP5/6 is phosphorylated by GSK3β and by CKIγ[47]. Axin, the scaffold protein that holds the other components of the β-catenin destruction complex in close proximity, is recruited by the phosphorylated cytoplasmic tail of LRP5/6 and is therefore no longer available to form the β-catenin destruction complex[48]. Although this reduced availability of Axin is enough to activate downstream components of the canonical pathway, there are other mechanisms that stabilize cytoplasmic β-catenin. Additionally a second scaffold protein, Dishevelled, is phosphorylated by CKIε and recruited by Frizzled upon Wnt ligand binding[49]. Phosphorylated Dsh and Axin interact and further inhibit the assembly of the destruction complex. Moreover, phosphorylated Dsh recruits GSK3-binding protein Frequently rearranged in advanced T-cell lymphomas (FRAT), which inhibits GSK3β and therefore prevents β-catenin from phosphorylation and degradation. β-catenin accumulates in the cytoplasm and is translocated into the nucleus[50]. A key event in Wnt signalling induction may be the proximity of Fzd and LRP5/6 cytoplasmic domains as indicated by the addition of the intracellular LRP domain to the intracellular domain of Fzd resulting in a constitutively active Wnt receptor[51].

Once in the nucleus, β-catenin binds to a variety of transcription factors. The T cell factor/lymphoid enhancer-binding factor (Tcf/Lef) family of transcription factors is the most important one. Tcf/Lefs bind sequence specific to DNA and in the absence of β-catenin they bind Transducin-like enhancer protein 1 (TLE)/Groucho and histone deacetylases (HDAC) to form a repressive complex and block the transcription of Wnt target genes[52,53]. If β-catenin is present in the nucleus, it replaces TLE1/Groucho directly from its binding site and converts the transcription repressive in a transcription activating complex[54]. The β-catenin/Tcf/Lef complex then recruits proteins that act as co-activators with histone acetyltransferase (HAT) activity, like cAMP-responsive element-binding protein (CREB) binding protein (CBP) or p300 and activates the transcription of Wnt target genes[55]. More than one hundred genes are known to directly dependent on β-catenin and Tcf/Lef transcription factors[56]. These Wnt target genes include genes from a broad range of important cellular functions, necessary for embryonic development and tissue homeostasis. The list of β-catenin/Tcf/Lef dependent

genes includes transcription factors, cell-cycle regulatory proteins, extracellular matrix proteins, cytokines and enzymes[56]. The Wnt signalling pathway is autoregulated, which means that proteins involved in the positive and negative regulatory mechanisms of the pathway are also Wnt target genes[56].

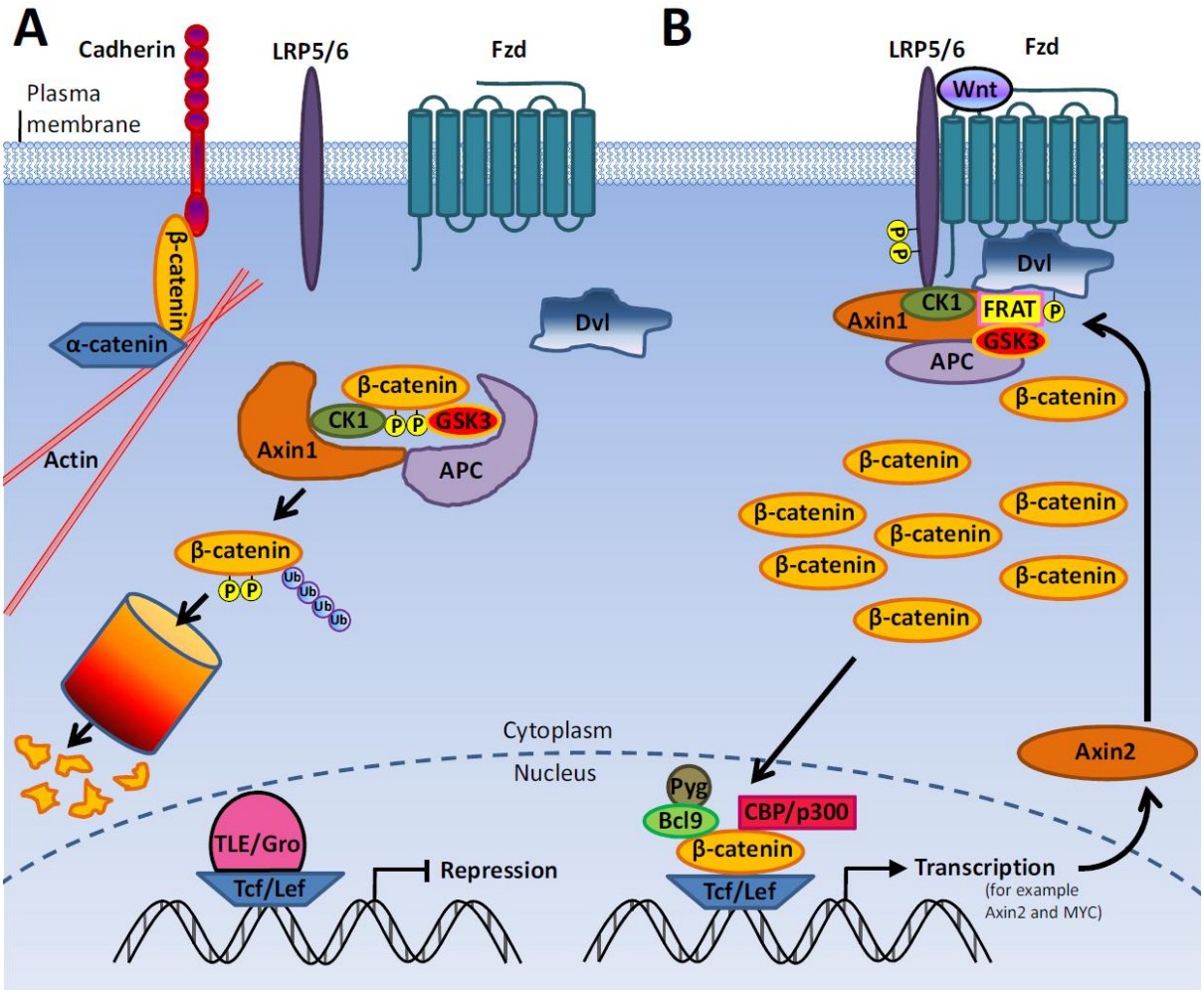


Figure 1: The canonical Wnt signalling pathway. A: In the absence of an extracellular signal the destruction complex consisting of Adenomatous Polyposis Coli (APC), Axin1, Glycogen Synthase Kinase 3 (GSK3) and Casein Kinase 1 (CK1) is formed. It binds and phosphorylates non-cadherin-associated β-catenin. Phosphorylated β-catenin is then ubiquitinated and targeted for proteasomal degradation. B: If an extracellular Wnt ligand binds to the Frizzled (Fzd) and its co-receptor low density lipoprotein receptor related protein 5 or 6 (LRP5/6), a conformational change is induced, LRP5/6 is phosphorylated by GSK3 and CK1 and recruits Axin1. The scaffold protein Dishevelled (Dvl) is phosphorylated by CK1 and bind to Fzd. Activated Dvl and LRP5/6 bound Axin1 disrupt the β-catenin destruction complex and allow cytosolic β-catenin to accumulate. β-catenin translocates to the nucleus binds to T-cell factor/Lymphocyte enhancer-binding factor (Tcf/Lef) transcription factors and displaces the Transducin-like enhancer protein/Groucho (TLE/Gro) transcriptional repressor. The β-catenin/Tcf/Lef complex recruits further transcriptional co-activators like the CREB-binding protein (CBP/p300), the B-cell CLL/Lymphoma 9 protein (Bcl9) and Pygopus (Pyg) to activate the transcription of Wnt target genes. The canonical Wnt signalling pathway is autoregulatory as indicated by the transcription of Axin2, a negative regulator. (adapted from McNeill and Woodgett, 2010[57] ;Baarsma et al., 2013[58], and Bauer et al., 2012[59])

1.1.2 T-cell factor / Lymphocyte enhancer-binding factor (Tcf/Lef)

The family of T cell factor/Lymphoid enhancer-binding factor (Tcf/Lef) proteins belongs to the high mobility group (HMG) box-containing superfamily of transcription factors. Tcf/Lefs bind sequence-specific and have no internal transactivation domain, and are therefore called architectural transcription factors[60]. They act in the canonical Wnt signalling pathway by recruiting β-catenin, which in turn has an internal transcription activation domain but no DNA-binding domain. This β-catenin/Tcf/Lef complex activates Wnt target gene transcription. Tcf/Lefs maintain both repression as well as activation activities. In absence of β-catenin they repress transcription by interaction with TLE1/Groucho, a chromatin repressor that interacts with histone deacetylases (HDACs) to condense DNA and inhibit transcription[52,55,61].

In most vertebrates four Tcf/Lefs, Tcf7 (Tcf1), Tcf7l1 (Tcf3), Tcf7l2 (Tcf4) and Lef1 (Tcf7l3) can be found, and one orthologous form each in flies, worms and hydra[62-67]. Although there is only a hand full of Tcf/Lefs, they generate a great diversity through extensive patterns of alternative splicing and alternative promoter usage[68].

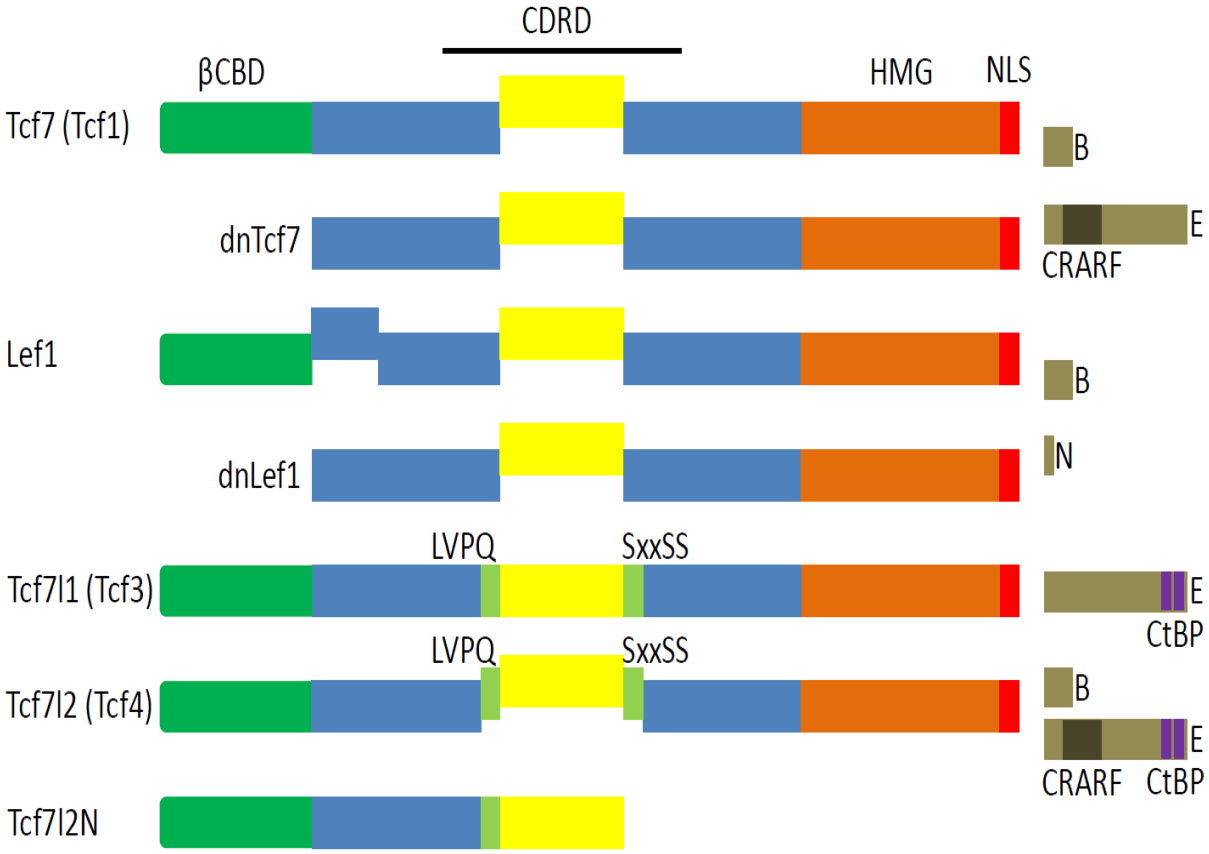


Figure 2: Vertebrate Tcf/Lefs and their diversity created by alternative splicing and promoter usage. In dominant negative isoforms of Tcf7 and Lef1 (dnTcf7 and dnLef1) the β-catenin binding domain (βCBD, green) is missing, making them negative regulators of the Wnt signalling pathway. The βCBD is followed by a context-dependent regulatory domain (CRDR) with an internal exon (yellow), which is alternatively spliced in all forms except in Tcf7l1. In Tcf7l1 and Tcf7l2 this exon is flanked by two short sequences (light green) also created by alternative splicing. In response to TGFβ signalling the CDRD of Lef1 is alternatively spliced[69].Different C-terminal ends are created by alternative splicing and attached to the DNA-binding high mobility group (HMG, orange) box domain and the nuclear localization signal (NLS, red). Only the most common tails are shown, some harboring a conserved CRARF domain and a C-terminal binding protein binding domain. Tcf7l2N, truncated by alternative spicing, lacks the HMG DNA-binding domain. (Image adapted from Arce et al. 2006[68])

Tcf/Lefs contain an N-terminal β-catenin binding domain (βCBD), a context-dependent regulatory domain (CDRD) it the mid part, a high mobility group box (HMG) DNA binding domain, followed by the nuclear localization signal (NLS) and a C-terminal tail dependent on the Tcf/Lef variant and alternative splicing. The CDRD is a domain functioning as a strong transcriptional activator, highly dependent on the context of its binding site within the enhancer[70,71].By the use of alternative promoters a dominant negative (dn), N-terminal truncated form of Lef1 and Tcf7. These dominant negative forms lack the βCBD and are therefore negative regulators by competing for the same DNA binding site, but being not able to bind β-catenin and activate transcription.

Tcf/Lefs recognize DNA sequence specific through a high mobility group box (HMG) domain[72,73]. The Tcf/Lef binding sequences are referred to as Wnt response elements (WREs). Transcription factors containing an HMG domain bind the DNA in at the minor groove and bend the double helix dependent on the nucleotide sequence between 85-130° [74,75]. This bend in DNA probably facilitates recruitment of other distant transcription factors, such as Bcl9 and Pygopus[76]. All Tcf/Lefs bind the DNA as monomers to a nearly identical nucleotide sequence with the consensus 5'-CCTTTGWW-3' (where W indicates A or T) [77,78,79]. Unbound, these transcription factors are unstructured in conformation, but when binding to DNA, they undergo a disorder-to-order transition, where three helices build up an L-shape[80]. The function of Tcf/Lef is tightly regulated in the nucleus by a variety of mechanisms, which include posttranslational modifications like phosphorylation, sumoylation, ubiquitination and acetylation. These modifications contribute to the interaction with other proteins or DNA[81].

1.1.3 Structure of β-catenin

β-catenin has diverse roles in the cell. It is a part of the cytoskeleton in adherens junctions, by linking the extracellular cadherins to the intracellular cytoskeleton via α-catenin[82]. Cytoplasmic β-catenin functions as key mediator in the canonical Wnt signalling pathway. β-catenin consists of 12 copies of an imperfect 42 amino acid motif called Armadillo repeat. Each of these armadillo repeats forms three α-helices (H1, H2 and H3), except repeat 7 which lacks H1[83]. The whole armadillo repeat domain forms a superhelical structure with a positively charged groove which provides the interaction surface for many of β-catenin's interaction partners[83]. These central repeats are flanked by mostly unstructured, flexible amino terminal (NTD) and carboxy terminal domains (CTD). The CTD is linked to the armadillo repeats by a conserved helix termed helix C and has a proposed function to distinguish between β-catenin in cell-cell adhesion or Wnt signalling by folding back and shielding the armadillo repeats[83,84,85]. The carboxy terminal domain harbors a transactivation domain and recruits additional co-activators of transcription. This co-activators comprise the TATA-binding protein (TBP), Brahma/Brahma-related gene-1 (Brg-1), the CREB-binding protein (CBP/p300), and others[86,87,88,89]. A second region with transcriptional activities is located near the N terminus, where B-cell CCL/Lymphoma 9 (BCL9) binds and recruits the nuclear Plant Homeo Domain (PHD)-finger protein Pygopus (Pyg)[90,91,92]. β-catenin does not only interact with transcriptional activators but also with transcriptional repressors like Chibby and Inhibitor of β-catenin and Tcf4 (ICAT)[93,94,95]. ICAT inhibits binding of β-catenin to Tcf/Lef by interacting with the armadillo repeats 5-12 via an N-terminal helical domain and an extended CTD. The C-terminal domain blocks the binding to Tcf/Lef and the N-terminal helices blocks selectively the binding of CBP/p300 to β-catenin[96]. β-catenin contains several sites for the phosphorylation by GSK3 and CK1[44,97]. The different phosphorylation patterns represent alternatives for the regulation of the distinct functions of β-catenin[98].

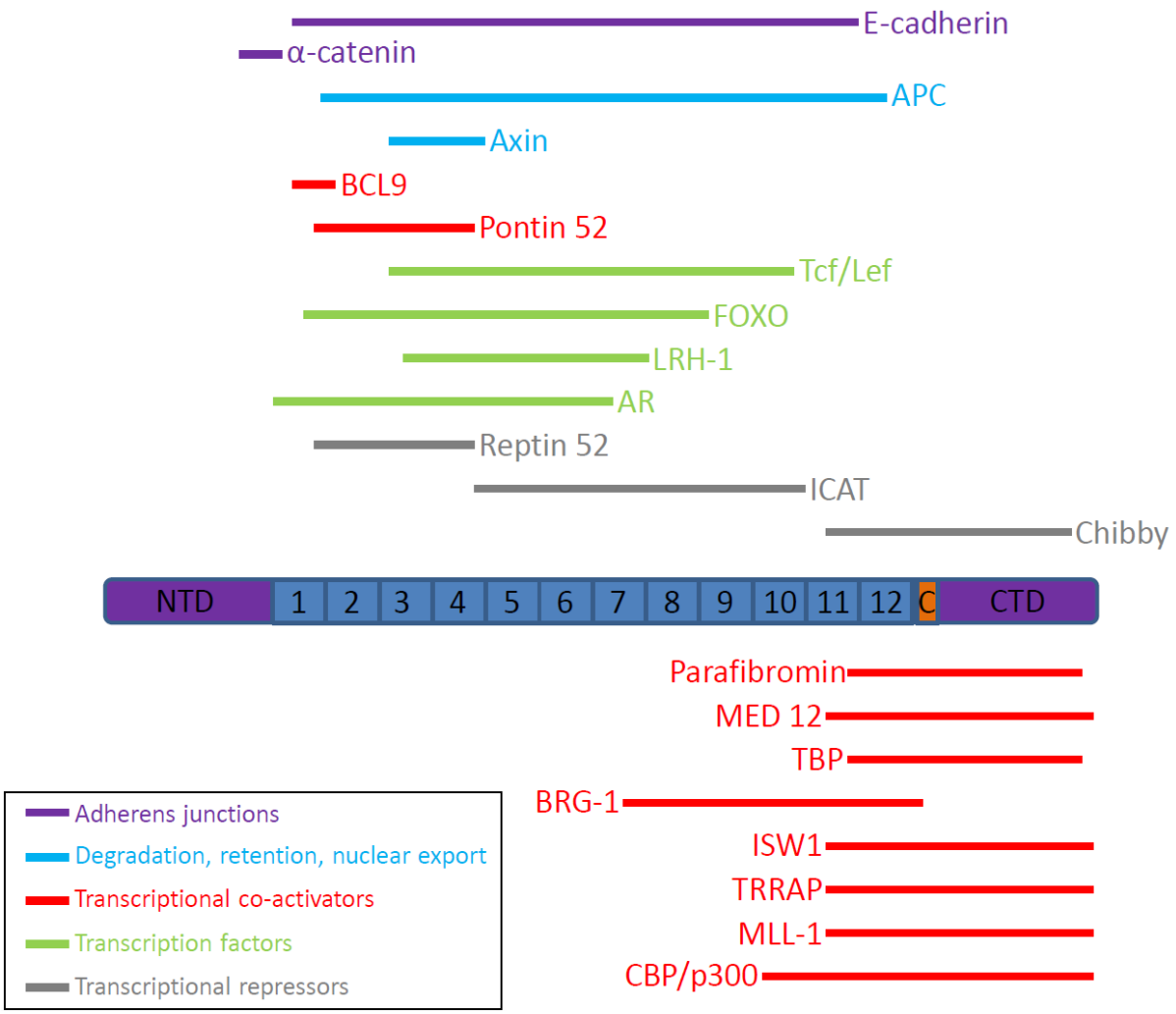


Figure 3: β-catenin and its reaction partners. β-catenin interacts with many different proteins in the cytoplasm, the nucleus and the cytoskeleton. The central β-catenin consists of 12 copies of an imperfect 42 amino acid motif termed Armadillo repeats (numbered boxes 1-12). The Armadillo repeats are flanked by an unstructured, flexible amino terminal (NTD) and carboxy terminal domain (CTD) which is linked to the Armadillo repeats via a conserved helix-C (C). The colored bars show the interaction sites of β-catenin and its partners. Purple bars indicate the interaction site with components in adherens junctions. The components of the destruction complex interaction with β-catenin are shown in blue. Interaction sites for transcription factors providing DNA binding are depicted by green bars. Transcriptional repressors are indicated by bars colored in gray. β-catenin binds a variety of transcriptional co-activators some interact with the Armadillo repeats, but the majority binds to the CTD (red bars). FOXO, Forkhead box protein O; LRH-1, Liver receptor homologue 1;AR, Androgen receptor; ICAT, Inhibitor of β-catenin; MED 12, Mediator of RNA polymerase II transcription subunit 12; TBP, TATA-box binding protein; Brg-1, Brahma/Brahma-related gene-1, also known as SMARCA4, SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 4;TRRAP, Transformation/transcription domain-associated protein; MLL-1, mixed lineage leukaemia; CBP, CREB binding protein. (adapted from Valenta et al., 2012[99])

1.1.4 Tcf/Lef-β-catenin interactions

Tcf/Lef transcription factors contain a highly conserved N-terminal β-catenin binding domain. Binding of β-catenin induces a change in conformation in the first 50 amino acids via formation of an alpha helix and salt bridges in the superhelical interaction groove of the central armadillo repeats 3-10 of β-catenin[100,101]. Although, this is the main interaction site of Tcf/Lef and β-catenin, there exists a second, low affinity binding site between the Tcf/Lef and β-catenin. This low affinity binding site is located N-terminal to the HMG binding domain and may contain a part of the HMG region[54].

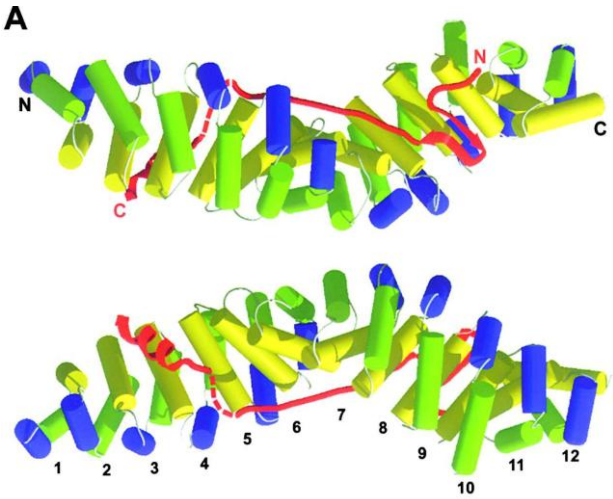


Figure 4: Structure of the β-catenin binding domain of Xenopus Tcf3 in complex with β-catenin armadillo repeats. β-catenin armadillo repeats are formed by 12 copies of the armadillo motif, each containing three helices, except repeat 7, which lacks one helix. The helices are shown in blue, green and yellow. The β-catenin binding domain consists of a β-hairpin module, an extended region containing an α-strand and a β-helix and follows the major axis of the superhelical structure of the armadillo repeats in an antiparallel fashion. (adopted from Graham et al., 2000[100])

1.2 Binding site selection

The identification of transcription factor target genes is critical to the understanding of gene regulatory networks. Therefore, multiple technologies have been developed to determine the binding sites of transcription factors. These include chromatin immunoprecipitation (ChIP)-based methods, like ChIP-on-chip or ChIP-Seq[102,103]. In ChIP-based methods, protein occupied DNA fragments are uncovered. These in vivo methods require high quality antibodies and cells or tissues with abundantly expressed and active transcription factors.

In contrast, there are in vitro methods for the determination of transcription factor binding sites. These allow the identification of directly bound DNA. In these methods the conditions under which the transcription factors are active are not necessarily important, because the purified transcription factor selects its binding site from an initially random pool of oligonucleotides. The selected oligonucleotides are extracted, amplified and subjected to further rounds of selection. Therefore, these methods are called polymerase chain reaction-based methods.

1.2.1 Polymerase Chain Reaction (PCR)-based selection of random oligonucleotides by specific targets

In 1989, a new PCR based method for the selection of nucleic acid molecules that bind specific targets was introduced by Oliphant, Brandl and Struhl in their publication "Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein"[104]. In august 1990 Ellington and Szostak published "In vitro selection of RNA molecules that bind specific ligands" in Nature[105]. In the same month Tuerk and Gold published "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase" in Science and called their method SELEX[106]. Both groups developed oligonucleotide aptamers, short single-stranded nucleic acid oligomers, which bind to their target with high selectivity and sensitivity because of their three-dimensional shape[105,106]. Both described nearly the same method[105,106,107]. In the same year Pollock and Treisman published "A sensitive method for the determination of protein-DNA binding specificities" where they used the same PCR based method for a binding site selection of transcription factors[108]. Two years later in 1992 this method was described by Wright, Binder and Funk as Cyclic Amplification and Selection of Targets in their publication "Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site"[109]. Since then, many of different methods were developed, each optimized for its purpose or target[107].

Although the names and purposes of these methods vary, the principle in all these methods is the same. All of them rely on mechanisms accredited to evolution, which are variation,

selection and replication. A pool of oligonucleotides containing a completely randomized stretch, flanked by PCR primer sequences, is used for the selection of the binding to a target[107]. These target molecules range from large targets like proteins over peptides to organic small molecules or even metal ions[107]. After selection, the bound oligonucleotides are recovered by immunoprecipitation or affinity purification of the target[108,110,111,112]. The selected oligonucleotides are then subjected to PCR amplification. The new pool of oligonucleotides is then used for further rounds of selection. Subsequent rounds of selection increase the affinity of the selected oligonucleotide to its target[105].

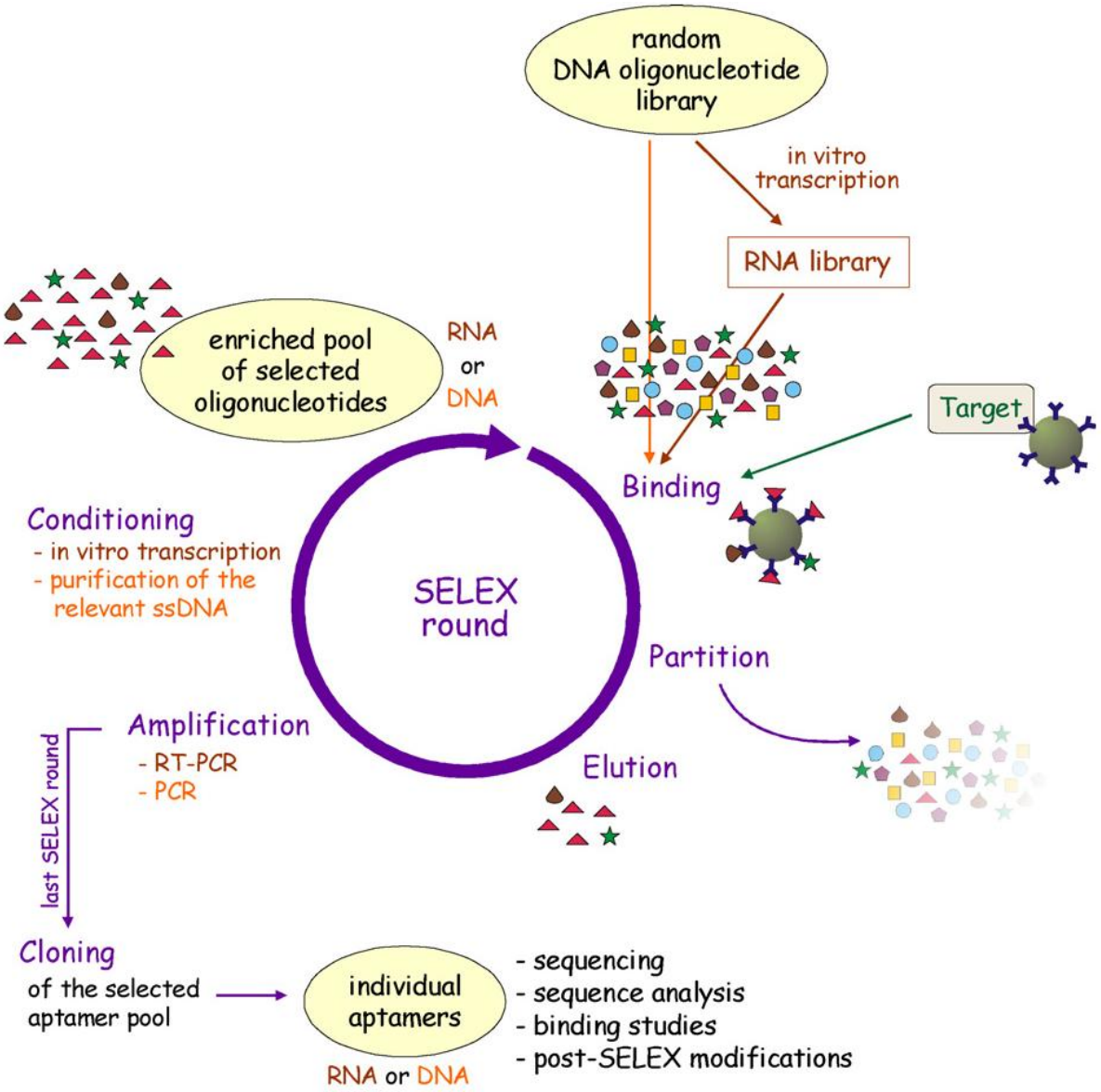


Figure 5: Schematic representation of the SELEX process. A random single strand (ss) DNA library is used as starting point for the selection. Depending on the purpose of the selection the random DNA pool can be converted to double stranded (ds) DNA or RNA prior selection. In the binding reaction the oligonucleotides bind specifically to the target. The bound oligonucleotides are then separated from the unbound, eluted and amplified by PCR. In a conditioning step, the amplified selected oligonucleotides are prepared for the use in the next round of selection depending on if they are used as ssDNA, dsDNA or RNA. This enriched pool of selected oligonucleotides is then subjected to another round of selection. After the final selection round the selected oligonucleotides are cloned into a plasmid vector and then sequenced. The obtained sequences are then analyzed and further binding studies are conducted. (from Stoltenburg et al. 2012)[107]

1.2.1.1 PCR-based selection of random oligonucleotides by using recombinant proteins

For the determination of binding sites of transcription factors the PCR-based method was optimized to suit this purpose. The random oligonucleotide pool consists of double stranded oligonucleotides containing the random nucleotide stretch with lengths between 14 and 35 bp[79,107,109-113]. To reduce unspecific binding of DNA to the transcription factors nonspecific competitors, like salmon sperm or poly(deoxyinosinic-deoxycytidylic) acid (poly(dI-dC)) may be added to the binding reaction[116]. There are different ways to separate the bound oligonucleotides from the remaining pool of unspecific oligonucleotides. With a previously labeled selection oligonucleotide, an electrophoretic mobility shift assay can be performed, the shifted band, containing the protein and bound oligonucleotide can be excised and the DNA eluted[113]. The transcription factors with bound DNA can also be immunoprecipitated with antibodies or the transcription factors contain tags for affinity purification and are separated by binding to affinity purification beads and washing[79,109,111,114,115]. Depending on the way of separation, the selected oligonucleotides need to be eluted before they are amplified. If they are bound to beads via antibody or affinity tag there is no need for elution, because in the first denaturing step in the PCR process the transcription factor is denatured and releases the selected oligonucleotide[79,109,111,114,115]. Because the product of the PCR is a double stranded oligonucleotide, the conditioning of the amplified selected oligonucleotides is reduced to a purification reaction to get rid of the remaining PCR primers[79,109,111,114,115]. After four to six rounds of selection the selected oligonucleotides are cloned into a plasmid vector and the binding sites are sequenced and analysed[79,109,111,114,115].

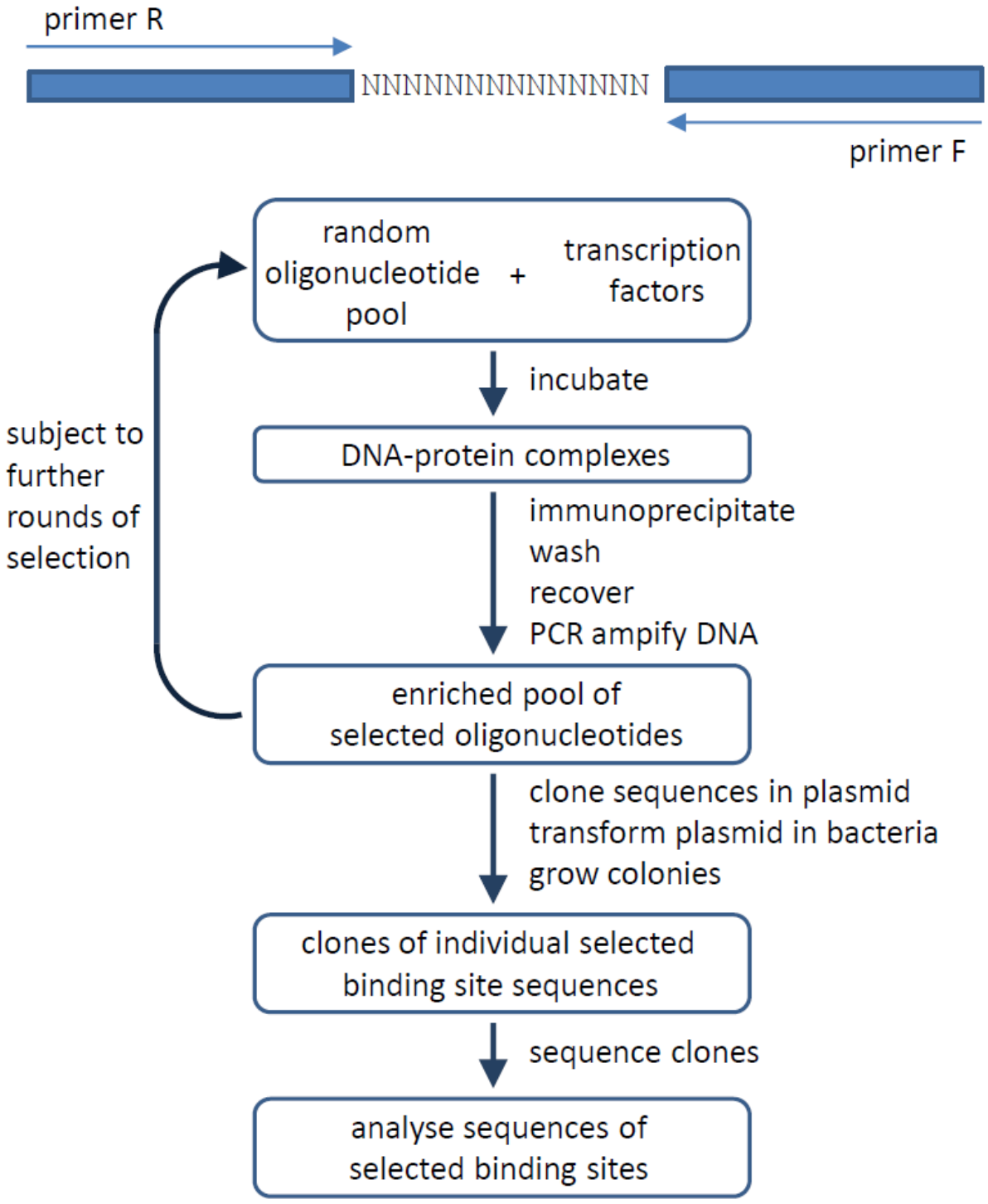


Figure 6: Flowchart of the PCR-based binding site selection of a transcription factor. At the top is a schematic representation of an oligonucleotide used for a binding site selection, consisting of a random nucleotide part flanked by two PCR-primer sequences.

1.2.1.2 Considerations about the random nucleotide site

Some consideration about the length of the random nucleotide stretch should be made. Considering a binding site selection for Lef1 with an b=8 base pair consensus sequence (CCTTTGAA) with an oligonucleotide with a random nucleotide stretch of l=14 bp[77,78,79]. As shown in Figure 7 there are l-b+1=7 possible positions for the 8 bp binding site of Lef1 in the 14 bp random nucleotide stretch.

```
5' - NNNNNNNNNNNNNN - 3'

1     CCTTTGAANNNNNN
2     NCCTTTGAANNNNN
3     NNCCTTTGAANNNN
4     NNNCCTTTGAANNN
5     NNNNCCTTTGAANN
6     NNNNNCCTTTGAAN
7     NNNNNNCCTTTGAA
8     TTCAAAGGNNNNNN
9     NTTCAAAGGNNNNN
10    NNTTCAAAGGNNNN
11    NNNTTCAAAGGNNN
12    NNNNTTCAAAGGNN
13    NNNNNTTCAAAGGN
14    NNNNNNTTCAAAGG
```

Figure 7: Possible binding sites of Lef1 in a random oligonucleotide with the length of 14 bp.

This circumstance complicates the analysis of the sequences of the selected binding sites. To further complicate this, there are two different possibilities to describe the same binding site depending on the orientation of the transcription factor and the DNA strand which is sequenced. Meaning the Lef1 binding site, CCTTTGAA, can also be described as TTCAAAGG. This fact should be considered in the analysis of the obtained sequences. Every base pair the random part exceeds the length of the binding site, complicates the subsequent analysis. For this reason the random nucleotide part should be kept as small as possible.

2 Aim of the study

The aim of this thesis was to establish the necessary tools for a binding site selection of Lef1 and Lef1/β-catenin fusion proteins.

Based on knowledge of the second binding site between β-catenin and Lef1, which interacts with the HMG domain and the fact, that β-catenin binds to Lef1 even if Lef1 is not bound to DNA, a hypothesis was formulated: Opposed to the general believe, that all of Lef1 is already bound to its binding site on DNA and represses the transcription, a portion of free, unbound Lef1 exists in the nucleus and is also bound to β-catenin. Upon β-catenin binding at the main interaction site, the C-terminal domain of β-catenin interacts with the HMG domain of Lef1 and changes its binding site affinity. This Lef1/β-catenin complex may then bind the DNA at different binding sites as Lef1 alone and activates a different set of target genes.

To prove this hypothesis, reporter assay experiments in cell culture were conducted by members of Dr. Czerny's lab with promising results. To further prove this theory a binding site experiment was designed, where the binding sites of Lef1 and Lef1 bound to β-catenin should be compared.

Part of the aim of this thesis was to prepare functional proteins expressed in E. coli and mammalian cell culture (HeLa, COP-8) in sufficient concentration, amount and purity. Furthermore, the establishment of a method for the multimerization of the selected binding sites was needed. In the normal binding site selection experiments the amplified oligonucleotide pool of selected binding sites is then cloned into a plasmid vector, transformed into E. coli and individual clones are sequenced. This method yields only the sequence of one selected binding site per sequenced clone. To increase the yield of sequencing, the aim of the present study was to establish a method to multimerize the binding sites before cloning them into the plasmid vectors.

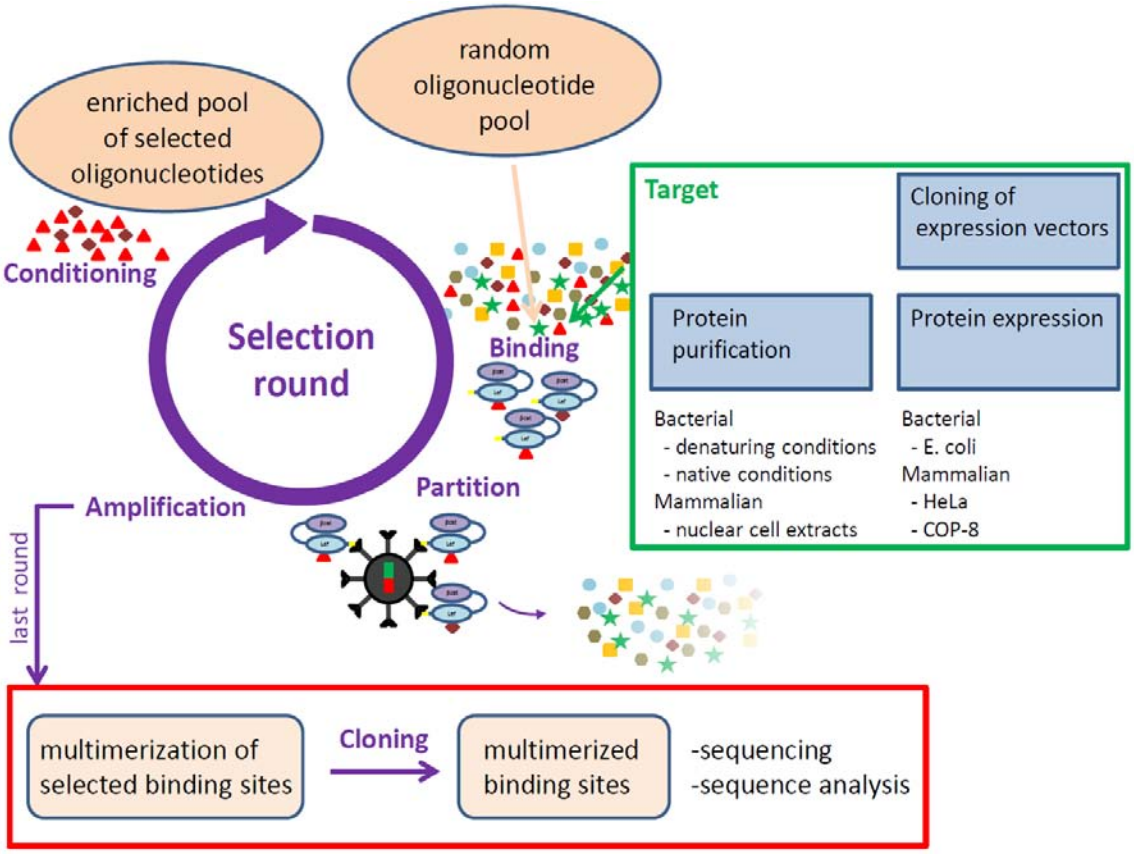


Figure 8: Schematic representation of the planned binding site selection experiment with the aim of this thesis framed in green and red rectangles. The first part of the thesis comprises the preparation of the target proteins, in this case the LefHMG/293 and LefHMG/293 ß-catenin fusion proteins. To achieve this, the expression vectors had to be cloned for the expression in bacterial and mammalian hosts. The proteins had to be expressed and purified. Mammalian proteins were extracted from the nucleus and bacterial proteins were purified under native and denaturing conditions. The second part of the thesis was to establish a method for the multimerization of the binding sites once selected. Therefore different strategies were used, like the simultaneous digestion and ligation of oligonucleotide.

3 Material und Methods

3.1 Enzymes

Restriction enzymes	Restriction site
Cla / Bsu15I	5'...A T↓C G A T...3' 3'...T A G C↑T A...5'
NheI	5'...G↓C T A G C...3' 3'...C G A T C↑G...5'
PstI	5'...C T G C A↓G...3' 3'...G↑A C G T C...5'
SmaI	5'...C C C↓G G G...3' 3'...G G G↑C C C...5'
XbaI	5'...T↓C T A G A...3' 3'...A G A T C↑T...5'
XhoI	5'...C↓T C G A G...3' 3'...G A G C T↑C...5'
HindIII	5'...A↓A G C T T...3' 3'...T T C G A↑A...5'
BamHI	5'...G↓G A T C C...3' 3'...C C T A G↑G...5'
EcoRI	5'...G↓A A T T C...3' 3'...C T T A A↑G...5'
SalI	5'...G↓T C G A C...3' 3'...C A G C T↑G...5'
Other DNA modifying enzymes	
T4 DNA Ligase	
FastAP Thermosensitive Alkaline Phosphatase	
Phusion® High-Fidelity DNA Polymerase	

Table 1: List of enzymes. All enzymes were purchased from Thermo Scientific. All enzymatic reactions were performed in the optimal buffers, provided by the manufacturer.

3.2 Cloning of Plasmid-Vectors

3.2.1 General procedure

For the generation of new plasmid vectors, a plasmid containing the designated backbone (called vector) and a plasmid containing the desired DNA sequence (called insert) were digested over night with appropriate restriction enzymes (Thermo), generating corresponding types of overhangs (Table 1). For a partial digestion, the plasmids were treated with 0.1, 0.3 and 1 U restriction enzyme for one hour at 37°C. The backbone vector DNA was then treated with alkaline phosphatase (FastAP, Thermo) to prevent self-ligation of the backbone. After heat-inactivation of the alkaline phosphatase the vector DNA was purified by using the MSB® Spin PCRapace Kit (INVITEK) according to the manufacturer's protocol. The restriction digested insert DNA was resolved in a 1% agarose gel. The fragment with the correct size

was excised and the DNA was purified using the Invisorb® Spin DNA Extraction kit (INVITEK). After purification, the amount and proper size of the vector and insert DNA was verified by agarose gel electrophoresis. The vector and insert DNA were ligated by using T4 DNA Ligase (Thermo). A ratio of 1:3, vector: insert, was used. The ligation was carried out at room temperature over night.

3.2.2 Transformation of E. coli Top10F'

After ligation the newly constructed plasmid were transformed into a $CaCl_2$-competent E. coli Top10F' (Invitrogen). Therefore the ligation reaction was incubated with the E. coli cells for 10 minutes on ice and subsequent transformation using a 42°C heat shock for 90 seconds. After 2 minutes on ice the bacteria were allowed to grow in LB-Medium (10 g/L Pepton (Roth), 5 g/L yeast extract, 5 g/L NaCl) for half an hour. Consecutively the E. coli cells were spun down (5000 rpm, 4 min, RT), resuspended in 100 µL LB-Medium and plated on LB-agar plates (LB medium + 1.5 % agar) containing 100 µg/mL Ampicillin. Optionally 10 µL IPTG (100µM, Thermo Scientific) and 10µL X-Gal (100µM, Thermo Scientific) for alpha complementation screening was added to the bacteria suspension directly before plating. The plates were incubated at 37°C overnight.

E. coli TOP10F' (Invitrogen) genotype:

F'[lacI^q Tn10(tet^R)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ

3.2.3 Miniprep: for small scale plasmid preparation

For the screening of colonies, containing the correct plasmid, a small scale DNA preparation and purification using a rapid alkaline extraction method, was carried out. Therefore, selected colonies, normally ten to twenty, were incubated in 2 mL LB-medium containing 100 µg/mL ampicillin (LB-Amp medium) overnight. The E. coli cells were harvested by centrifugation and resuspended in resuspension buffer P1 (50 mM Tris, 10 mM EDTA, 100 µg/mL RNase A). Bacteria were lysed in buffer P2 (0.2 M NaOH, 1% SDS) and the reaction was stopped by addition of neutralization buffer P3 (3 M KAc, 11.5 % glacial acetic acid). Cellular residues, denatured proteins and cross-linked, denatured chromosomal DNA was removed by centrifugation (14000 rpm, 20 min, RT). The plasmid DNA was precipitated in a 12 % PEG solution; washed with ethanol (70 %) and resuspended in 20 µL dH_2O. 2µL of DNA was used for restriction digestion analysis.

3.2.4 Midiprep: for mid-scale plasmid preparation

After positive restriction digestion analysis of the Miniprep one colony was inoculated in 75 mL LB-Amp medium and incubated at 37°C over night. The cells were harvested by centrifugation and the plasmid DNA was extracted and purified using the JETSTAR 2.0 Midi Kit (Genomed) according to the manufacturer's protocol except: only the half of the recommended volume for cell lysis and wash steps were used. To determine DNA yield, OD_{260} measurement and consecutive restriction digestion analysis were conducted.

3.2.5 Agarose gel electrophoresis

To check the correct size of digestion-fragments of plasmids, DNA was separated by agarose gel electrophoresis. For fragment sizes ranging from 200 bp to 2000 bp a gel containing 1% agarose (BIOZYME) in SB-buffer (2 M NaOH, 7.3 M Boric Acid) was used. For fragments smaller than 200 bp a higher concentration of 2% agarose was used. For the determination of the fragment size a mixture of two DNA ladders (100 bp ladder + λ BstEII (Thermo)) denoted as λ ladder were also loaded onto gels. The gels were run at 120 V for 20 minutes as standard settings. For visualizing the DNA fragments the intercalating dye ethidium bromide (10 µg/mL in SB-buffer) was used. The gels were exposed to UV light and pictures were taken by an Alphaimager® Mini (Cell Biosciences).

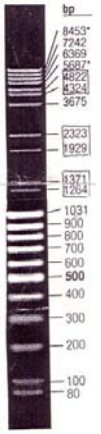


Figure 9: Band pattern of the λ ladder used to determine fragment lengths of DNA on agarose gels.

3.2.6 Plasmids

3.2.6.1 Plasmids for protein expression in E. coli

For protein expression in E. coli the target protein sequences were cloned into the pET expression vector which contains a T7 promoter under the control of a lac operator. This makes the protein expression inducible upon addition of IPTG. This vector also contains an artificial ATG-translation start codon, followed by a 6x histidine tag (His-tag) and the multiple cloning site. It confers resistance to Ampicillin and harbors a ColEI origin of replication.

The plasmids were generated by cloning the target protein sequences form preexisting plasmids into the pET backbone by using the restriction enzymes Bsu15I and XhoI.

pET LefHMG

pET LefHMGmβcat(521-781C)

pET Lef293m

pET Lef293mβcat(521-781C)

3.2.6.2 Plasmids for protein expression in HeLa cells

To use animal cells as expression hosts, the target proteins had to be cloned into a vector suited for this task. The pMC vector contains a simian cytomegalovirus (CMV) promoter and a simian virus 40 (SV40) polyadenylation site. For subsequent protein purification using a His-tag, the target protein sequence, including the 6x Histidine tag from the pET expression vector was cloned into the pMC backbone. Therefore the insert was obtained by digestion of the pET vector with XbaI and XhoI restriction enzymes and the final plasmid was generated by ligation of the insert into the NheI, XhoI restriction sites of the pMC vector. By ligating the compatible ends of the XbaI and NheI restriction sites, both sites are destructed.

pMC LefHMG 6His

pMC LefHMGmβcat(521-781C) 6His

pMC Lef293m 6His

pMC Lef293mβcat(521-781C) 6His

3.2.6.3 Plasmids for protein expression in COP-8 cells

To enhance protein expression, the target protein sequences were cloned into the pKC backbone. It contains a human CMV promoter, a SV40 polyadenylation site and also a polyoma origin of replication. Transfected into a polyoma virus transformed cell line, like COP-8, the plasmid gets amplified and therefore a higher protein yield is achieved. This

plasmid was constructed by excision of the target protein sequence, including the His-tag form the pMC vector and ligation into the pKC backbone by using SmaI and XhoI restriction enzymes. The insert and vector DNA were only partially digested by SmaI.

pKC LefHMG 6His
pKC LefHMGmβcat(521-781C) 6His
pKC Lef293m 6His
pKC Lef293mβcat(521-781C) 6His

3.3 Protein expression and purification

3.3.1 E. coli as expression system

3.3.1.1 Transformation of E. coli BL21(DE3) and BL21(DE3)RIL

The transformation procedure was the same, as described in 3.2.2, except that the transformed E. coli BL21(DE3)RIL (Agilent) were plated on LB-agar plates containing 100 µg/mL Ampicillin and 34 µg/mL Chloramphenicol for additional selection pressure. E. coli BL21(DE3) was cultivated on LB-agar plates containing only 100 µg/mL Ampicillin.

E. coli BL21(DE3) (Agilent) genotype:
E. coli B F^-, dcm, *ompT*, hsdS(r_B^-,m_B^-), *gal*, λ(DE3)

E. coli BL21(DE3)RIL (Agilent) genotype:
E. coli B F^-, *ompT*, hsdS(r_B^-,m_B^-), *dcm*$^+$, *Tet*^{r}, *gal*, λ(DE3), *end*A, *Hte*, [*argU ileY leuW*Cam^r]

3.3.1.2 Inoculation of an expression culture and induction of protein expression

One colony was inoculated in 3 mL LB-medium in a 15 mL culture tube and was incubated overnight in a shaking incubator at 37°C with 140 rpm. One milliliter of the overnight culture was diluted 1:4 in LB-Medium and the OD_{600} was measured (usually in the range of 3.0-3.5 in the undiluted sample).For proteins that were purified under native conditions 0.75 mL of a OD_{600} = 3.0 bacteria suspension was inoculated in 75 mL prewarmed LB-medium in a 1 L Erlenmeyer flask and incubated at 37°C, 250 rpm. For expression culture for protein purification under denaturing conditions 500 mL LB media were inoculated accordingly. Upon an OD_{600} of 0.6 the protein expression was induced by adding IPTG to a final concentration of 100 µM or 1 mM IPTG. In later experiments the incubation temperature was lowered to 25°C during protein expression. After protein expression the OD_{600} was measured and the bacteria were harvested by centrifugation (4000 rpm, 10 min, 4°C). The supernatant was discarded and the pellet was frozen at -80°C.

3.3.1.3 Cell lysis by sonication

The pelleted bacteria, for purification under native conditions, stored at -80°C were thawed on ice and resuspended in precooled Lysis/Binding Buffer (50mM Tris pH 7.5, 50mM NaCl) according to: the bacteria contained in 10 mL expression culture with an OD_{600} of 0.6 were resuspended in 0.5 mL Lysis/Binding Buffer. 2 mL aliquots in 15 mL Falcon tubes were made. The frozen pellet from the 500 mL expression culture, for purification under denaturing conditions was also thawed and resuspended in 40 mL Buffer A (6 M Guanidine HCl, 100 mM NaH_2PO_4 pH 8.0, 20 mM β-mercaptoethanol). The bacteria suspensions were sonicated six times with 50% power, 50% continuity for 10 seconds with one minute interval. The samples were cooled on ice between the sonications. For protein purification under native conditions the cell debris was removed by centrifugation (4500 rpm, 30 min, 4°C) and clear cell extract were immediately processed or frozen at -80°C.

3.3.1.4 Purification under denaturing conditions

The bacteria cell slurry from sonication was stirred for 1 h on a magnetic stirrer. The cell debris was then removed by centrifugation at full speed for 15 minutes at 4°C. The supernatant was collected and 1 mL Ni-NTA agarose (Qiagen), preequilibrated in Buffer A, was added. The suspension was stirred for 3 h at room temperature. The Ni-NTA agarose was spun down for 5 min at 2000 rpm, the supernatant was discarded and the Ni-NTA washed in 20 mL Buffer B (8 M Urea, 100 mM NaH_2PO_4 pH 8.0, 10 mM Tris, 20 mM Imidazole, 20 mM β-Mercaptoethanol) and finally resuspended in 3 mL Buffer B. the Ni-NTA slurry was loaded in a gravity flow column (Chromabond Macherey-Nagel) and washed with 15 mL Buffer B. The proteins were eluted in 10 mL Buffer B with an Imidazole gradient ranging from 25 mM to 250 mM. 16 fractions were collected and analyzed by SDS-PAGE for their protein content. Protein containing fractions were pooled and processed immediately or stored at 4°C.

3.3.1.4.1 Dialysis and concentration with spin columns

The pooled fractions of one protein were then loaded with a syringe into a dialysis chamber (Slide-A-Lyzer Dialysis Cassette, Thermo Scientific)) with a cutoff of 10 kDa for the LefHMG and Lef293 proteins and 50 kDa for the fusion proteins LefHMGβcat and Lef293βcat. The dialysis buffer (10 mM Tris pH 8.0, 50 mM NaCl, 10% glycerol, 1 mM β-mercaptoethanol, 3 mM Urea) was changed every eight hours. With every buffer change the urea concentration was lowered at 1 M steps to a final concentration of 0 M.

After dialysis the protein solution was loaded into spin filters (Amicon Ultra, Merck Millipore, MWCO: 10kDa) and were centrifuged at 4000 rpm for 10-15 min at 4°C. The proteins were stored at 4°C.

3.3.1.5 Purification under native conditions

The cell extract was thawed on ice, meanwhile 100 µL of His Mag Sepharose™ Ni (GE Healthcare) were pre-equilibrated in 500 µL Binding Buffer (50 mM Tris pH 7.5, 50 mM NaCl, 5 mM Imidazole). Imidazole was added to the cell extract to a final concentration of 5 mM and was then incubated with the equilibrated beads for 30 minutes at 4°C with 850 rpm shaking. To get rid of loosely bound, nonspecific proteins the magnetic beads were washed twice in 1 mL Wash Buffer (50 mM Tris pH 7.5, 50 mM NaCl, 20 mM Imidazole). The beads with bound protein were stored in Storage Buffer (50mM Tris pH 7.5, 50 mM NaCl) at +4°C in the fridge. For quantification by using the Bradford assay and SDS-Page protein was eluted from the beads in Elution Buffer (50 mM Tris pH 7.5, 50 mM NaCl, 500 mM Imidazole).

3.3.1.6 SDS-PAGE

For protein analysis on a SDS-Page a BioRad Protean II minigel apparatus was used. A 10% running and a 5% stacking gel was casted. The samples were mixed with 2x Lämmli Buffer (120 mM Tris pH 6.8, 20% Glycerol, 4 % SDS, 0.02% bromophenole blue), heated to 95°C for 5 minutes and centrifuged for 2 minutes at 14000 rpm. The samples and 5 µL PageRuler Prestained Protein Ladder (Thermo) as marker were loaded and the gels were initially run at 80 V in 1x Running Buffer (25mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). After the proteins have staked at the stacking / running gel border the voltage was increased to 200 V and the gel was run until the bromophenole blue band left the gel. The gel was stained with Coomassie Brilliant Blue G250 and pictures were taken with a FluorChem E (ProteinSimple)

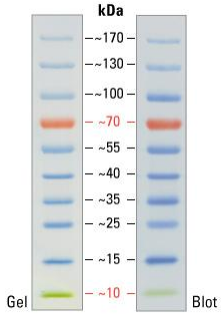


Figure 10: Band profile of the PageRuler Prestained Protein Ladder (Thermo Scientific)[117] used for the size determination of Proteins by SDS-PAGE

	Running gel	Stacking gel
Water	6.25 mL	4.25 mL
Acrylamide (Roth)	4 mL	935 µL
3M Tris pH 8.8	1.5 mL	------
3M Tris pH 6.8	------	225 µL
10% SDS	120 µL	55 µL
10% APS	120 µL	28 µL
TEMED	2,5 µL	2.8 µL
Total	11.9925 mL	5.4958 mL

Table 2: Recipe for SDS-polyacrylamide gels

3.3.1.7 Bradford assay for the determination protein concentration

Aliquots of the sample were diluted with water to a volume of 800 µL. 200 µL of 5x Bradford reagent was added to each tube, vigorously vortexed and incubated for 15-30 minutes at room temperature. The samples were measured with a photometer (Genesis 6, Thermo Scientific) at a wavelength of 595 nm. For each measurement a new standard curve was made. Therefore BSA was diluted in 800µL water to final concentrations of 1, 2, 4, 6, 8 and 10 µg/mL. If the sample was contained in a buffer, this buffer was also added to the blank and standards to ensure comparable results.

3.3.2 Mammalian cells as expression system

3.3.2.1 Cell lines

3.3.2.1.1 HeLa (ATCC: CCL-2)

Human epithelial cervix adenocarcinoma cell line (female)

3.3.2.1.2 COP-8

Polyoma virus transformed mouse fibroblasts

3.3.2.2 Passaging of cells

The cells were cultured in an incubator at 37°C, 5% CO_2 and saturated H_2O atmosphere in a 75 cm^2 cell culture flask with high glucose Dulbecco's Modified Eagle Medium (PAA) (DMEM) with 10% fetal calf serum (PAA), 100°U/mL penicillin and 100°U/mL streptomycin sulfate (1xPenicillin/Streptomycin PAA). Every two to three days, when a confluency of 100% was reached the cells were washed with PBS followed by trypsinization for 5 minutes at 37°C. The detached cells were resuspended in fresh medium and then sub-cultured at a rate of 1:10-1:15.

3.3.2.3 Transfection

For transient transfection experiments 1x 10^6 HeLa cells or 1.2x 10^6 COP-8 cells per 10 cm dish were seeded prior to transfection. 10 µg DNA (2,5µg plasmid plus 7.5 µg backbone vector pBluescript) was prepared in 1.5 mL of serum free DMEM and mixed with another 1.5 mL serum free DMEM containing 20 µL Turbofect (Thermo) transfection reagent. This mix was incubated at room temperature for 30 minutes. Subsequently the medium was removed from the 10 cm dishes and replaced by the 3 mL transfection mixture. The cells were incubated at 37°C for two hours. The transfection was stopped by the addition of 10 mL DMEM containing 10% FCS and the cells were incubated for another 24 hours before they were harvested.

3.3.2.4 Preparation of nuclear cell extracts

All steps were performed on ice and the buffers were precooled. The 10 cm cell culture dishes were taken from the incubator and immediately placed on ice and the medium was sucked off. The cells were washed twice with ice cold PBS and then scraped in 1mL ice cold PBS. The cell clumps were transferred into a 2 mL reaction tube and centrifuged (2800 rpm, 7 min, 4°C). The supernatant was discarded and the pelleted cells were resuspended in 400 µL Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA; 1 mM DTT, 1x Proteinase Inhibitor EDTA-free (Roche) and 0,5 mM PMSF were added directly before use)) and lysed on ice for 15 minutes. After lysis 25 µL of 10 % NP-40 Substitute was added and vigorously vortexed for 10 seconds. The nuclei were spun down in a centrifuge (14000 rpm, 10 sec, 4°C), the supernatant was discarded and the nuclei were lysed by the addition of 50 µL Buffer C (20 mM HEPES pH 7.9, 400 mM NaCl, 1 mM EDTA; 1 mM DTT, 1x Proteinase Inhibitor EDTA-free (Roche) and 0.5 mM PMSF were added directly before use) and 15 minutes of vigorous shaking on a rocking platform at 4°C. To remove the cell debris the extracts were centrifuged (14000 rpm, 2 min, 4°C) and aliquots from the clear extract were frozen at -80°C.

3.3.3 Radioactive electrophoretic mobility shift assay (EMSA)

The radioactive EMSA was performed by Dr. Thomas Czerny at the Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, 1030 Vienna.

A 4% or 6% non-denaturing polyacrylamide gel in 0.25x TBE was cast and prerun for 30 minutes at 180 V. The protein-DNA binding reaction prepared by mixing the binding buffer (10 mM HEPES pH 7.9, 100mM NaCl, 4% Ficoll400, 1 mM EDTA) with the protein and 2 µg poly(dI-dC) as nonspecific competitor. At last the radiolabeled probe was added and the binding reaction was incubated for 30 minutes on ice. The reactions were loaded onto the gel and run for 3-3.5 h at 230 or 290 V. The gels were then exposed to an x-ray film for at least 24 hours.

Sequence of the probe:

5′ AGGGCGTCAACAACTATTTCTTTTCCTTTGATCTTTGTCAACAAAGAGG 3′

3.4 Binding site multimerization

3.4.1 Double strand synthesis of the oligonucleotide containing the binding site

5µg (166.6 pmol) selection oligo were mixed with two fold excess of reverse primer and PCR reaction mixture (80 mM Tris, 20 mM ammonium sulfate, 0.02% Tween 20, 2.5 mM magnesium chloride, 0.2 mM dNTP (Thermo), 0.025U Taq polymerase (AGROBIOGEN)) to a final volume of 50 µL.

Temperature program:

98°C	30 sec	
55°C	30 sec	15 cycles
72°C	90 sec	
72°C	30 min	

After the PCR reaction the double stranded oligonucleotide was purified by phenol/chloroform extraction (25:24:1 of phenol:chloroform:isoamylalcohol) and ethanol precipitation. The precipitated double stranded oligonucleotide was then resuspended in water to a final concentration of 1µg/µL. The concentration was measured by comparing the

intensity of the bands on a 2% agarose gel to an oligonucleotide with the same length and known concentration.

3.4.2 PCR amplification of the oligonucleotide

The oligonucleotide, 25 pmol of each primer, 1 mM dNTPs were amplified in 1x HF-Buffer (Thermo) by 0.5 U Phusion polymerase (Thermo).

Temperature program:

98°C	30 sec	⎫
55°C	30 sec	⎬ 25 cycles
72°C	90 sec	⎭
72°C	5 min	

3.4.3 DNA purification with spin columns

To get rid of the primers after PCR or small digestion fragments the oligonucleotide DNA was purified by the PCRapace Kit (INVITEK), which has a fragment length cutoff of 80 bp. The kit was used according to the manufacturer's protocol.

3.4.4 A-Tailing of blunt end insert DNA and ligation in the pGemTEasy vector (Promega)

6 µL of the PCR reaction was incubated with 0.5 U polymerase and 0.2 µM dATP in the according buffer for 20 minutes at 72°C. The A-tailed insert DNA was ligated into the pGemTEasy (Promega) vector in a molar ratio of 3:1 over night. The vector was then transformed into E. coli as described in 3.2.2 Transformation of E. coli Top10F'.

3.4.5 Phenol/chloroform extraction

For the phenol/chloroform extraction the same volume of phenol:chloroform:isoamylalcohol (25:24:1) as the sample was added to the sample. The tube was mixed vigorously by vortexing and then centrifuged (5 min, 14000 rpm) to separate the two phases. The aqueous (top) phase was transferred to a new tube and directly proceeded with the ethanol precipitation.

3.4.6 Ethanol precipitaion

Up to a volume of 200 µL sample 133 µL 7.5 M ammonium acetate, 3 µL mussel glycogen (Thermo) and 1 mL of ethanol (EtOH) 100% were added. If the sample volume was bigger than 200 µL, aliquots were made and treated separately. The sample was placed in the freezer at -80°C for at least 60 minutes and was then centrifuged (30 min, 14000 rpm) at 4°C. The supernatant was sucked off and the pellet washed twice with 1 mL ice cold EtOH 70%. Afterwards the pellet was air-dried for 10 minutes. DNA was resuspended in water or an appropriate buffer for the next step.

3.4.7 DNA-PAGE

Polyacrylamide gels were prepared by mixing acrylamide (30%, 29:1) (Roth) and 10x TBE in water to obtain the appropriate pore size. 6% and 16% acrylamide was used and 0.25x TBE. To start the polymerization of acrylamide APS and TEMED was added. The gel was run in 0.25 x TBE at 176V (20 V/cm).

The same DNA-Ladder, as for agarose gel electrophoresis, was used. For specifying the 34 bp band an oligo with the same size was annealed and also loaded onto the gel.

PAGE	30% Acrylamide (29:1)	H_2O	5x TBE	10% APS	TEMED
6%	4 mL	11.86 mL	4 mL	120 µL	20 µL
12%	8 mL	7.86 mL	4 mL	120 µL	20 µL
16%	10.66 mL	5.2 mL	4 mL	120 µL	20 µL

Table 3: Recipe for 20 mL non denaturing polyacrylamide gel for DNA-PAGE.

The gels were post gel stained with either ethidium bromide or GelRED™. For post gel staining with ethidium bromide the gel was incubated 15 minutes in a 0.5 µg/mL ethidium bromide in 0.25x TBE and then destained in ultrapure water for another 15 minutes. The bands were visualized by the Alphaimager® Mini (Cell Biosciences)

For the staining with GelRED™ a 3x staining solution of GelRED™ in 0.25x TBE was used. The gels were incubated 30 minutes to one hour in this staining solution. No destaining is necessary. Pictures were taken by the Alphaimager® Mini (Cell Biosciences)

3.4.8 Multimerization of the binding site (final version)

2.1 nmol (~67.5 µg) of the purified double stranded oligonucleotide was digested with XhoI (Thermo) in a total volume of 100µL overnight at 37°C. The digested oligo was resolved on a 16% PAGE. The 34bp band was excised and fragmented by centrifugation through a PCR tube with a hole poked into the bottom. The DNA was then eluted in 150 µL LoTE (3mM Tris-HCl pH 7.5, 0.2mM EDTA pH 7.5) containing ammonium acetate (LoTE : ammonium acetate; 125:25) at 37°C for 2 hours or at 4°C overnight and was subsequently ethanol precipitated. The DNA-pellet was resuspended in ligation buffer (Thermo) and the binding sites were self-ligated by the addition of T4 DNA Ligase (Thermo) at room temperature (RT) for 4 hours to overnight. These multimerized, longer molecules are called “concatemers”. To linearize circulated ligation products a partial digestion with XhoI (10 U, 37°C, 1 min) was performed. The linearized concatemers were resolved on a 6% PAGE. Gel pieces containing concatemers with different length were excised: <300 bps, 300-500 bps and >500 bps concatemers. DNA was eluted and ethanol precipitated. The concatemers were then ligated (10U T4 DNA Ligase, RT, 4 o/n) into a SalI digested, alkaline phosphatase treated pUC19 vector. The plasmids containing the multimerized binding sites were then transformed into E. coli TOP 10F’. To screen for positive clones alpha complementation was used. Therefore IPTG and X-Gal was added prior to plating. White colonies were picked for miniprep (see 3.2.3). The size of the inserted concatemer was determined by the digestion of the Miniprep DNA with PstI and XbaI restriction enzymes. Plasmids containing large inserts were then subjected to sequencing.

3.4.9 Sequencing of plasmid DNA

1 µg of plasmid DNA from miniprep was mixed with 25 pmol sequencing primer and sent to Microsynth AG for sequencing.

Plasmid	Primer sequence
pBSII KS(+),pGemT®Easy	TAATACGACTCACTATAGGG
pUC19	TGTAAAACGACGGCCAGTG

Table 4: Primers used for sequencing the inserted binding sites in different plasmids.

3.4.10 Analysis of the obtained sequences and alignment of the binding sites

The obtained sequences were checked for correctness by analysis of the sequencing-chromatogram. The sequences of the concatemers were then divided into individual binding sites by splitting the sequence at the ligation sites. Subsequently, the orientation of the binding sites was checked and, if needed, adjusted, simply by inversion of the complementary sequence. Sequences with missing bases in the random segment or mutations in the known Tcf/Lef consensus binding site were omitted. The sequences were aligned using MEGA5.1[118,119].

3.4.11 Calculation of position specific scores and creation of a sequence logo

To calculate position specific scores and generate a sequence loge from the aligned binding site sequences the web based application WebLogo3 (http://weblogo.threeplusone.com) was used. The .fasta file from the alignment, created by MEGA5.1, was submitted to WebLogo3 to generate the sequence logo and score matrix form the obtained sequences[118,119].

4 Results

The binding site selection can be seen as a process consisting of four consecutive steps. At first the proteins of interest have to be prepared in sufficient concentration, amount, purity and after all, they must be correctly folded and bind DNA. The next step is the PCR-based binding site selection itself, were through cyclic amplification and selection of targets (CASTing), the oligonucleotides with the highest affinity to the protein are enriched from a library of random oligonucleotides. These high affinity binding sites are then sequenced. To increase the yield of sequencing the selected short oligonucleotides are multimerized resulting in longer, multiple binding sites containing DNA fragments called concatemers. The final module could be seen as the interpretation of the acquired sequences.

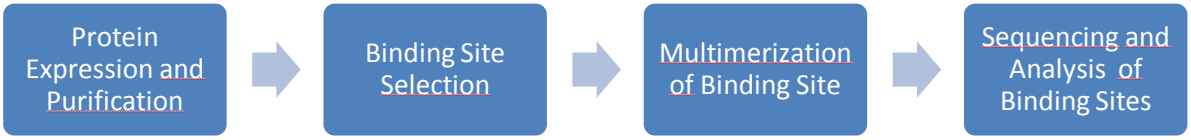


Figure 11: Scheme of the process of a binding site selection.

The aim of this diploma thesis was to prepare functional proteins expressed in E. coli in sufficient concentration, amount and purity. The proteins were expressed in E. coli and in mammalian cells, because it cannot be ruled out that posttranslational modifications of β-catenin affect the interaction with Lef1 or the DNA-binding.

A further goal was to establish a method for the multimerization of the selected binding sites.

4.1 Protein Expression

In these experiments truncated versions of murine Lef1 and human β-catenin were used, both lacking the main site of interaction between each other. Two different truncated forms of Lef1 were used. The one consists only of the high mobility group (HMG) DNA binding domain of Lef1, designated LefHMG, and the other starting at the HMG domain, at amino acid 293, including the whole C-terminal end. This second form of Lef1 is designated Lef293. In cell culture reporter assays, with different Tcf/Lef binding sites in the promoter for a firefly luciferase reporter gene, the transcription activation of Lef and Lef-β-catenin complexes were investigated. The results of these reporter assays pointed at the possibility of a change in the affinity of Lef to its DNA binding sites upon β-catenin binding. To be able to verify, if the binding of β-catenin to Lef changes only the affinity of Lef to its DNA binding site, or if it also changes the preferred nucleotide sequence of its DNA binding site, fusion proteins with β-catenin were made. These fusion proteins contain one form of truncated Lef1 fused via a myc linker to an N-terminal truncated β-catenin. This N-terminally truncated β-catenin contains the armadillo repeats 10-12 and the C-terminal end, therefore lacking the primary Lef interaction domain which would interact with the Lef1 N-terminus. The fusion proteins are named after the truncated form of Lef they contain followed by βcat.

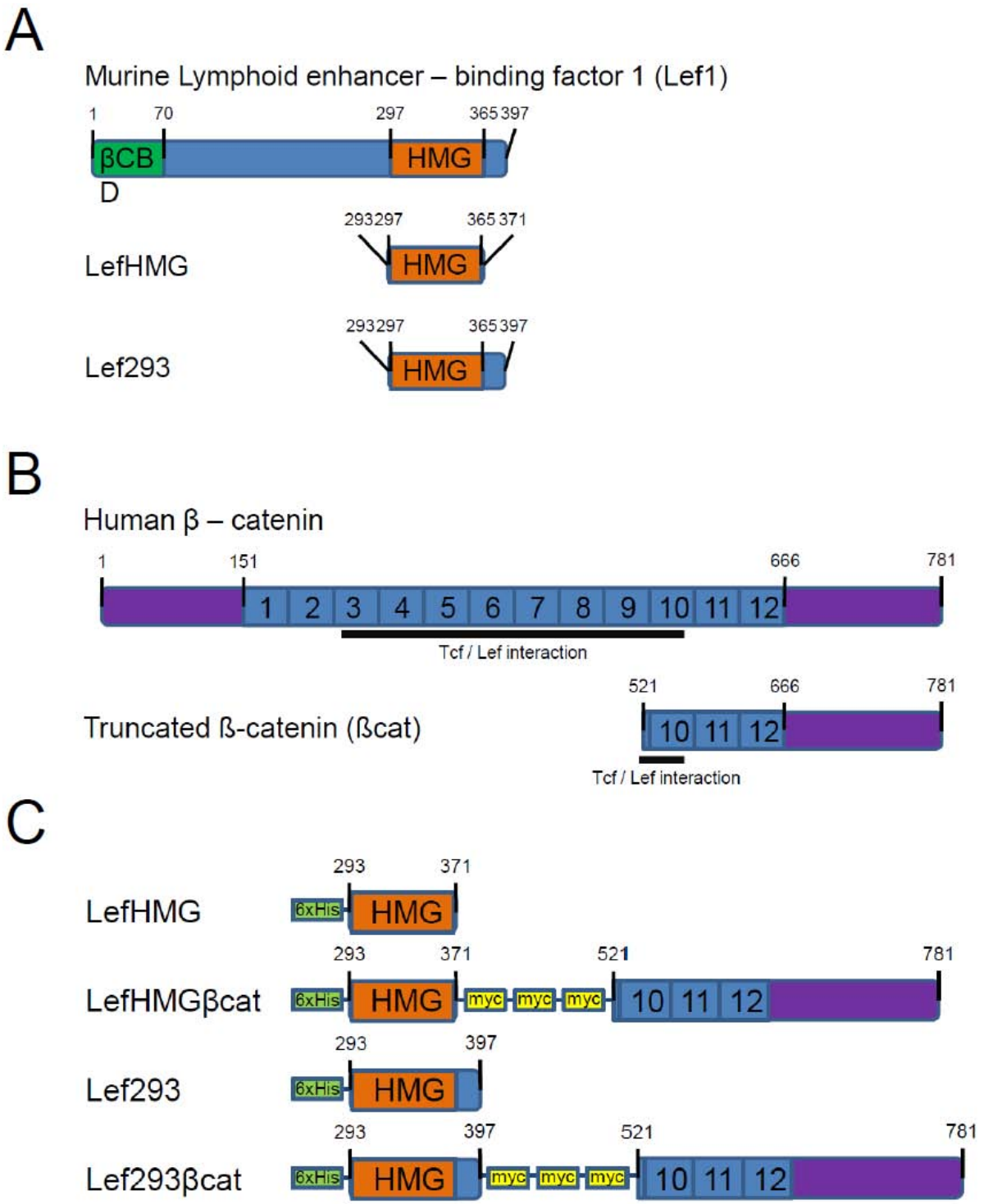


Figure 12: Proteins for the binding site selection. A: murine Lymphoid enhancer-binding factor 1 (Lef1) complete its truncated forms. B: human full length and truncated β-catenin C: Proteins for the binding site selection that were expressed

4.1.1 Plasmids for protein expression

The first step was to clone the DNA coding for the proteins of interest into a backbone vector suited for the expression of proteins in the desired host. For the expression in E. coli Novagen's® pET vector was chosen. This vector is optimized for high level protein expression due to a strong T7 promoter under control of the lac operator. The T7 polymerase has to be provided by the host. For an easier purification of the expressed protein an N-terminal His-tag is added by the pET vector.

Originating from the pET vector, the ORF was cloned into the pMC and subsequently into the pKC backbone vector. Both of these vectors are designed for the protein expression in mammalian cells. In pMC as well as in pKC the protein expression is driven by the strong constitutive CMV promoter, originating from the simian and human cytomegalovirus, respectively. Despite the different origin of the CMV promoters the main difference between these backbone vectors is a polyoma origin of replication provided by pKC which leads to an amplification of the plasmid when transfected into a polyoma transformed host cell. This increased plasmid copy number results in a higher expression level of the desired protein.

Protein	E. coli	HeLa cells	COP-8 cells
LefHMG	pET LefHMG	pMC LefHMG	pKC LefHMG
LefHMGβcat	pET LefHMGmβcat(521-781C)	pMC LefHMGmβcat(521-781C)	pKC LefHMGmβcat(521-781C)
Lef293	pET Lef293m	pMC Lef293m	pKC Lef293m
Lef293βcat	pET Lef293mβcat(521-781C)	pMC Lef293mβcat(521-781C)	pKC Lef293mβcat(521-781C)

Table 5: Plasmids that were cloned for the expression of the desired proteins in different host cells.

4.1.2 Protein expression in E. coli

4.1.2.1 Time course experiment with E. coli BL21(DE3)

To investigate the kinetics of protein expression a time course experiment was conducted. The pET vectors were transformed into E. coli BL21(DE3), an optimized E. coli strain, carrying a λ prophage containing the T7 polymerase under the control of a lac operator. Therefore, protein expression is repressed until induction with IPTG. To monitor the bacterial density of the expression culture, the absorption of the cell suspension was measured with a photometer at a wavelength of 600 nm (OD_{600}). The expression cultures were grown to an OD_{600} of approximately 0.6 before induction. After induction, samples were taken for OD_{600} and SDS-PAGE analysis at an interval of one hour (Figure 13). Since IPTG can have negative effects at higher concentrations, two different concentrations were tested.

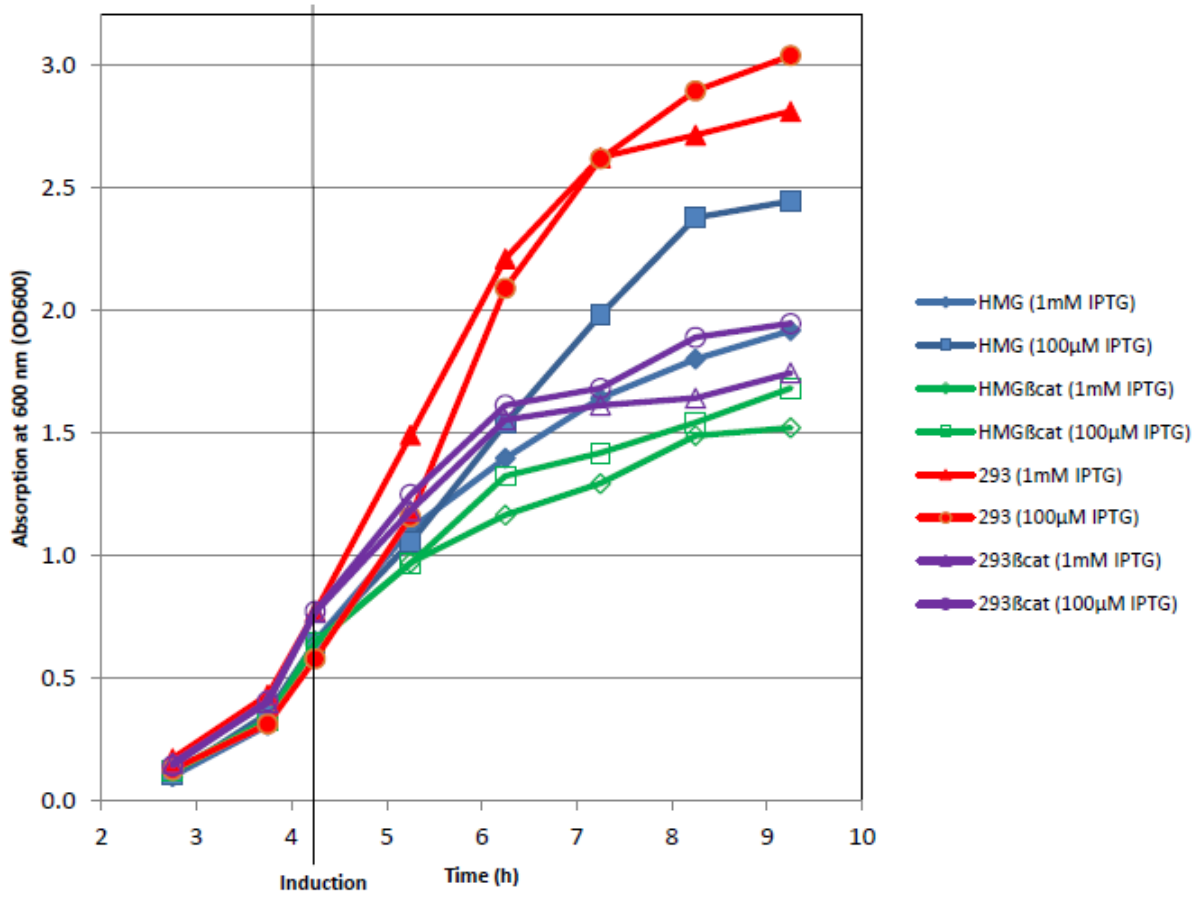


Figure 13: Optical density at 600 nm (OD600) of the different E. coli BL21(DE3) expression cultures measured over the time. The protein expression was induced by either 1 mM or 100 µM IPTG and the temperature was kept at 37°C. Lines in the same color designate the same protein at different IPTG concentrations. The mark differs from Lef to Lef-β-catenin by being filled or being empty. The Zero hour point of the x-axis is at the time point of inoculation. The vertical line gives the time point of induction of the protein expression by adding IPTG.

The OD_{600} values (Figure 13) show the growth characteristics of the different expression cultures. The graph starts somewhere at the beginning logarithmic growth phase and ends at the stationary phase. It can be seen that the stationary phase is reached in most cases after about three hours of expression. Cultivation after this point is unnecessary because the protein yield does not increase anymore.

To check if the proteins were expressed properly, crude extracts of E. Cole BL21 (DE3) cultures were analyzed by SDS-PAGE followed by Coomassie staining. To obtain comparable results, the samples were diluted according to their OD_{600} value before analysis.

The SDS-PAGE showed (Figure 14) that the LefHMG and Lef293 proteins were not expressed at all (Figure 14A, 2C, 2E and 2G). The Lef-β-catenin fusion proteins (Figure 14B, 2D, 2F and 2H) showed thin bands after four hours of expression.

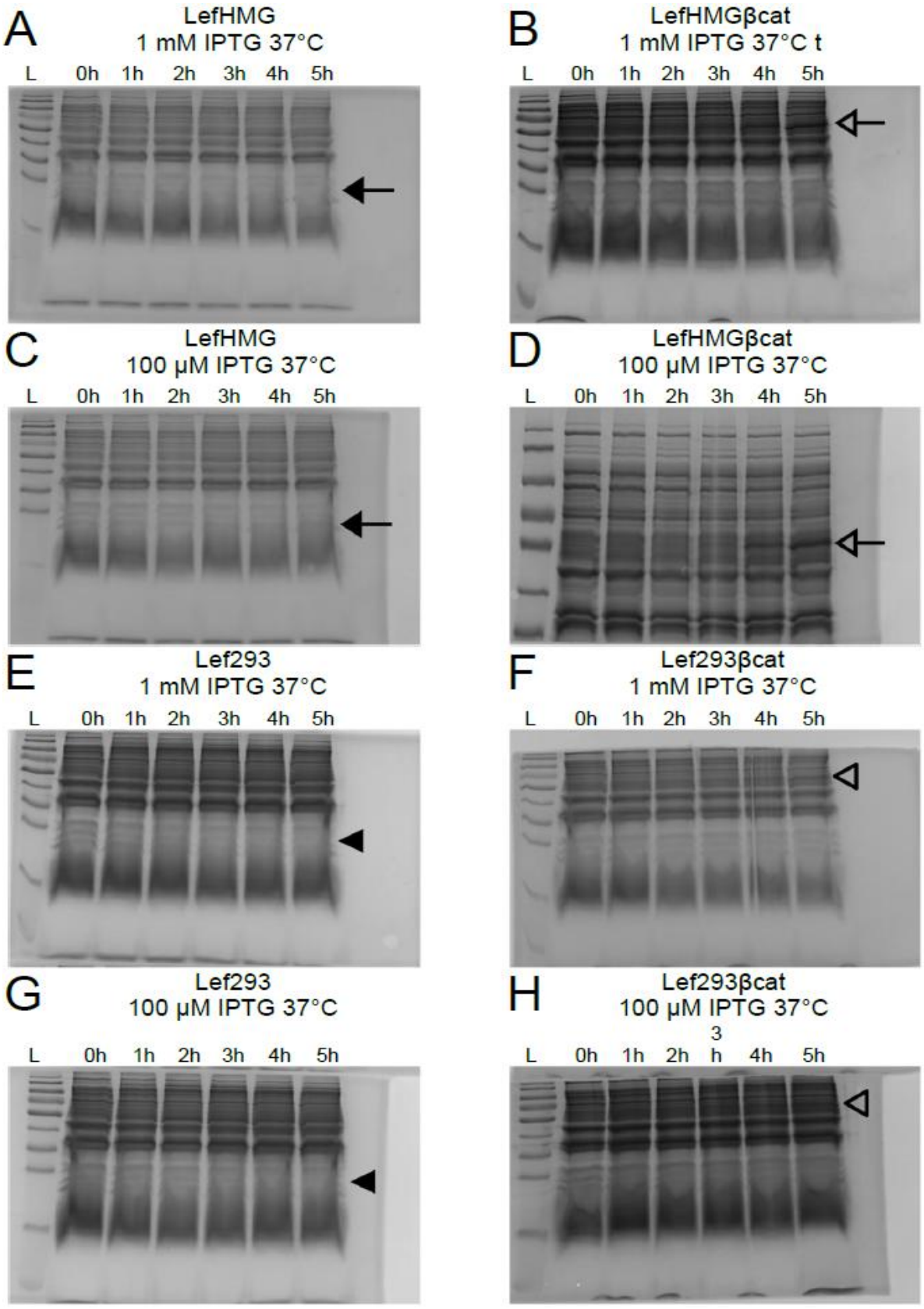


Figure 14: Time course of E. coli BL21(DE3)protein expression. Coomassie stained SDS-PAGE's of crude extracts from the different E. coli expression cultures. Samples were collected in one hour intervals and were diluted according to their OD_{600} value (Figure 13) to give comparable results The arrow indicates where bands of LefHMG, the empty arrow the ones of LefHMGβcat, the arrowhead where the bands of Lef293 and the empty arrowhead that of Lef293βcat would be expected. As ladder (L) Thermo Scientific's PageRuler™ was used.

4.1.2.2 Codon usage of E. coli and the expressed proteins

Because this first time course experiments showed that the target proteins were not expressed at a reasonable rate, the cause of the problem had to be found and solved. A possible reason for this could be the differences in codon usage of humans and E. coli. Therefore the amino acid composition of the open reading frame coding for the target proteins were calculated and compared with values from literature[120]. In fact this examination revealed huge discrepancies between the codon usages of these species. The biggest differences in codon usage can be observed by codons coding for arginine, isoleucine and leucine. Bottleneck codons for all four protein open reading frames are shown in Table 6, the whole codon usage tables are shown in the appendix (7.3 Protein properties). To overcome this problem a different E. coli strain, E. coli BL21(DE3)RIL, was used. This E. coli strain is a descendant from the E. coli strain BL21(DE3) used before, with the advantage of carrying a plasmid coding for the tRNAs that code for these rare codons. The tRNAs code for the codons shown in Table 6 and should greatly increase the translation rate and therefore increase the protein expression.

Amino acid	tRNA	Codon	LefHMG	LefHMG βcat	Lef293	Lef293 βcat	E. coli [h^{-1}]		
							0.4	1.07	2.5
Arginine	*argU*	AGA	7 / **45.16**	10 / **24.04**	9 / **49.73**	12 / **27.15**	1.12	0.84	0.63
		AGG	0 / **0.00**	1 / **2.40**	1 / **5.52**	2 / **4.52**	0.09	0.05	0.03
Isoleucine	*ileY*	AUA	0 / **0.00**	1 / **2.40**	0 / **0.00**	1 / **2.26**	0.93	0.75	0.52
Leucine	*leuW*	CUA	4 / **25.81**	7 / **16.83**	5 / **27.62**	8 / **18.10**	2.15	1.53	0.83

Table 6: A subset of codons used by E. coli and the expressed proteins. E. coli codons are given as per thousand codons in coding regions of the genome, at different growth rates (doublings per hour). The amount of codons used by the expressed proteins (codons per ORF/codons per thousand codons).

4.1.2.3 Time course experiment with E. coli BL21(DE3)RIL

A second time course experiment was carried out. The setup was the same as the first time course experiments, except the expression host was changed from E. coli BL21(DE3) to E. coli BL21(DE3)RIL. The time course was again analyzed by OD_{600} measurement (Figure 15) and by SDS-PAGE (Figure 16).

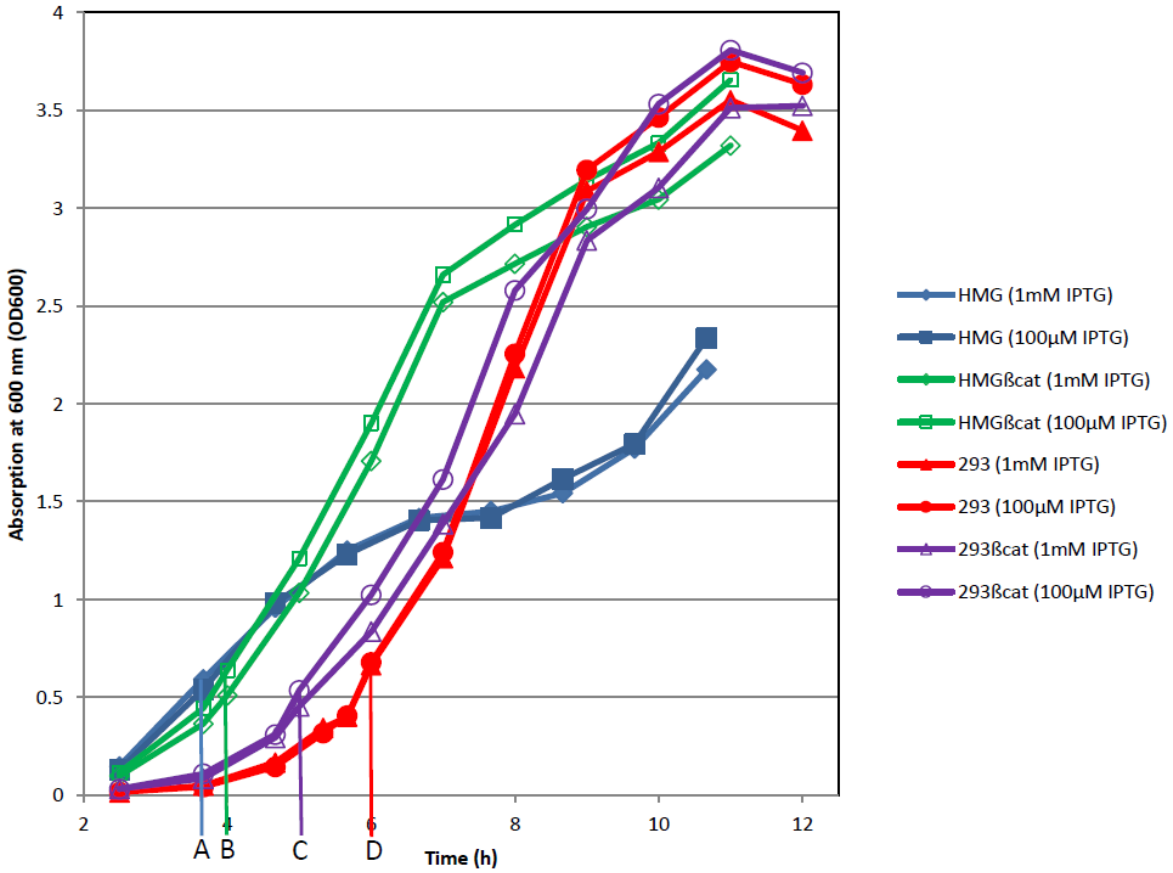


Figure 15: Cell density of the different E. coli BL21(DE3)RIL expression cultures over time. The protein expression was induced by either 1 mM or 100 µM IPTG and the temperature was kept at 37°C. Lines in the same colour designate the same protein at different IPTG concentrations. The mark differs from Lef to the Lef-β-catenin fusion protein by being filled or being empty. The zero hour point of the x-axis is at the time point of inoculation. The vertical lines give the time points of induction. A: point of induction of LefHMG, B, C, D, points of induction of LefHMGβcat, Lef293, Lef293βcat, respectively.

The different time points of induction of the expression cultures resulted from the different length of the lag phases. The course of the OD600 values of the LefHMG expression culture (Figure 15, blue lines) could indicate a contamination by different bacteria of the overnight culture, because both lines run nearly parallel. It could also indicate the loss of the plasmid in culture and the growth advantage of the bacteria not containing the plasmid. The SDS-PAGE analysis showed that all four proteins are expressed at high levels at both IPTG concentrations, so that it is unclear why the LefHMG culture has such growth behavior.

The OD_{600} graph shows further, that after two hours protein expression the bacterial growth slows down and reaches the stationary phase after approximately three hours. The height of these plateaus varies between the different cultures and batches. The samples taken during protein expression were diluted to the same OD_{600} and analyzed by SDS-PAGE. The course of the OD600 values of the LefHMG expression culture (Figure 15, blue lines) could indicate an internal contamination, meaning that the E. coli got rid of its plasmid and is outgrowing the E. coli still carrying the plasmid. Because both lines run nearly parallel this loss of the plasmid is supposed to have occurred at the overnight culture step.

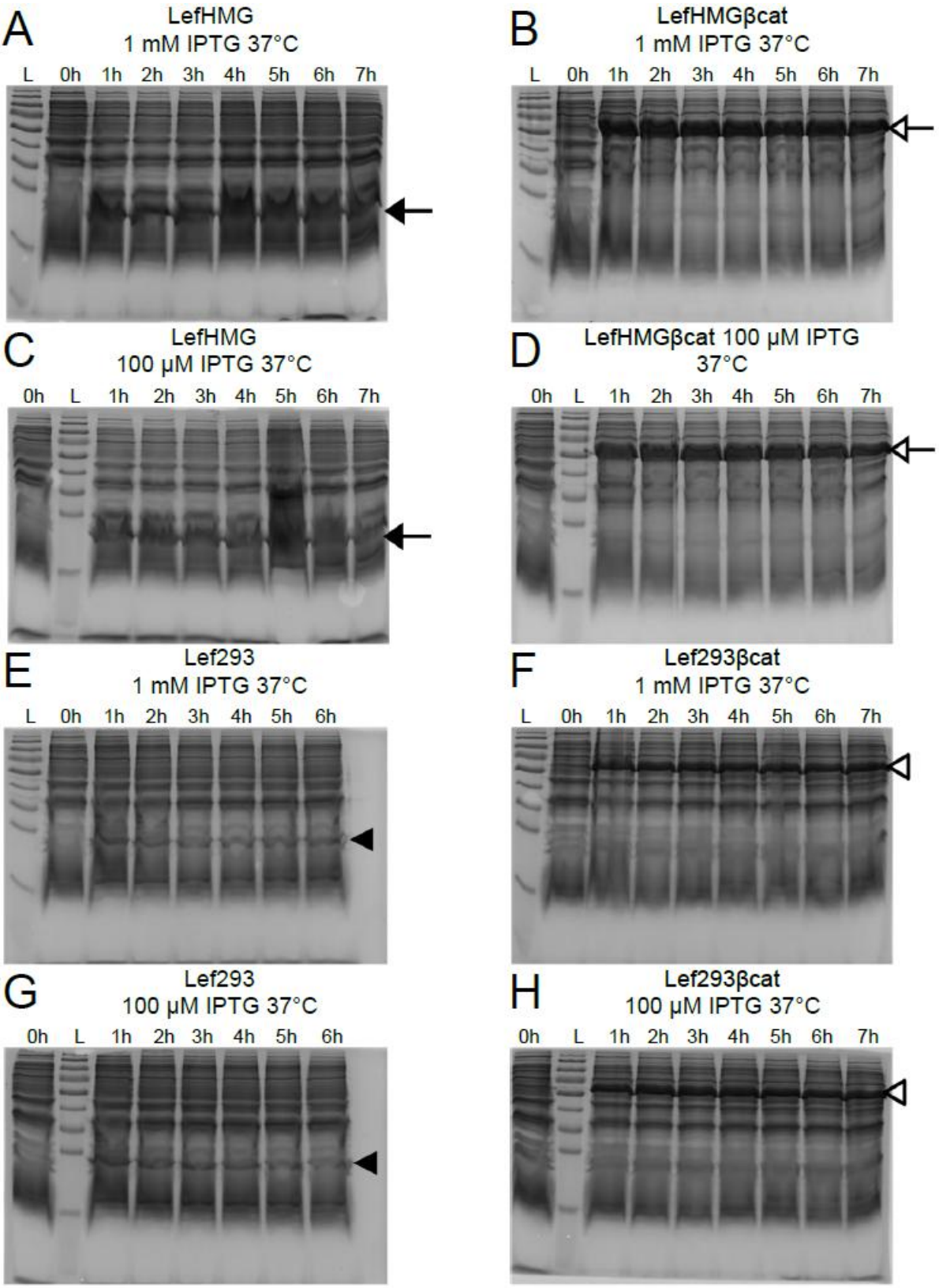


Figure 16: SDS-PAGE analysis of the E. coli BL21(DE3)RIL time course experiments. Samples were collected in one hour intervals. To give comparable results the samples were diluted to the same OD_{600} value (Figure 15). Arrowheads indicate the bands of LefHMG, Lef293, and the empty arrowheads indicate the LefHMGβcat and Lef293βcat bands. The uninduced control is denoted as 0h. As ladder (L) Thermo Scientific's PageRuler™ was used.

The SDS-PAGE gels showed that after one hour of expression the maximum protein concentration in the expression culture were reached and the protein yield only increases with bacterial cell densities.

4.1.2.4 Protein purification under denaturing conditions

When proteins are expressed in high concentrations they tend to agglomerate and form insoluble inclusion bodies. In this case the proteins are denatured and have to be refolded in vitro. To achieve this, the expressed proteins, as inclusion bodies, are isolated, solubilised and denatured by the addition of a denaturing agent (in this case guanidine). Subsequently the proteins are refolded by removing the denaturing agent by dialysis. After dialysis the proteins can be concentrated with a centrifugal filter.

4.1.2.4.1 Preparative expression of proteins in BL21(DE3)RIL

The proteins were expressed in BL21(DE3)RIL and the expression was induced by the addition of 1 mM IPTG and the proteins were expressed at 37°C for 6 hours. As can be seen in Figure 17, a high yield was achieved.

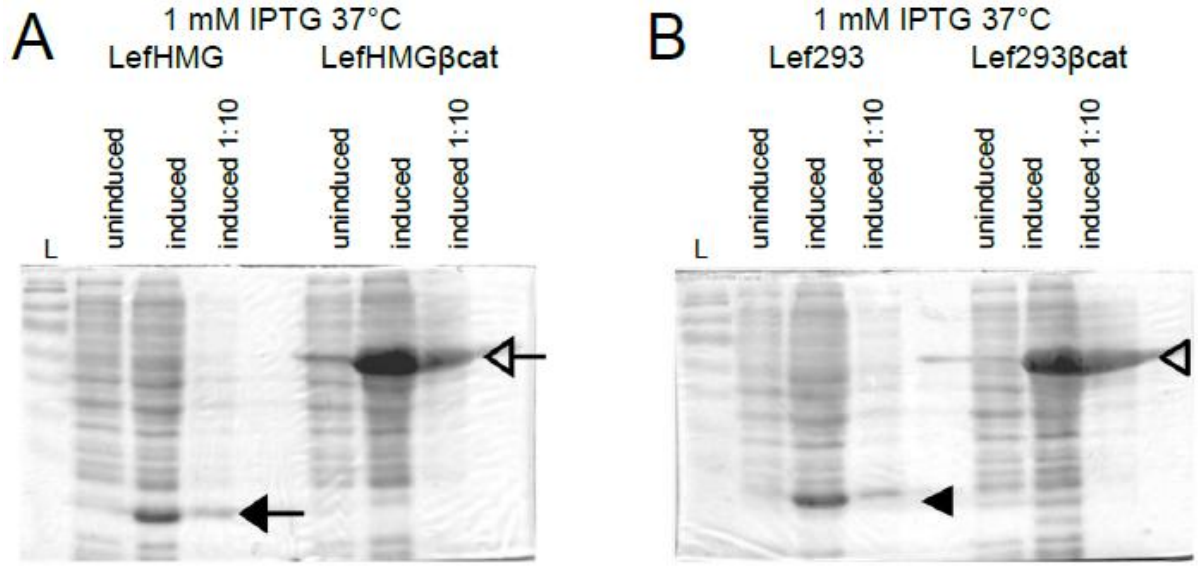


Figure 17: SDS-PAGE of the different expression cultures. Comparison of uninduced control sample with expressed protein and its 1:10 dilution. The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

4.1.2.4.2 Purification of the proteins by metal ion affinity chromatography

After the proteins were solubilised and denatured they were bound to Ni-NTA Agarose. Elution occurred with an increasing imidazole gradient ranging from 25 mM to 250 mM. 16 fractions were collected and checked for protein content with a SDS-PAGE analysis. The gels (Figure 18) show that LefHMG and Lef293 are contained in fraction 2 to 7 and that LefHMGβcat and Lef293βcat in fraction 1 to 6. The gels of the fractions 9 to 16 contained no proteins and are not shown.

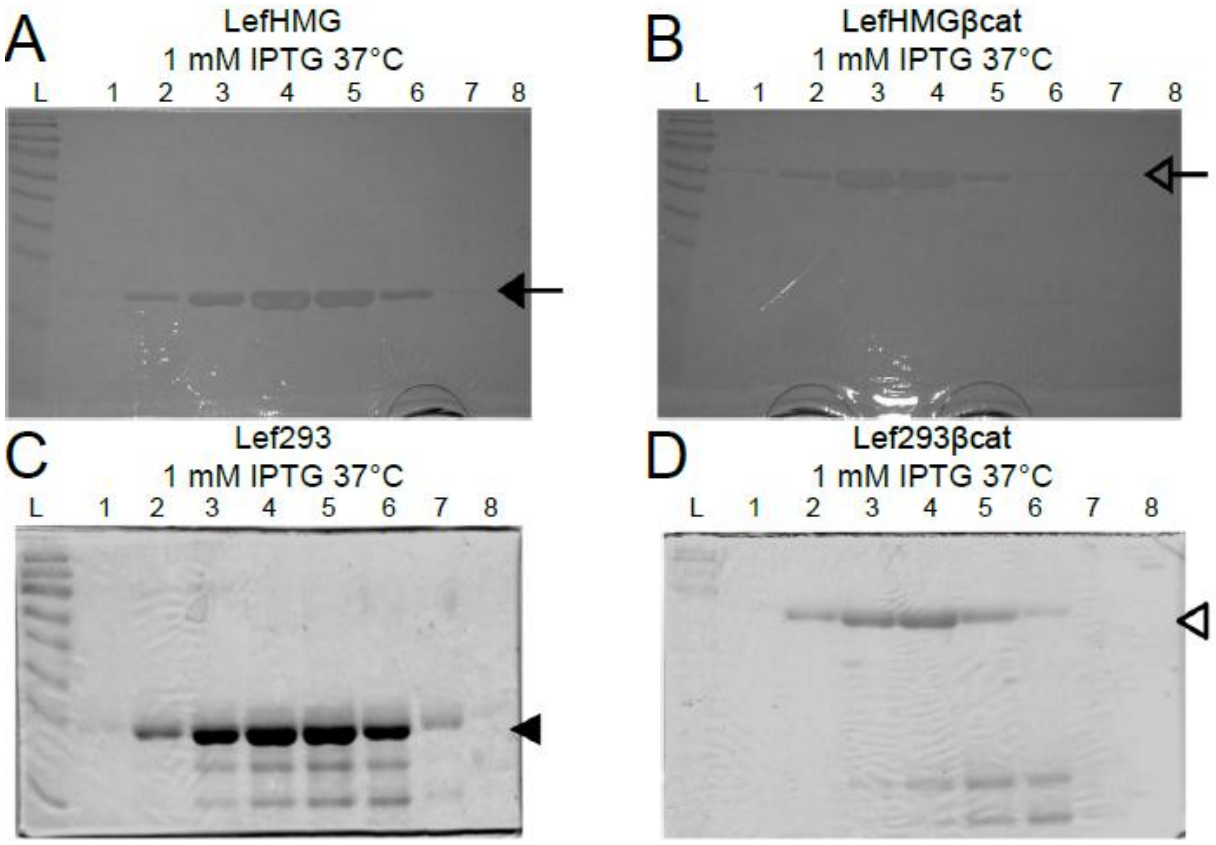


Figure 18: Collected fractions of the gradual elution from the Ni-NTA agarose gravity flow columns. The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

4.1.2.4.3 Refolding of the protein by dialysis and concentration via spin filter

To refold the denatured proteins the protein containing fractions (Figure 18) were pooled and subjected to dialysis. To allow refolding the denaturing agent, in this case urea, was removed gradually. After dialysis the protein solutions of Lef293 and Lef293βcat were both turbid from precipitated protein. The precipitation could indicate incorrect folding or low solubility of the protein. The protein solutions of LefHMG and LefHMGβcat were clear with no precipitate. The Lef293 and Lef293βcat suspension was centrifuged to clear the solution of precipitated protein and a test to concentrate the solution with centrifugal filters, yielded again a turbid solution, indicating that the solution is saturated with protein and no further concentration is possible. LefHMG and LefHMGβcat protein solutions were concentrated by centrifugal filters. After centrifugation the solutions were still clear. To check the result of the concentration step the concentrated protein solutions were analyzed by SDS-PAGE. For Lef293 and Lef293βcat the unconcentrated cleared solutions were analyzed.

The gel in Figure 19A shows that the spin column concentration procedure yielded an approximate tenfold increase in protein concentration of LefHMG and LefHMGβcat proteins. Figure 19B shows that Lef293 and Lef293βcat are still in solution even the concentrations are very low. The protein amount run on the gel Figure 19B) is fifteen times bigger than that of LefHMG unconc and LefHMGβcat unconc in Figure 19A.

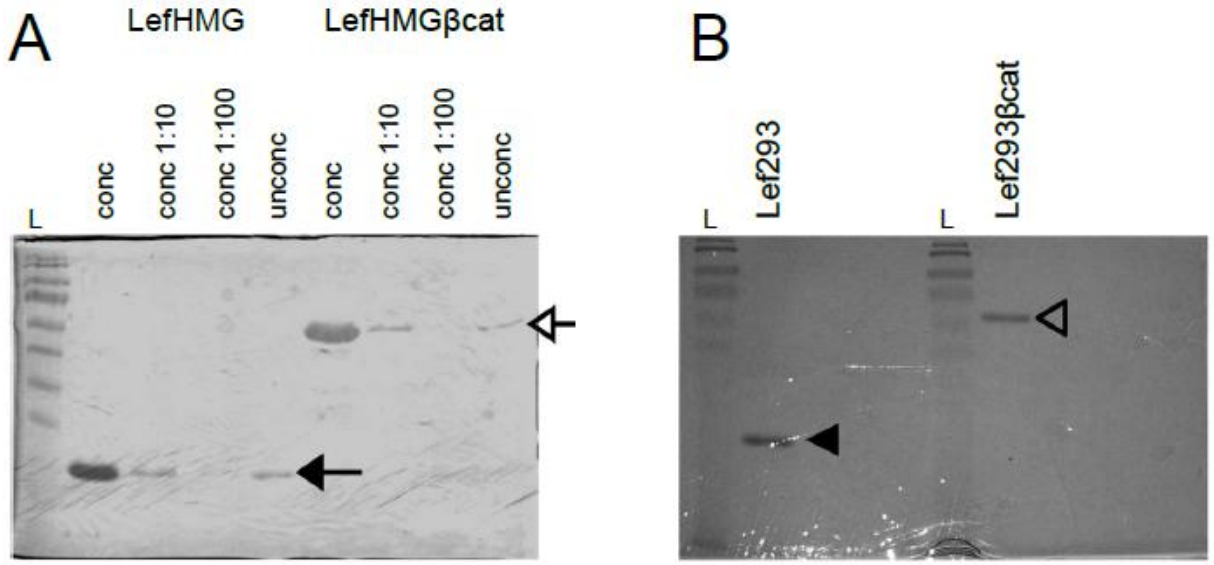


Figure 19: Purified and refolded proteins after dialysis and centrifugal concentration. A shows LefHMG and LefHMGβcat directly after dialysis (unconc) and after concentration. Conc denotes the undiluted concentrated protein solution, conc 1:10 and conc 1:100 the 1:10 and 1:100 dilution, respectively, of the concentrated solution. B shows Lef293 and Lef293βcat after dialysis in a fifteen fold higher protein amount than in A unconc. The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler^{TM} was used.

Finally, all four proteins were analyzed by SDS-PAGE to compare the different protein concentrations. The result (shown in Figure 20) shows, as suggested, a strong difference in the concentration between the two Lef variants, LefHMG and Lef293, regardless of their associated fusion proteins.

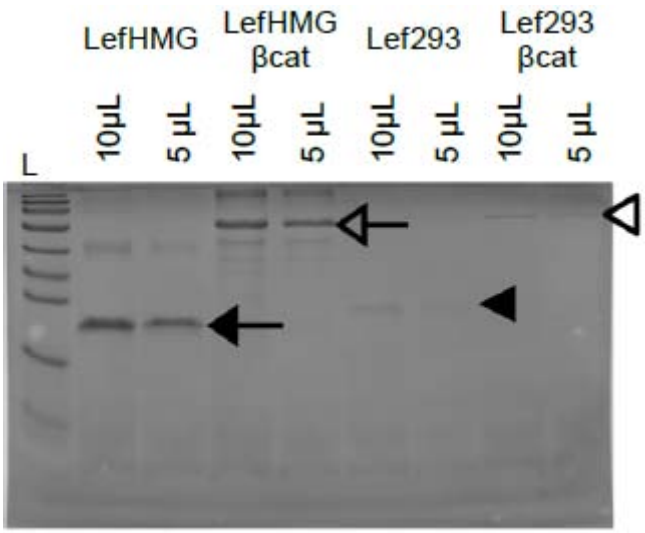


Figure 20: Comparison of the concentration of the different purified and refolded proteins. Two amounts of protein were loaded for each of the four different proteins. The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler^{TM} was used.

4.1.2.4.4 Electrophoretic mobility shift assay of the expressed proteins

To check if the purified proteins contain correctly folded, DNA-binding proteins an electrophoretic mobility shift assay was conducted. As probe a known Tcf/Lef consensus binding site, radiolabeled with ^{32}P-dATP, was used. The electrophoretic mobility shift assay (EMSA) showed a rather surprising result (Figure 21), to be specific, a shift of DNA with Lef293 and Lef293βcat, but no shift with the LefHMG proteins, meaning that they are not able to bind DNA and can therefore not be used for the binding site selection.

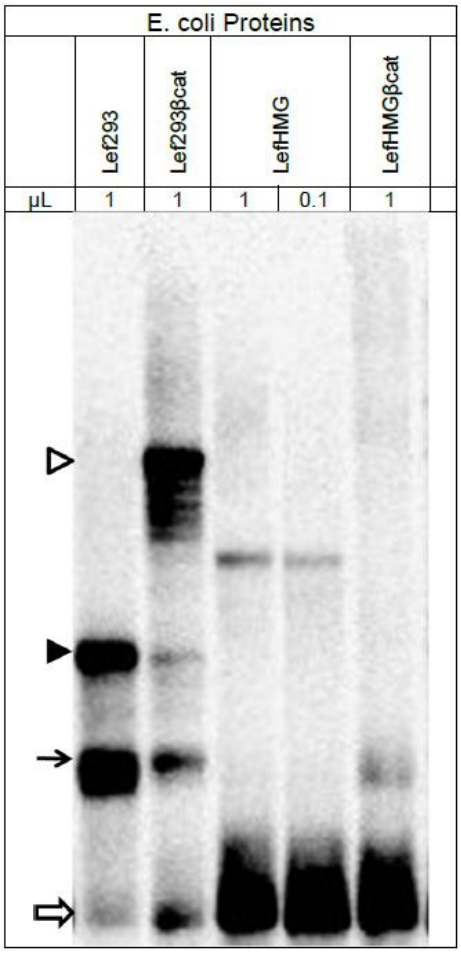


Figure 21: Electrophoretic mobility shift assay of the purified, refolded and concentrated proteins. As probe a radiolabeled Tcf/Lef consensus binding site was used. After incubating the proteins and the probe on ice the reaction was loaded on a 6% polyacrylamide gel and run at 290V for 3h. The results were obtained by exposing an x-ray film to the gel. The empty arrow indicates the free, unbound probe and the arrow some unspecific shift. The arrowhead points at the specific Lef293 shift and the empty arrowhead at the Lef293βcat shift. The lanes with LefHMG and LefHMGβcat show no specific DNA shifts since LefHMG is smaller than Lef293.

4.1.2.5 Protein purification under native conditions

Protein purification under native conditions has the advantage that the proteins once folded inside the cell, under prevalent intracellular conditions, into their three dimensional structure, this native conformation is retained during purification. For the native purification only soluble proteins are accessible once precipitated, they are lost in later steps. To achieve a high soluble protein concentration, different parameters in protein expression and purification can be modified.

4.1.2.5.1 Effect of the lysis buffer on protein solubility

To further work with the expressed proteins they must be extracted from the bacterial cells and subsequently be purified. At first the cells were broken by sonication and the insoluble content of the cells and the bacterial cell wall residues were removed by centrifugation. All the proteins contained in the supernatant are available for further purification steps. Therefore the expressed protein should be soluble in the lysis buffer.

This buffer is important because the proteins should be soluble and stable and the buffer should not interfere with the subsequent procedures. These buffers typically consist of a buffer substance, mono-and bivalent salts at various concentrations, and most of the times a detergent or low concentration of urea. In high concentrations, urea, as chaotropic substance, is a good denaturant, in low concentration though, urea stabilizes the hydrophobic parts of proteins, reducing the affinity of the proteins to form insoluble agglomerates[121].

To check if inclusion bodies are formed preferentially at higher inducer concentration or in the later stages of the expression the solubility of the proteins at different time points were tested. Therefore, the samples, taken during the time course experiment (see Time course

RIL), were resuspended in PBS (according to their OD_{600}), sonicated and the insoluble parts were separated by centrifuging. Subsequently, the insoluble pellets were resuspended in the same volume and the soluble and insoluble fractions were analysed by SDS-PAGE.

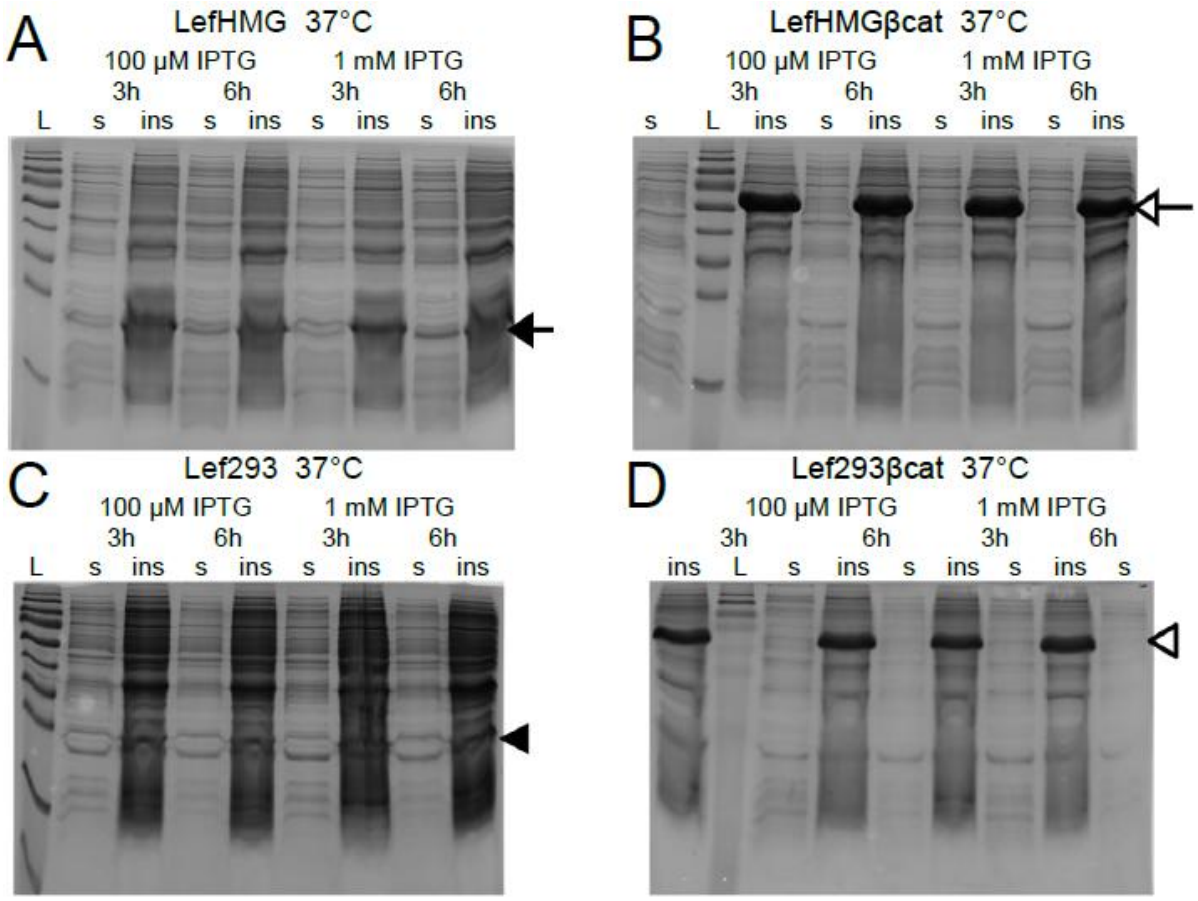


Figure 22: Solubility test of expressed proteins with different inducer concentrations and at different time points. Samples were sonicated in PBS, soluble (s) and insoluble (ins) fractions were separated and analysed. The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

The results from this analysis (Figure 22) showed that only a small portion of the proteins was soluble in PBS. LefHMGβcat and Lef293βcat (Figure 22B and 10D) may be insoluble at all. To increase the solubility a different lysis buffer is needed, therefore four different buffers were tested. Three of these buffers contained 20 mM Tris, pH 7.5, 50 mM NaCl, and 0.2% of one of the following mild, non-ionic detergents: Triton X-100 (TN-TX), NP40 (TN-NP40) or Tween20 (TN-T). Because these buffers should not interfere with subsequent purification procedures, the binding buffer (20 mM Na_3PO_4, 500 mM NaCl) (MB), suggested by the manufacturer of the magnetic beads used for the purification, was also tested. Therefore the insoluble fractions from the first solubility experiment (Figure 22) were tried to be solubilised by the different buffers and again analyzed by SDS-PAGE.

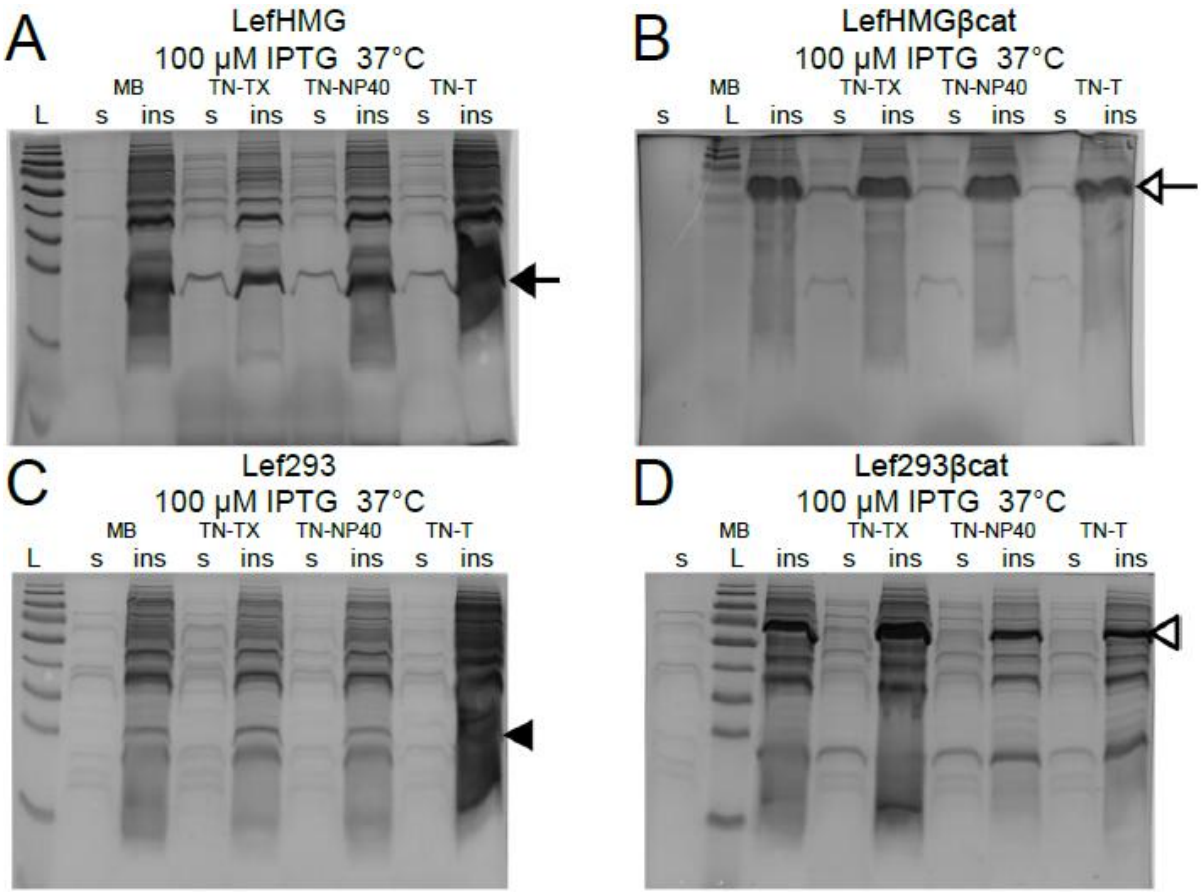


Figure 23: Solubilisation test of insoluble protein fraction in different buffers. Insoluble protein fraction were solubilised by different buffers and analysed. MB indicates the Magnetic bead binding buffer (20 mM Na_3PO_4, 500 mM NaCl) and in all other buffers TN stands for 20 mM Tris pH 7.5 and 50 mM NaCl and the latter acronym for 0.2% of the following detergents: TritonX-100 (TX), NP40 and Tween20 (T). The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

The results of this analysis clearly showed (Figure 23) that the proteins are generally more soluble in buffers containing detergents. It can also be seen that LefHMG and LefHMGβcat is better soluble than their corresponding Lef293 proteins. This result conforms with the observation from the dialysis step (4.1.2.4.3) of the protein purification under denaturing conditions, were Lef293 proteins formed a precipitate during dialysis and further concentration attempts, resulting in lower protein concentration opposed to LefHMG.

Although its artificial setup this experiment showed promising results and a further solubility test, where the cells were lysed in the correct buffer, was conducted. As buffers MB and TN-T were chosen, because MB might have beneficial properties by the later purification and TN-T because all buffers with detergents performed equal and Tween20 was chosen for further experiments. It was also tested if a solubilisation step between sonication and separation had a positive influence on the protein yield. Therefore the sonicated cells were shaken on a rocking platform for 15 min at 4°C. It was only tested with Lef293 and Lef293βcat (samples were taken from the time course experiment (4.1.2.3) because the solubility is less than that of the LefHMG proteins.

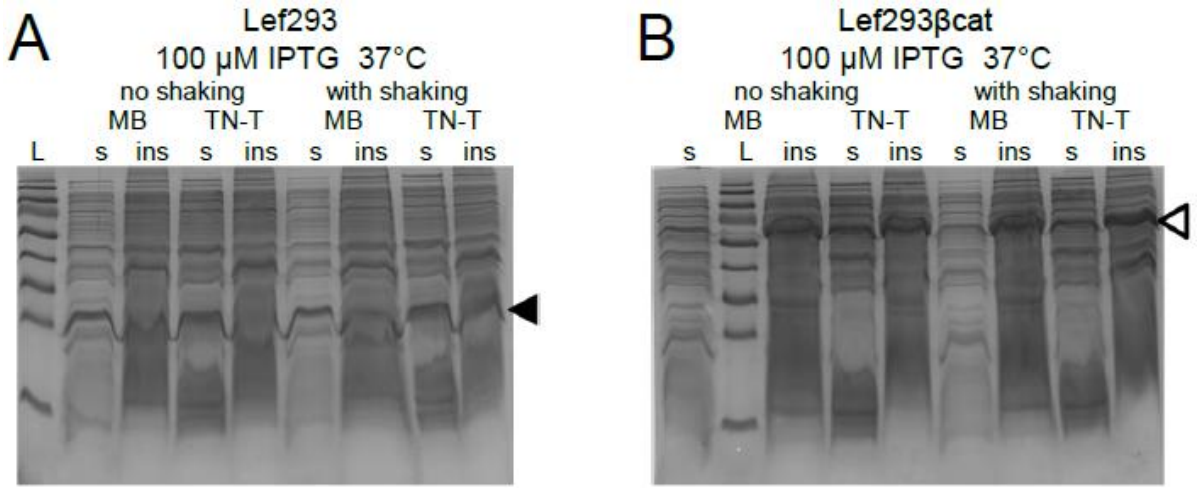


Figure 24: Solubility test with sonication in the correct lysis buffers with and without an extra solubilisation step between sonication and separation. Bacterial cells were resuspended in the correct lysis buffer before sonication. Half of the samples had an extra solubilisation step between sonication and separation (denoted as with shaking) the others were separated immediately after sonication (no shaking). The soluble (s) and insoluble (ins) fractions were loaded on the gel. MB indicates the Magnetic bead binding buffer (20 mM Na_3PO_4, 500 mM NaCl) and TN-T stands for 20 mM Tris pH 7.5, 50 mM NaCl and 0.2% Tween20. The arrowhead points at the bands of Lef293 and the empty arrowhead at that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

The results show a clear improvement in solubility in contrary to the first solubility test (Figure 24). The magnetic beads binding buffer was not able to solubilise as many protein as the TN-T buffer and therefore this approach was omitted. There was also no difference between the samples with and without the extra solubilisation step distinguishable, and therefore, for further experiments, the shorter version without extra solubilisation was used.

The solubility could be also increased by a slower expression rate. This is achieved by a lower expression temperature

4.1.2.5.2 Time course experiment with E. coli BL21(DE3)RIL at 25°C

A slower expression rate could yield a bigger fraction of soluble proteins. To reduce the expression rate the temperature during protein expression was lowered upon induction to 25°C. The protein expression was induced by 100 µM IPTG, because lower inducer concentration means a lower expression rate. To verify, if the expression rate had dropped a third time course experiment was conducted.

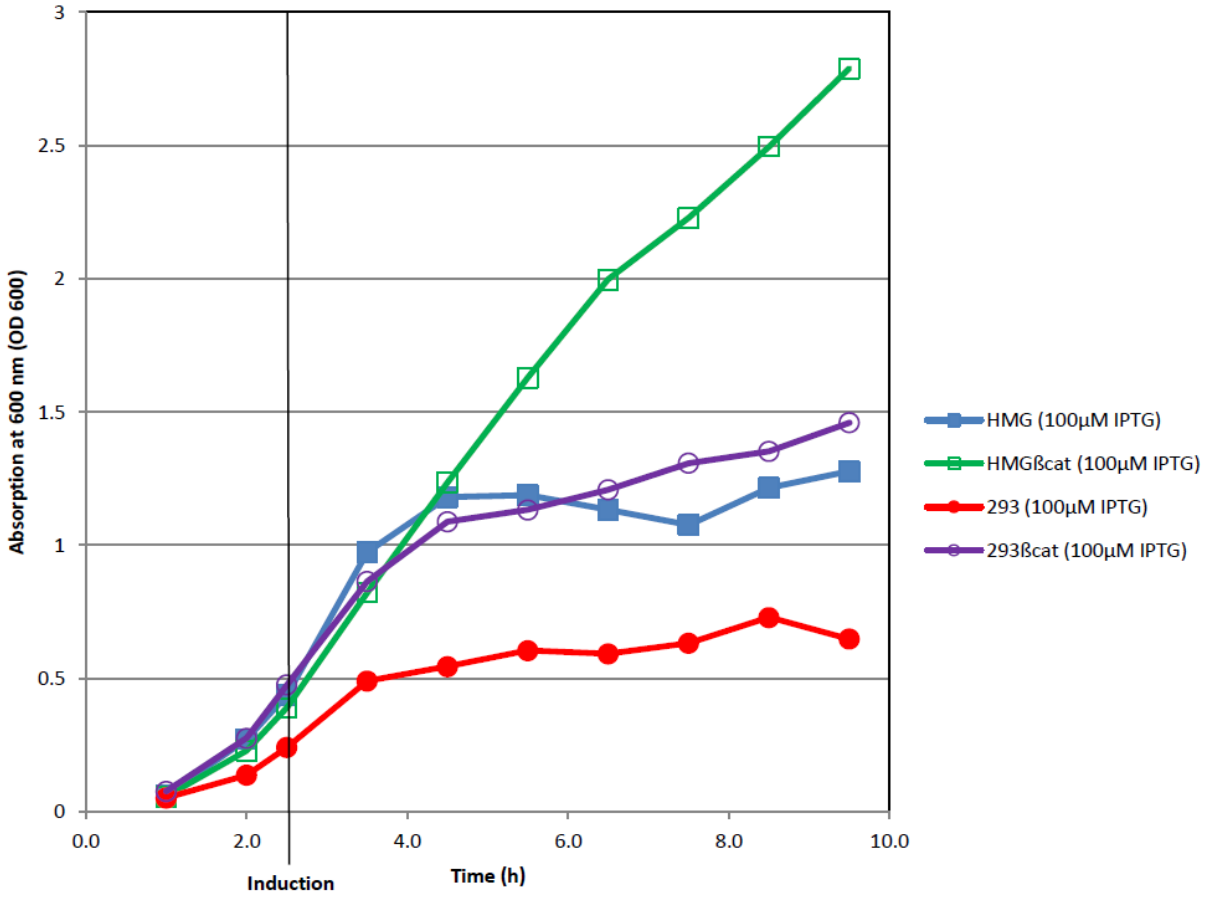


Figure 25: OD600 of the different E. coli BL21(DE3)RIL expression cultures over time. Protein expression was induced by 100 mM IPTG and the temperature during expression was lowered to 25°C. The zero hour point of the x-axis is at the time point of inoculation. The vertical line indicates the time point of induction.

OD_{600} (Figure 25) and SDS-PAGE analysis (Figure 26) of the third time course experiment showed that these parameters had the desired effect and slowed the protein expression down. The graph depicting the OD_{600} (Figure 25) of the 25°C expression cultures showed that the stationary phase is reached earlier than in the expression cultures at 37°C. The plateaus of the stationary phases seem to be generally lower than in 37°C cultures and may also depend on the OD_{600} value at the time point of induction. The expression culture of LefHMGβcat does not show these characteristics, which could indicate a contamination with a different E. coli strain. In spite of this unusual growth behavior the expression of LefHMGβcat was confirmed by SDS-PAGE analysis.

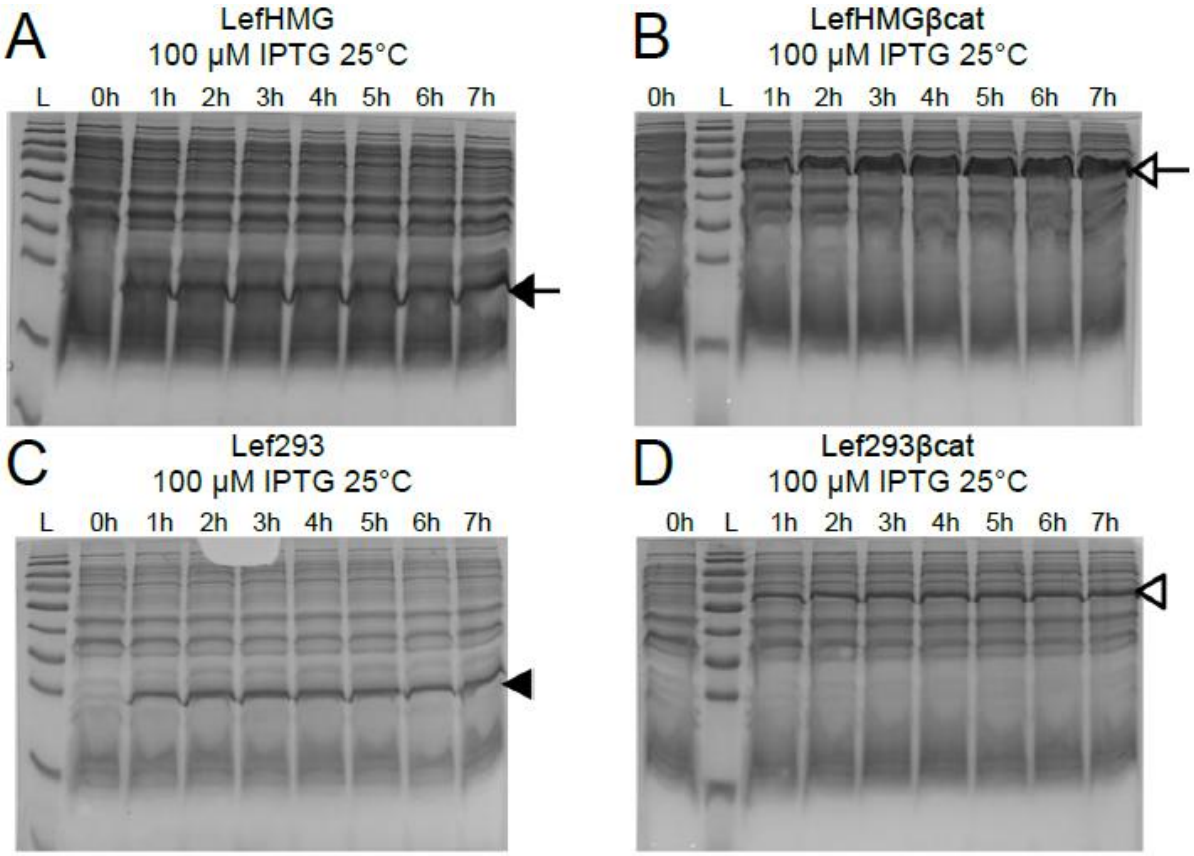


Figure 26: Time course experiments with reduced expression temperature. Protein expression was induced by 100 mM IPTG and the temperature during expression was lowered to 25°C. Samples were collected in one hour intervals. To give comparable results the samples were diluted to the same OD_{600} value (Figure 25). The arrow indicates the bands of LefHMG, the empty arrow LefHMGβcat, the arrowhead Lef293 and the empty arrowhead that of Lef293βcat. The uninduced control is denoted as 0h. As ladder (L) Thermo Scientific's PageRuler™ was used.

In fact the reduced expression rate can be high as seen in the case of LefHMGβcat (Figure 26B) were the band increases steadily during the first three hours. The other three proteins seem to reach the maximum concentration again after the first hour. Since all the SDS-PAGE analyses are comparable, because the samples were all diluted in the same way, it could be concluded that the thinner bands in the 25°C expression cultures mean a decrease in the overall protein concentration and not due to the lower OD_{600}. This could also help keeping the proteins from aggregating to inclusion bodies.

4.1.2.5.3 Solubility of proteins expressed at 25°C

To check if the fraction of soluble protein has increased the cells were lysed by sonication in TN-T buffer and the soluble and insoluble fractions were separated by centrifuging. Two different time points, after three and after six hours expression, were analyzed. The separated fractions were analyzed by SDS-PAGE.

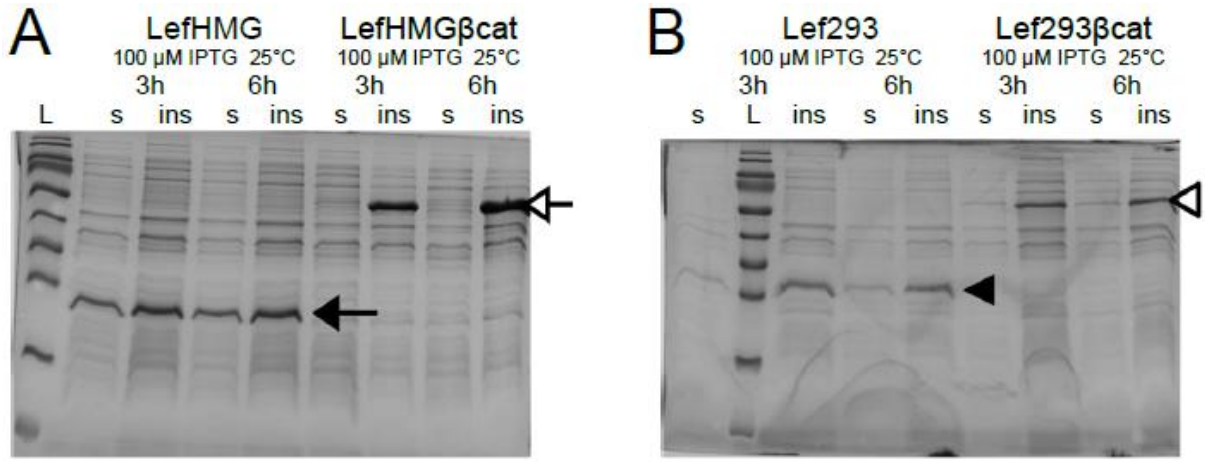


Figure 27: Solubility test of proteins expressed at a reduced temperature of 25°C and at different time points. Bacteria from the time course experiment were resuspended and sonicated in TN-T lysis buffer, soluble (s) and insoluble (ins) protein fractions were separated by centrifugation and loaded onto the gels. TN-T buffer consists of 20 mM Tris pH 7.5, 50 mM NaCl and 0.2% Tween20. The arrow indicates the bands of LefHMG, the empty arrow that of LefHMGβcat, the arrowhead points at the bands of Lef293 and the empty arrowhead to that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used.

This analysis showed (Figure 27) that the proteins are again soluble in TN-T buffer, but it cannot clearly be stated that the solubility had increased. The overall protein amounts loaded on these gels are the same as the amounts loaded on the other gels. These gels appear weaker than the others because they were destained for a longer time. This can also be noticed in reduced smear and brighter background bands.

Although the increased solubility could not be shown the proteins expressed at 25°C were taken for further experiments.

One possible explanation is that the protein extraction was still not complete, because not all cells were broken by sonication. Therefore 1mg/mL lysozyme was added to the lysis buffer to aid the disruption of the bacterial cell walls. To see if Tween20 in the lysis buffer has an effect on the activity of lysozyme, the cells were also lysed in the same buffer without Tween20.The resuspended E. coli cells from the 25°C time course were sonicated, separated and then loaded onto a SDS-PAGE.

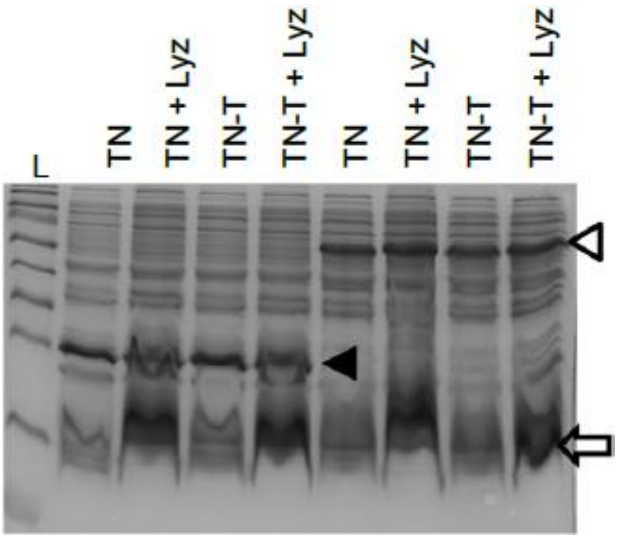


Figure 28: Analysis of crude extracts of the expressed Lef293 and Lef293βcat proteins obtained by sonication with and without the addition of lysozyme. TN stands for the protein lysed in 20 mM Tris pH 7.5, 50 mM NaCl and TN-T for the addition of 0.2% Tween20 to the TN lysis buffer. The +Lyz indicates the addition of 1 mg / mL lysozyme to the lysis buffers The arrowhead points at the bands of Lef293 and the empty arrowhead at that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used

The result of this experiment (Figure 28) was that additional cell disruption by lysozyme had no effects and was therefore omitted. Surprisingly Tween20 seems not to be responsible for the increased solubility, because the bands of the lysis buffer with and without Tween20 had the same strength for both Lef293 and Lef293βcat. Therefore Tween20 was no longer added to the lysis buffer.

4.1.2.5.4 Purification with His Mag Sepharose™ Ni

To remove the other E. coli proteins that may interfere with the binding site selection the target proteins were purified from the crude extracts. The proteins were again purified by metal ion affinity chromatography, but this time in the form of magnetic beads. This native protein purification method has some important advantages over the gravity flow columns used before. First of all the time used for the purification is reduced drastically because no long gradual elution from the column is required and there is no need for refolding during dialysis. Second, the protein amounts needed for the binding site selection are very small, in the range of 100-200 ng per reaction, so there is no need for larger scale purification methods like gravity flow purification. The third advantage becomes clear by looking at the process of the binding site selection: DNA binds specifically to the protein and is then purified by binding the protein to magnetic beads and washing off all non-specific DNA. Since the proteins are already bound to the magnetic beads any losses during this step in the binding site selection can be minimized.

To reduce binding of unspecific proteins, low concentrations of imidazole are added to the lysis buffer right before the binding reaction. The needed concentration of imidazole had to be determined first. Therefore, different imidazole concentrations were added to the lysis buffer and the protein was bound to the magnetic beads, washed with lysis buffer containing the same concentration of imidazole, then eluted and analyzed by SDS-PAGE in contrast to the remaining protein.

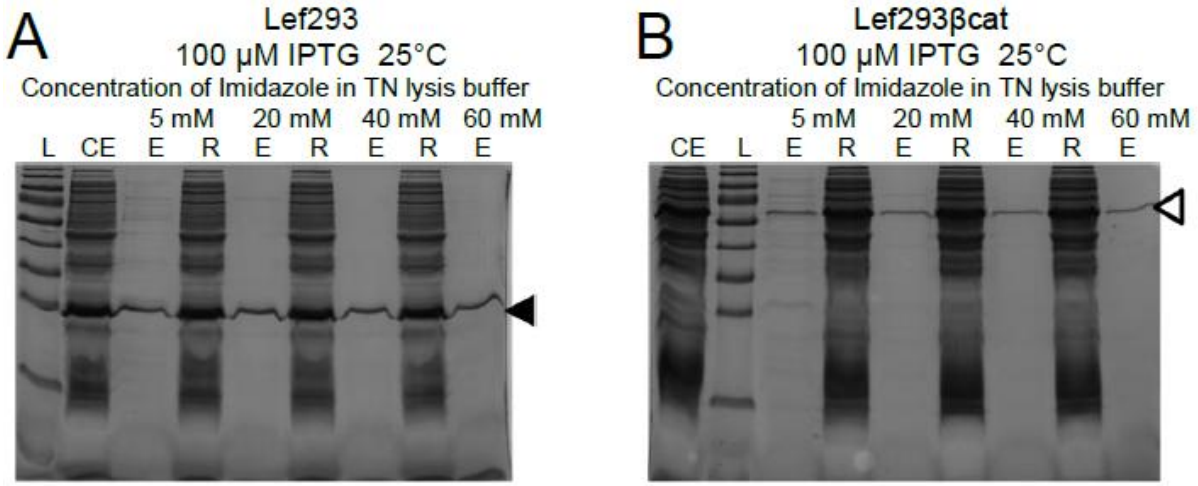


Figure 29: Purification of Lef293 and Lef293βcat with different imidazole concentrations in the TN lysis buffer to reduce unspecific protein binding. 5 mM-60 mM imidazole was added to the lysis buffer to prevent the unspecific binding of proteins. Bound proteins were then washed with lysis buffer containing the same amount of imidazole, eluted (E) and analysed in contrast to the protein remaining in the crude extract (R). For 60 mM imidazole in the lysis buffer only the eluted protein is shown. CE stands for the crude extract of the protein. The TN lysis buffer consists of 20 mM Tris pH 7.5, 50 mM NaCl. The arrowhead points at the bands of Lef293 and the empty arrowhead at that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler™ was used

This analysis showed (Figure 29) that the addition of 20 mM imidazole is enough to reduce unspecific binding. And that the ratio of the bound protein to unbound, residual protein is very small, indicating, that higher amount of magnetic bead suspension per volume crude extract could be used.

From the time course experiments, solubility studies and the purification test the parameters for the preparative expression cultures were specified as:

- Expression in E. coli BL21(DE3)RIL
- Induction of expression with 100 µM IPTG at OD_{600} 0.4-0.6
- Temperature during expression lowered to 25°C
- Harvesting proteins after three hours expression
- Cell lysis with sonication in TN buffer
- The use of magnetic beads for protein purification
- Addition of 20 mM imidazole to the lysis buffer to reduce unspecific protein binding to the magnetic beads

4.1.2.6 Preparative protein expression and purification under native conditions

With all parameters defined in previous experiments, the preparative protein expression and purification was done. Because the EMSA of the E. coli proteins, purified under denaturing conditions (Figure 21), showed that LefHMG and LefHMGβcat don't bind to DNA, no preparative protein expression and purification was done.

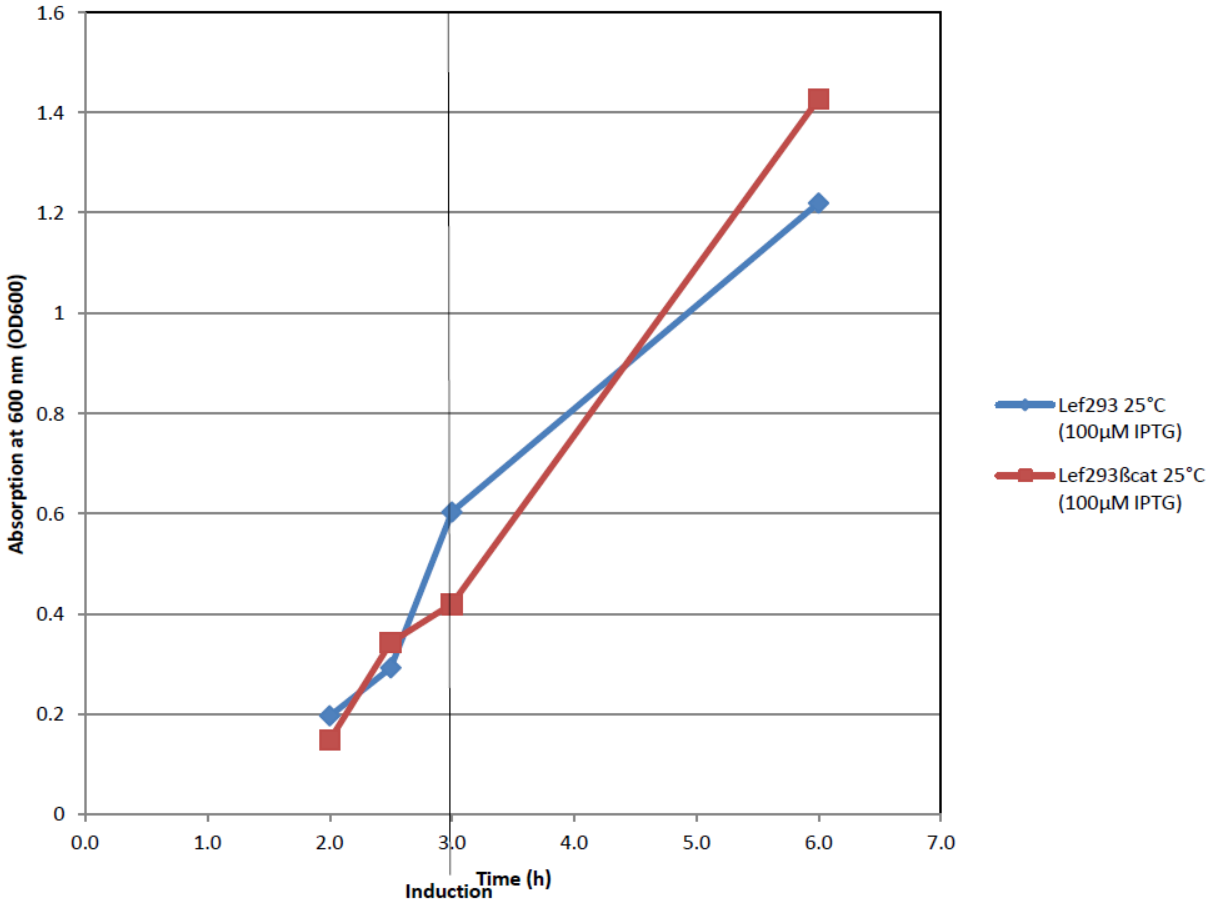


Figure 30: Cell density of the Lef293 and Lef293βcat expression cultures. The cell density was measured photometrically at a wavelength of 600nm. The vertical line indicates the time point of induction. Protein expression was induced by the addition of 100 µM IPTG. The proteins were expressed at 25°C and the cells were harvested after three hours after induction.

The harvested cells were then diluted according to their OD_{600} (Figure 30) in TN-buffer, aliquoted and stored at -80°C. One aliquot was thawed and purified by magnetic beads.

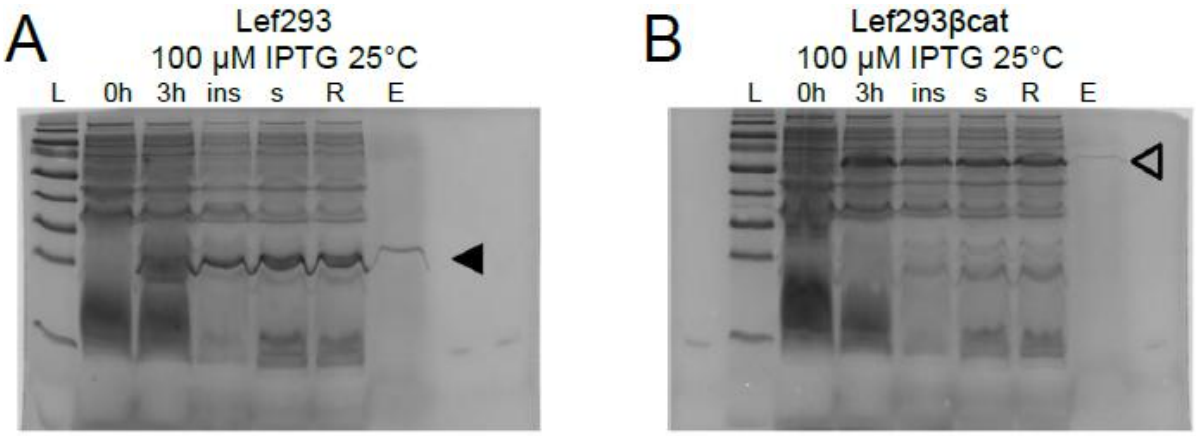


Figure 31: The course of Lef293 and Lef293βcat from expression to purification. Both gels show the uninduced (0h) and the induced cell culture at harvest (3h). Further, the insoluble (ins) and soluble (s) fraction of the crude extract, the protein (R) that is remaining after the purification and finally the purified protein (E) eluted from the magnetic beads. TN was used as lysis buffer and 20 mM imidazole prevented unspecific protein binding. The arrowhead points at the bands of Lef293 and the empty arrowhead at that of Lef293βcat. As ladder (L) Thermo Scientific's PageRuler^{TM} was used

Figure 31 shows the SDS-PAGE of all steps from protein expression, extraction to purification.

4.1.2.6.1 Quantification of the protein concentration by Bradford assay

To determine the concentration of the purified proteins a Bradford assay was carried out. Because the gel shown in Figure 31 indicated a low amount of bound protein bigger amounts of magnetic bead suspensions were used to elute enough protein for the Bradford assay. In the end proteins from 40 µL magnetic bead suspensions were eluted in 500 mM imidazole to get reasonable results.

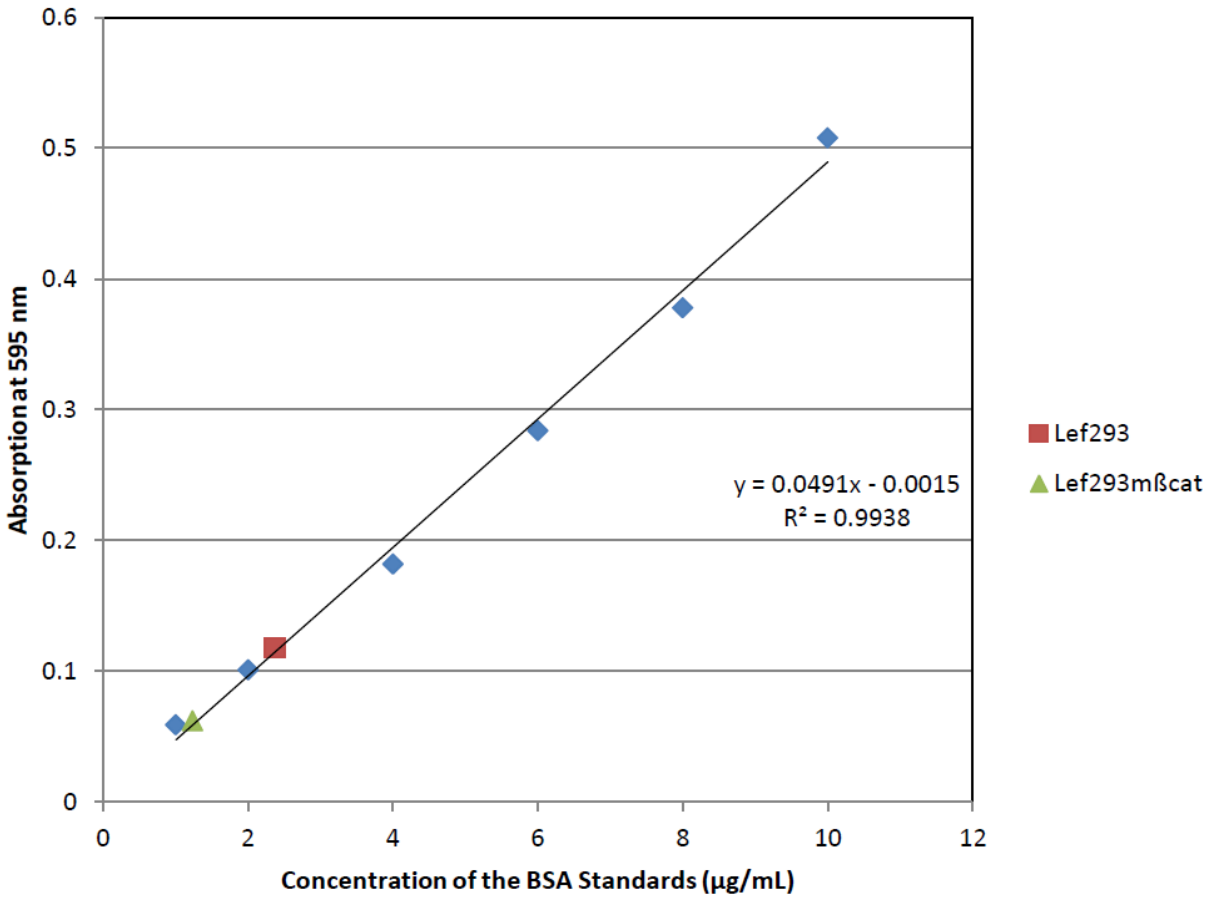


Figure 32: Estimation of the concentration of Lef293 and Lef293βcat purified by magnetic beads by the Bradford assay. Blue diamonds are values of the BSA standard curve connected by a trend line. The equation and the coefficient of determination of the trend line are shown. The measured values of Lef293 (red square) and Lef293βcat (green triangle) proteins, each eluted from 40 µL of magnetic bead solution, are depicted on the trend line. The BSA standard contained the same amount of imidazole used for the elution of the proteins.

The graph shown in Figure 32 depicts the trend line of the BSA standard curve, its equation and coefficient of determination (R^2). It also shows the measured values of Lef293 (red square) and Lef293βcat (green triangle) each eluted from 40 µL magnetic bead suspension.

Protein	Mass concentration	Molar concentration
Lef293	59.3 ng / µL magnetic beads	2,85 fmol / µL magnetic beads
Lef293βcat	30.8 ng / µL magnetic beads	0,68 fmol / µL magnetic beads

Table 7: Mass and molar concentrations of purified Lef293 and Lef293βcat protein bound to magnetic beads determined by the Bradford Assay.

The obtained bound protein amount per volume of magnetic beads (Table 7) seems low, but only 100-200 ng protein is needed per round of binding site selection. Considering this fact the protein amount is in the ideal range, by using between 1.7 µL and 6.5 µL magnetic bead suspensions per binding site selection reaction.

4.1.3 Protein expression in mammalian cell culture

To obtain proteins from their most natural environment, they were expressed in mammalian cell culture. This was necessary, because it cannot be ruled out that any of the several possible posttranslational modifications of Lef1 or β-catenin has effects concerning DNA-binding.

4.1.3.1 Nuclear extracts from HeLa cells

LefHMG, Lef293 and their β-catenin fusion proteins were expressed in HeLa cells. Therefore HeLa cell culture plates were transfected with the various pMC vectors (see Table 5) and incubated for 24 hours at standard culturing conditions (37°C, 5% CO_2 in a saturated water vapour atmosphere). The cells were harvested by scraping, lysed, then the nuclei were spun down and the cytoplasmic proteins were discarded. The nuclear proteins were extracted by a high salt buffer and analysed by EMSA.

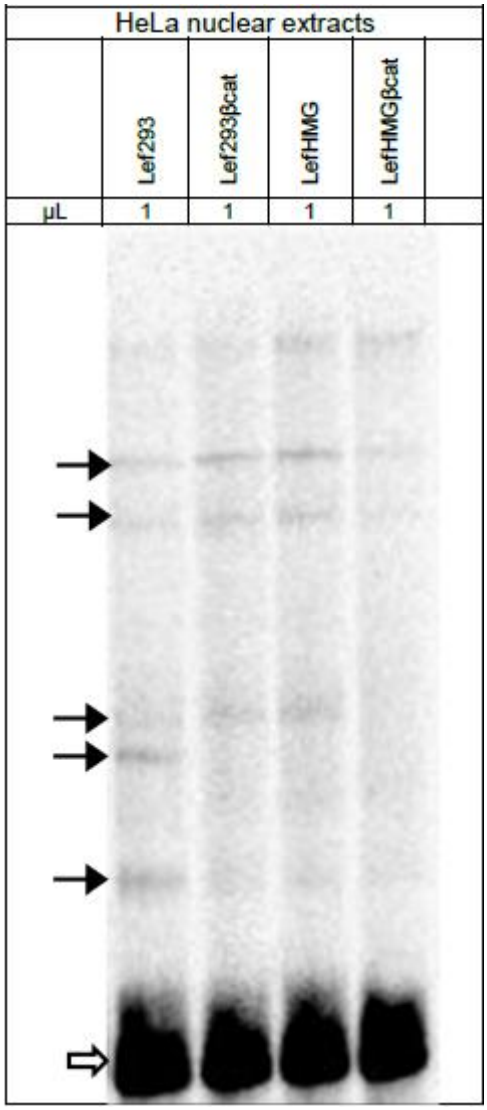


Figure 33: Electrophoretic mobility shift assay of nuclear extracts from transfected HeLa cells. As expression vectors the corresponding pMC vectors (see Table 5) were used. As probe a radiolabeled Tcf/Lef consensus binding site was used. After incubating the proteins and the probe on ice the reaction mixtures were loaded on a 6% polyacrylamide gel and run at 290 V for 3 h. The results were obtained by exposing an x-ray film to the gel. The empty arrow indicates the free unbound probe and the arrows some unspecific shifts.

The results of the electrophoretic mobility shift assay are shown in Figure 33. None of the four nuclear extracts showed specific DNA binding properties. Low transfection efficiency could cause low protein yield. The transfection efficiency was determined by transfection of a control plasmid containing a gene coding for the green fluorescent protein (GFP) into an equal cell culture plate, with the same cells and the same conditions. The cells expressing

GFP were evaluated by florescence microscopy and indicated a transfection efficiency of 15%.

4.1.3.2 Nuclear extracts from COP-8 cells

Because the transfection efficiency was only about 10-15% in the first few experiments, a different cell line than HeLa was used. This polyoma virus-transformed mouse fibroblast cells, COP-8, transfected with a plasmid containing a polyoma virus origin of replication, replicate the plasmid thus increasing the intracellular plasmid copy number. A high plasmid copy number means in general a higher yield of expressed protein. To use this ability the protein coding DNA had to be cloned into a third expression vector denoted as pKC. After cloning, COP-8 cells were transfected with an efficiency of 30%. Nuclear proteins were extracted and analysed by EMSA.

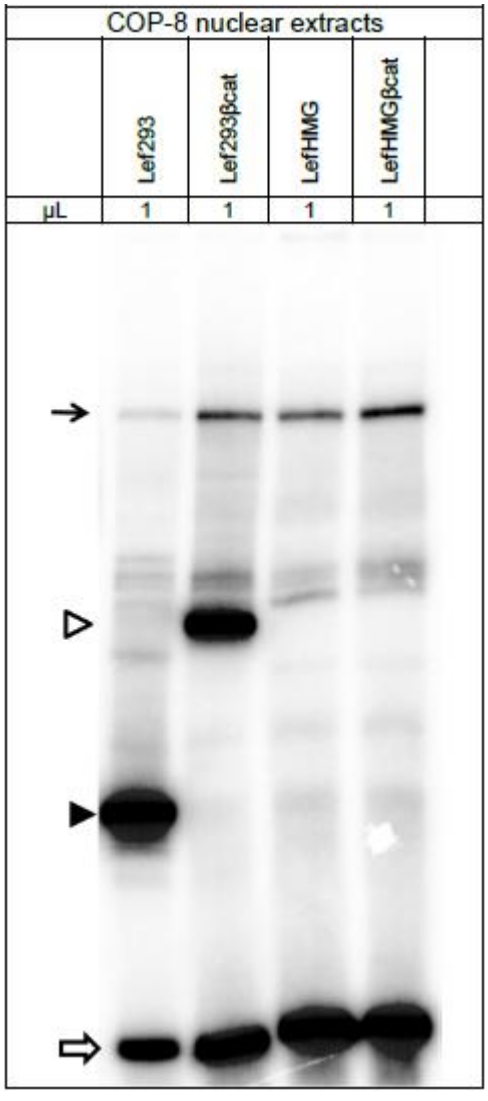


Figure 34: Electrophoretic mobility shift assay of nuclear extracts from transfected COP-8 cells. As expression vectors the corresponding pKC vectors (see Table 5) were used. As probe a radiolabeled Tcf/Lef consensus binding site was used. After incubating the proteins and the probe on ice the reaction mixtures were loaded on a 4% polyacrylamide gel and run at 230 V for 3.5 h. The results were obtained by exposing an x-ray film to the gel. The empty arrow indicates the free unbound probe and the arrow unspecific DNA shifts. The arrowhead points at the specific Lef293 shift and the empty arrowhead at the Lef293βcat shift. The lanes were LefHMG and LefHMGβcat ran show again no specific DNA shift.

The EMSA clearly shows (Figure **34**) that nuclear extracts containing Lef293 and Lef293βcat are able to bind and shift DNA. LefHMG and LefHMGβcat showed again no DNA binding properties. Therefore only Lef293 and Lef293βcat were selected for the binding site selection.

4.2 Multimerization of the selected Binding Sites

In classical binding site selection experiments the enriched oligonucleotides containing the binding sites are cloned into a backbone vector and then sequenced. This method yields the sequence of one binding site per sequenced positive clone. To increase the yield of sequences per clone, a method to multimerize the selected binding sites before cloning into the backbone vector was established. To achieve this, an oligonucleotide, containing 10 to 14 random nucleotides in the middle were the binding site selection would take place, and two ends that can be removed by digestion with a restriction enzyme, was designed. Therefore the oligonucleotides consist of the binding site part in between two ends. Throughout the different experiments the oligonucleotides had to be adapted according to the results (Table 8).

S1	*GCGAAGACGG* AGCTGCCGAG **NNNNNGNNNN** ACAGGC*AGCT* *CCGTCTTCGC* *CGCTTCTGCC* *TCGA*CGGCTC **NNNNNCNNNN** TGTCCGTCGA *GGCAGAAGCG*	30422.5 to 31143.1 Da
S1 Primer	GCGAAGACGG AGCTGCC	
S2	*CGGAAGACCC* ACGTCGCGAG **NNNNNGNNNN** ACAGGC*AGCT* *CCGTCTTCGC* *GCCTTCTGGG* *TGCA*GCGCTC **NNNNNCNNNN** TGTCCGTCGA *GGCAGAAGCG*	30422.6 to 31143.1 Da
S2 Primer1	GCGAAGACGGAGCTGCC	
S2 Primer2	CGGAAGACCCACGTCGC	
S3	*CGGAAGACGC* TGCAGGCGAG **NNTTTGNNNN** ACACCG*TGCA* *CGGTCTTCGC* *GCCTTCTGCG* *ACGT*CCGCTC **NNAAACNNNN** TGTGGCACGT *GCCAGAAGCG*	30539.7 to 31020.1 Da
S3 Primer1	CGGAAGACGC TGCAGGC	
S3 Primer2	GCGAAGACCG TGCACGG	
S4	*AGCTGC*TCGA GTACGACG**NN** **NNTTTGNNNN** **NN**ATCATGCC *TCGAGTTCGC* *TCGACGAGCT* CATGCTGC**NN** **NNAAACNNNN** **NN**TAGTACGG AGCT*CAAGCG*	33037.6 to 31178.2 Da
S4 Primer1	AGCTGCTCGAGTACGAC	
S4 Primer2	GCGAACTCGAGGCATGA	

Table 8: Different oligonucleotides used for the establishment of the method to multimerize the selected binging sites. The bold letters indicate the random part of the oligonucleotide were the binding of Lef would occur. The single nonrandom nucleotide inside the random part in S1 and S2 is used for the orientation of Lef on the random part and the four nonrandom nucleotides in S3 and S4 are conserved through Lef/Tcf binding sites. The blue colored nucleotides are cut off by restriction enzymes and referred to as ends.

4.2.1 1st Approach: Simultaneous digestion and ligation

The first approach was to simultaneously digest and ligate the oligonucleotide to form concatemers, a continuous DNA molecule that contains multiple copies of the same DNA sequences linked in series. Therefore the oligonucleotide S1 has two BpiI recognition sites on either end. The restriction digestion yields the selected binding site with 5' overhangs which can be easily ligated. Because BpiI is a restriction enzyme that cuts the DNA at a different point than the recognition site (Table 9), the binding sites can be ligated without the restriction enzyme cutting the DNA again at this site. If an end piece is ligated to the multimerized binding sites it is removed again by BpiI.

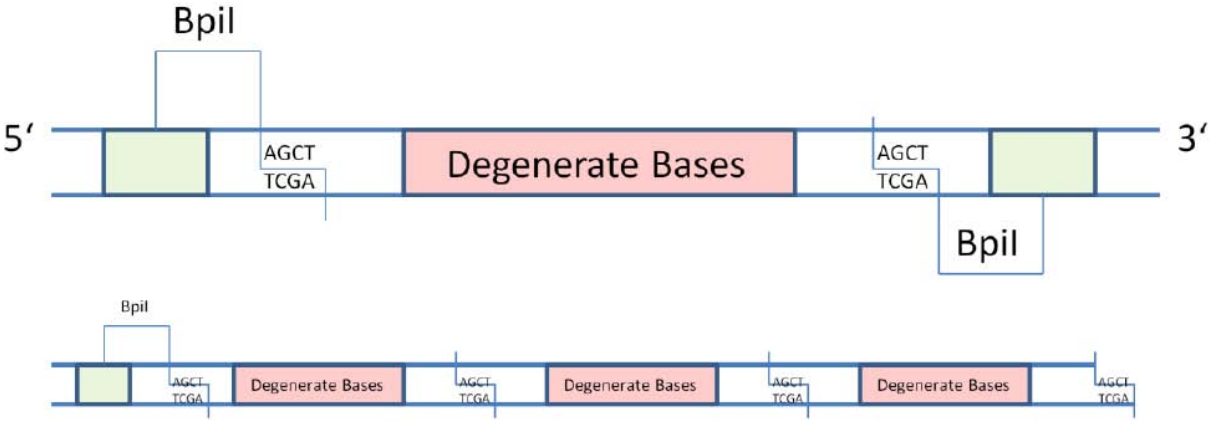


Figure 35: Schematic representation of S1 with one single oligonucleotide in the top and a concatemer, consisting of three ligated binding sites with one end still attached, below. The red areas indicate the degenerate nucleotides used for the binding site selection. The green areas show the recognition site of BpiI and the thin blue line points at the restriction site and the generated 5' overhangs.

The obtained concatemer with its 5' overhangs was then ligated into a backbone vector, in this case pBSII KS(+), a small cloning vector. This vector with the multimerized binding sites as insert was then transformed into E. coli TOP 10F'. Some colonies were then selected for miniprep. The obtained DNA was analysed by restriction digestion and then, if they looked promising, sequenced

Enzyme	Recognition site	Concentration
BpiI (Thermo)	5'...G A A G A C $(N)_2$↓...3' 3'...C T T C T G $(N)_6$↑...5'	10 U/ 10 µL reaction
XhoI (Thermo)	5'...C↓T C G A G...3' 5'...G A G C T↑C...5'	10 U/ 10 µL reaction
T4 DNA Ligase (Thermo)		1 U/ 10 µL reaction

Table 9: Enzymes used for the multimerization of the binding sites, their recognition and restriction site and the concentrations used in the different experiments.

4.2.1.1 Different times of digestion and ligation

As there was no information on how fast this concatenation reaction would take place, first it was necessary to determine the optimal period of time for the simultaneous digestion-ligation step (Figure 36). The idea was that the restriction enzyme activity outcompetes the ligase activity and cuts off all remaining end parts to yield concatemers with cohesive ends, which can be easily cloned into a backbone vector.

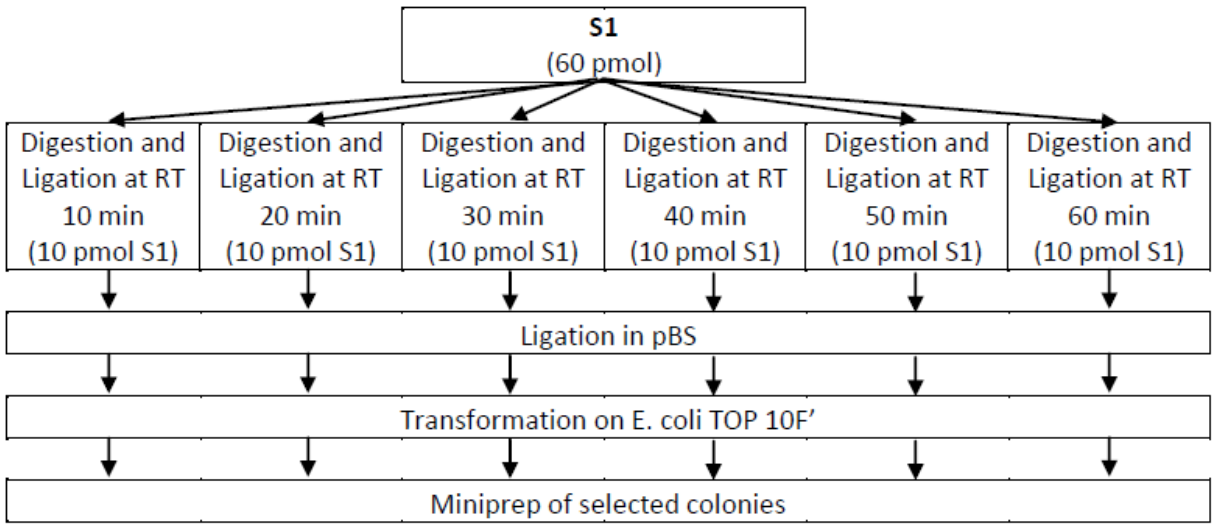


Figure 36: Experimental design to estimate the optimal simultaneous digestion-ligation time period.

Therefore six reactions were started in ten minute intervals, each containing 10 pmol of S1, the restriction enzyme BpiI and T4 Ligase (Table 9). After 60 minutes all reactions were heat inactivated at 80°C for 20 minutes. The newly formed concatemers were then ligated in pBSII KS(+) (Table 10), transformed in E. coli and plated on 1.5% agar plates containing LB media and ampicillin. Seven colonies of each time period were taken for miniprep and restriction digestion analysis, to verify if multimerized binding sites were inserted into the vector. The enzymes used for the restriction digestion analysis are shown in Table 10.

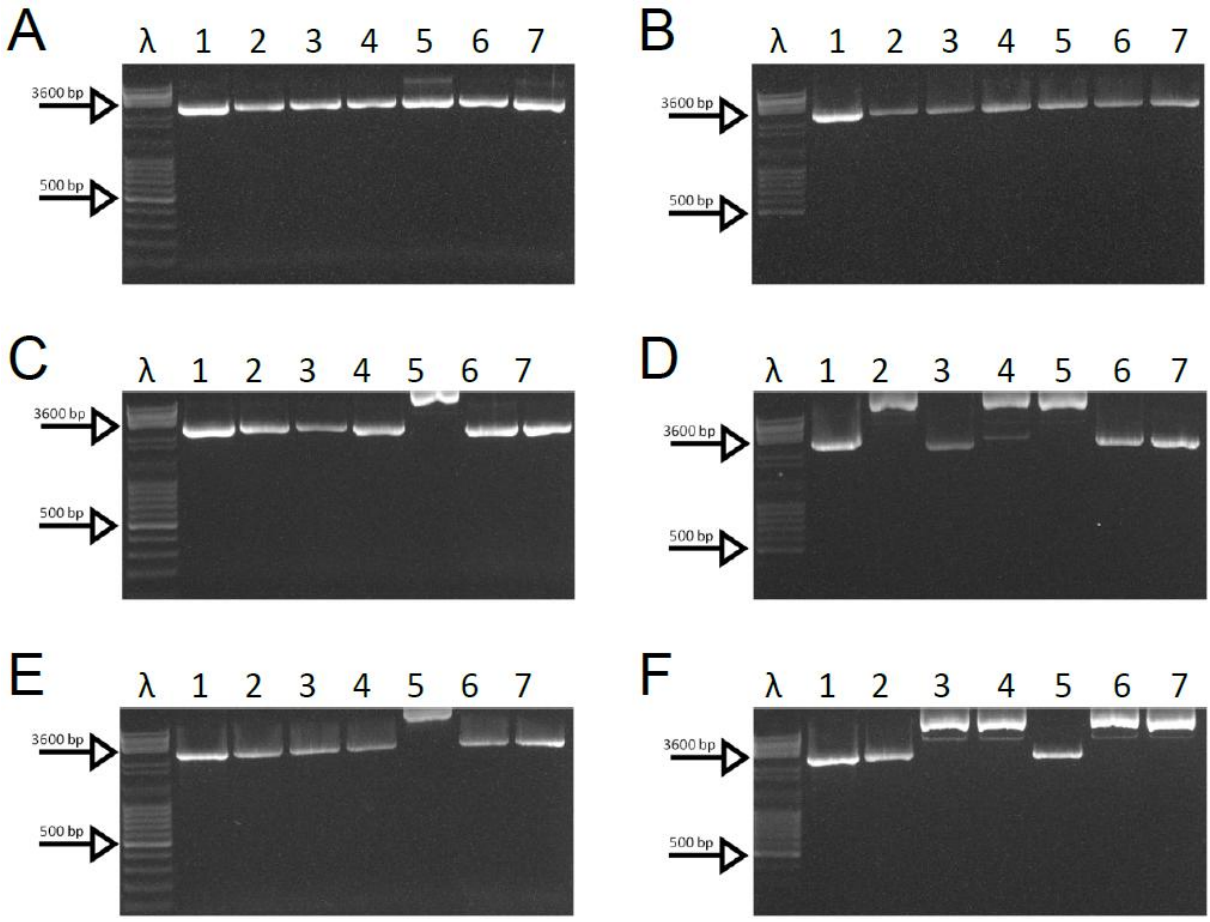


Figure 37: Agarose gel electrophoresis of digested minipreps of different digestion-ligation time spans. The gels from A to F depict the digested minipreps from 10min to 60 min digestion-ligation time, respectively. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

Figure 37 shows that none of the selected plasmids contained any multimerized binding sites, because all visible bands are at the height of the empty vector.

Vector	Enzymes	Concentration	Sequencing primer
pBSII KS(+) (HindIII)	BamHI / XhoI	10 U each / 10 µL rxn	TAATACGACTCACTATAGGG
pGemT®Easy	EcoRI	10 U each / 10 µL rxn	TAATACGACTCACTATAGGG
pUC19 (SalI)	PstI / XbaI	10 U each / 10 µL rxn	TGTAAAACGACGGCCAGTG

Table 10: Vectors used for transformation, the enzymes used for the restriction digestion analysis of the corresponding plasmids obtained by miniprep and the sequence of the primer used to sequence the insert of the vector. The restriction enzyme in braces in the vector column denotes the restriction site in which the concatemers were inserted.

4.2.1.2 Different temperatures and times of digestion and ligation

Because the first experiment failed to give positive results, a further experiment was designed. This experiment takes into account that the restriction enzyme BpiI and the T4 DNA Ligase have different optimal working temperatures. The optimal temperature for the BpiI digestion is 37°C and for the T4 DNA Ligase at 22°C. Therefore three temperatures were chosen, 37°C for optimal digestion, RT (about 23°C) for optimal ligation and 30°C as an intermediate temperature.

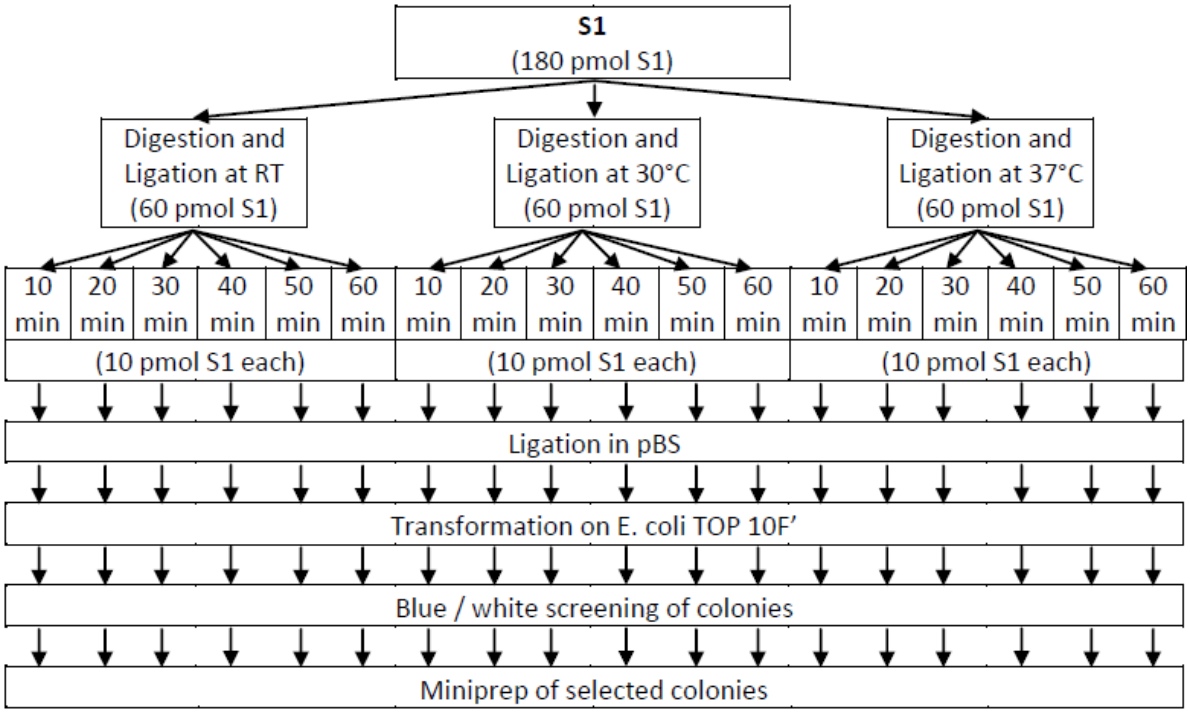


Figure 38: Experimental design to estimate the optimal simultaneous digestion-ligation time span at different temperatures for optimal digestion or ligation.

This experiment was performed the same way as the first one. Six reactions for each temperature were started in ten minute intervals and then after 60 minutes stopped by heat inactivation at 80°C for 20 minutes. The concatemers were again cloned into pBSII KS(+), transformed into E. coli TOP10F' and plated on 1.5% agar plates, this time containing LB media, ampicillin and IPTG and X-gal. Alpha complementation was used to improve the screening for positive colonies carrying an insert. Table 11 shows the numbers of blue and white colonies grown.

Temperature	Digestion/ligation time	Blue colonies	White colonies
23°C (RT)	10 min	5	0
	20 min	9	0
	30 min	45	1
	40 min	12	0
	50 min	6	0
	60 min	59	0
30°C	10 min	34	2
	20 min	12	1
	30 min	2	0
	40 min	4	0
	50 min	41	0
	60 min	22	1
37°C	10 min	15	0
	20 min	7	0
	30 min	9	0
	40 min	18	1
	50 min	29	0
	60 min	6	0

Table 11: Blue and white colonies obtained from different temperatures and time spans of simultaneous digestion and ligation.

To determine the fragment length of the positive (white) colonies, they were taken for miniprep and restriction digestion analysis (Table 10). The enzymes used for the restriction digestion analysis are shown in Table 10.

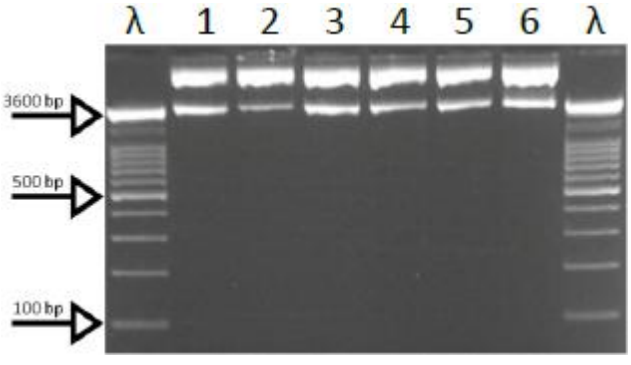


Figure 39: Agarose gel electrophoresis of digested minipreps of different temperature and digestion-ligation time spans. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described Material and Methods (3.2.5. Agarose gel electrophoresis). Lane1:23°C 30 min; Lane2+3: 30°C 10min; Lane4:30°C 20 min; Lane5: 30°C 60min; Lane6: 37°C 40min

The blue/white screening and the miniprep digestions showed that not only that there are very few positive colonies, but also that they are false positives, because no concatemers were incorporated into the vector.

4.2.1.3 PCR amplification of S1 concatemers

One possibility that no concatemers were cloned into the backbone vector is, that there are only a “few” concatemers and the odds of incorporating them are low. To overcome this, PCR amplification was done, right after the digestion-ligation step.

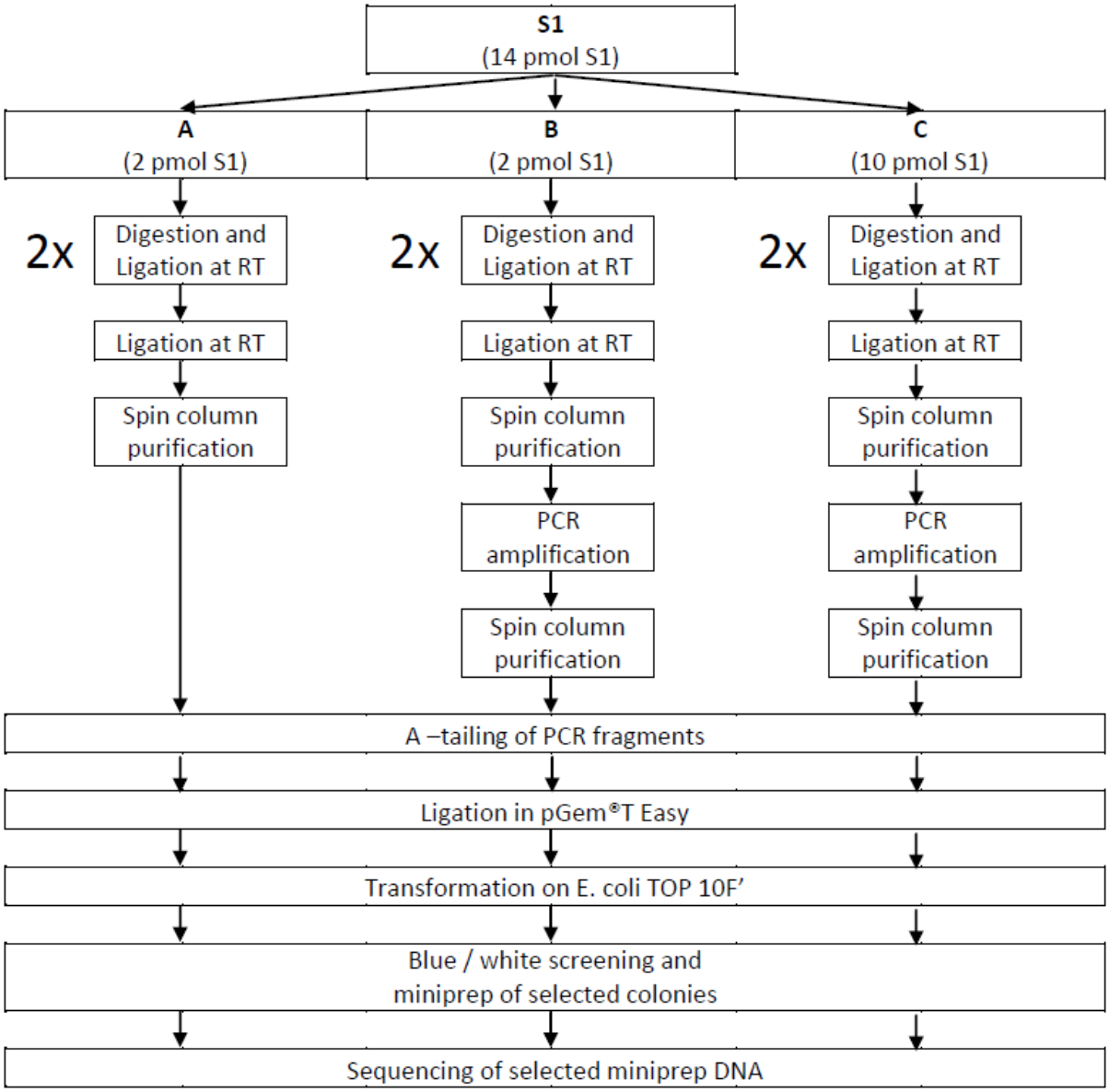


Figure 40: Experimental design of the PCR amplification of the concatemers. The difference between B and C is the starting concentration of oligonucleotide S1.

Three slightly different experiments were performed (see Figure 40). The starting molar amount of S1 was reduced in two experiments (Figure 40 experiment A and B) and kept the same in Figure 40 experiment C. In all three experiments the simultaneous ligation and digestion step was carried out twice to get a high number of multimerized binding sites. Both of these steps were performed at room temperature, because to be able to amplify the concatemers, the end parts, containing the primer sequences, have to be ligated to the multimerized binding sites. So it does not matter if the end parts are ligated directly after the digestion-ligation step. Subsequently to the digestion-ligation steps an extra ligation step was performed to ensure the ligation of the end parts to the concatemers. After the ligation of the ends to the concatemers a spin-column purification (for all spin-column purification procedures the MSB® Spin PCRapace Kit from INVITEK was used) with a cut-off of 80 bp

was done to get rid of the remaining ends. For every two ligated binding site, two end parts become redundant. A single end part is 14 bp and two ligated ends are 24 bp long and therefore too small to bind to and be purified by the spin column.

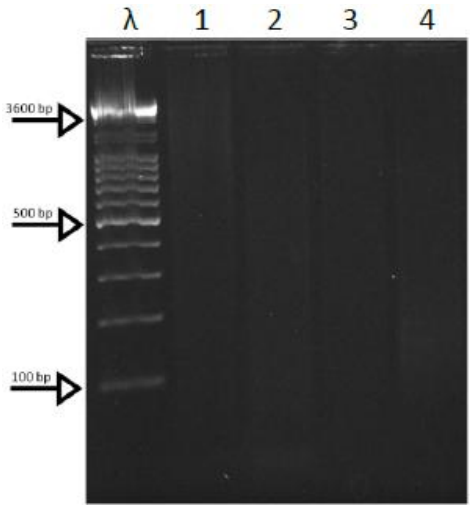


Figure 41: Agarose gel electrophoresis of the PCR amplified concatemers. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis). Lane1: PCR product of B; Lane2: purified PCR product of B; Lane3: PCR product of C; Lane4: purified PCR product of C

The purified concatemers (Figure 40 experiment B and C) were then amplified by 25 cycles of PCR, then again purified by spin columns. The crude and purified PCR products were analysed by agarose gel electrophoresis (Figure 41). Figure 41 showed a smear which could indicate DNA fragments of various sizes. Therefore the PCR products were then A-tailed and cloned into Promega's pGemT®Easy vector, a vector especially for the cloning of PCR products. The vectors were then transformed into E. coli TOP 10F' and plated on 1.5% agar plates containing LB media, ampicillin, IPTG and X-gal. The alpha-complementation showed a lot of positive (white) clones from the PCR amplified concatemers and even a few by the non PCR control. A few positive colonies were selected for miniprep and restriction digestion analysis. The enzymes used for the restriction digestion analysis are shown in Table 10.

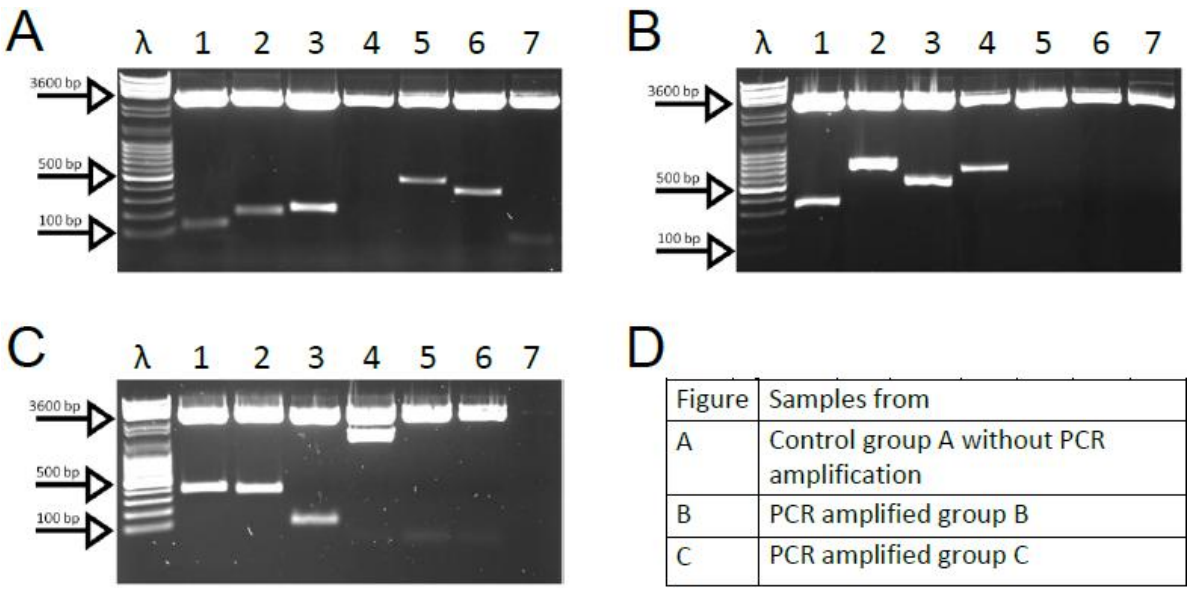


Figure	Samples from
A	Control group A without PCR amplification
B	PCR amplified group B
C	PCR amplified group C

Figure 42: Agarose gel electrophoresis of digested minipreps of PCR product ligated into pGemT®Easy. After PCR amplification of the concatemers, they were A-tailed and cloned into pGemT®Easy. The plasmid was then transformed in E. coli, incubated and screened for positive colonies. Seven positive colonies of each group were selected for miniprep and restriction digestion analysis, Figure A, B and C show the restriction digested miniprep DNA of the different groups shown in Figure D. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis).

The results of the miniprep digestion shown in Figure 42 showed that DNA fragments with various lengths could have been generated. It also showed that the PCR amplification did not increase the quality of the concatemers but the quantity of positive clones. The concatemers contained in plasmids isolated from the colonies Figure 42A5, 30B2 and 30C1 were sequenced. The obtained sequences (Appendix 7.4.1) revealed the multimerization of only one end part in all three sequenced plasmids.

Looking closely at the nucleotide sequence of S1 a horrible error came apparent: it has two 17 bases long complementary parts on each end, and therefore forms a perfect hairpin structure.

4.2.1.4 PCR amplification of S2 concatemers

A new oligonucleotide was designed (Table 8) and the experiment with the PCR amplification (Figure 40) was repeated with slight modifications (Figure 43). S2 was digested by BpiI at 37°C prior the simultaneous digestion and ligation step at room temperature (RT). The difference between Figure 43 experiment A, C and B, D is the molar amount of S2 at the start, which is 10 and 20 pmol respectively. And the difference between Figure 43 experiment A, B and C, D is a second simultaneous digestion-ligation step.

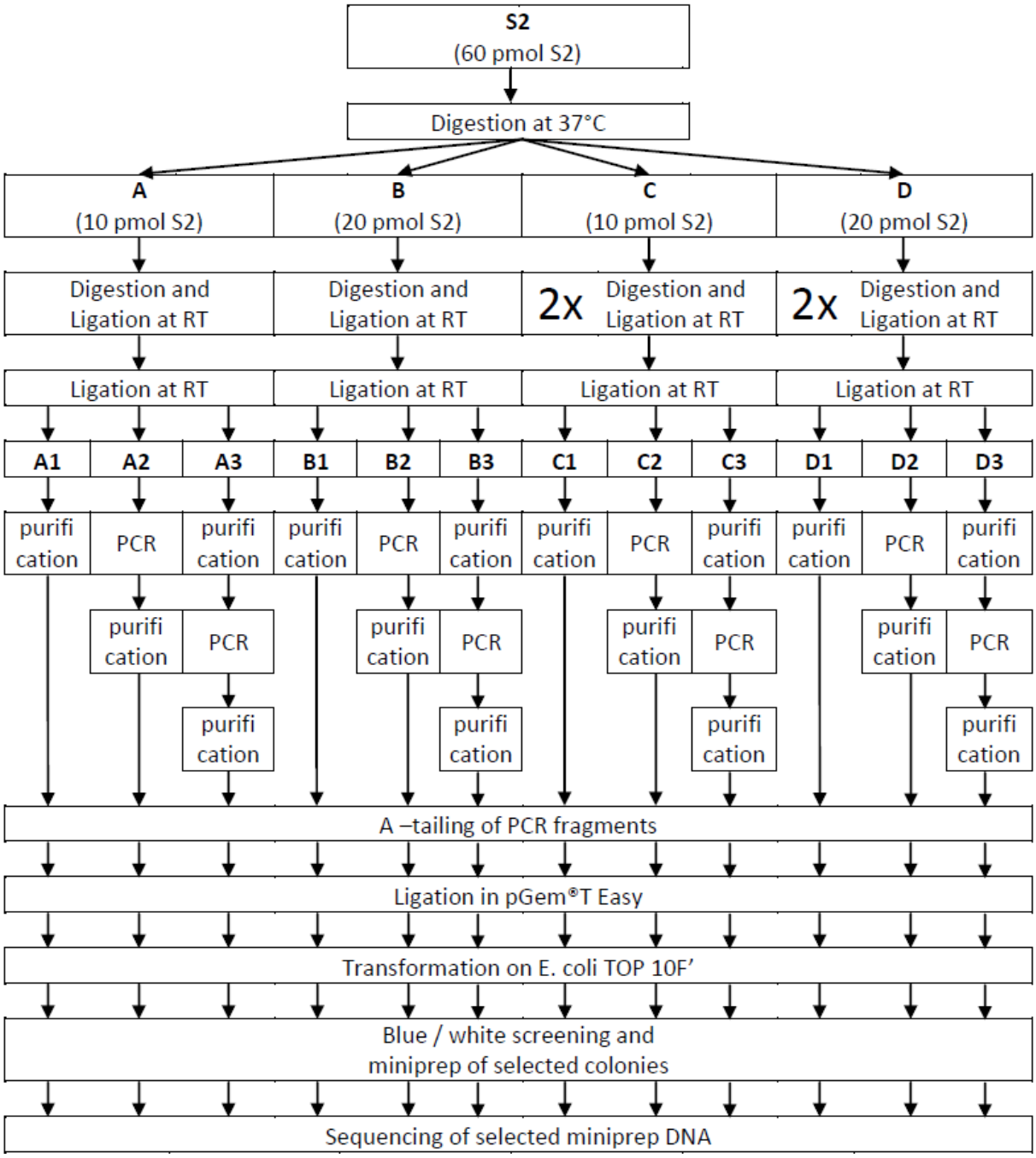


Figure 43: Design of the experiment to amplify the concatemers by PCR using the S2 oligonucleotide.

S2 was once or twice (see Figure 43) simultaneously digested and ligated at room temperature for at least one hour. As in the previous experiment the end parts were ligated to

the concatemers in an extra ligation step. The single experiments were then further subdivided into three groups. The first group was spin-column purified, A-tailed and cloned in the pGemT®Easy vector. The other two groups were amplified by 25 cycles of PCR, with the third group being spin-column purified before the amplification. The PCR products were analysed by agarose gel electrophoresis (Figure 44).

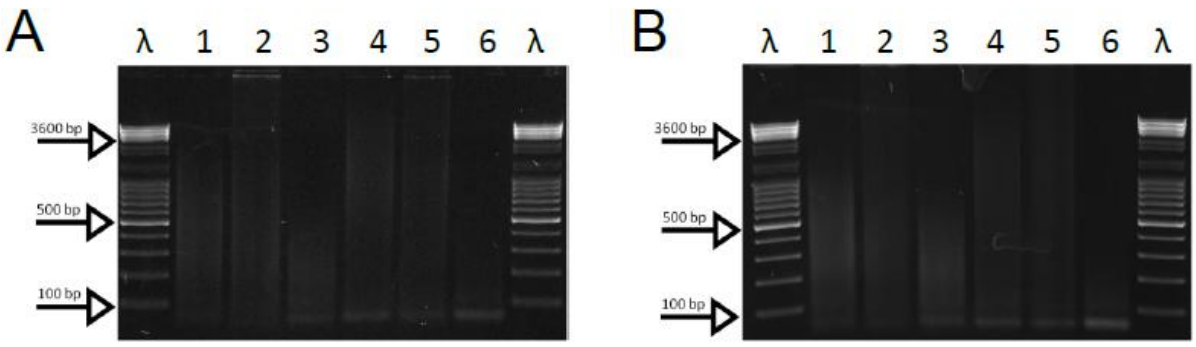


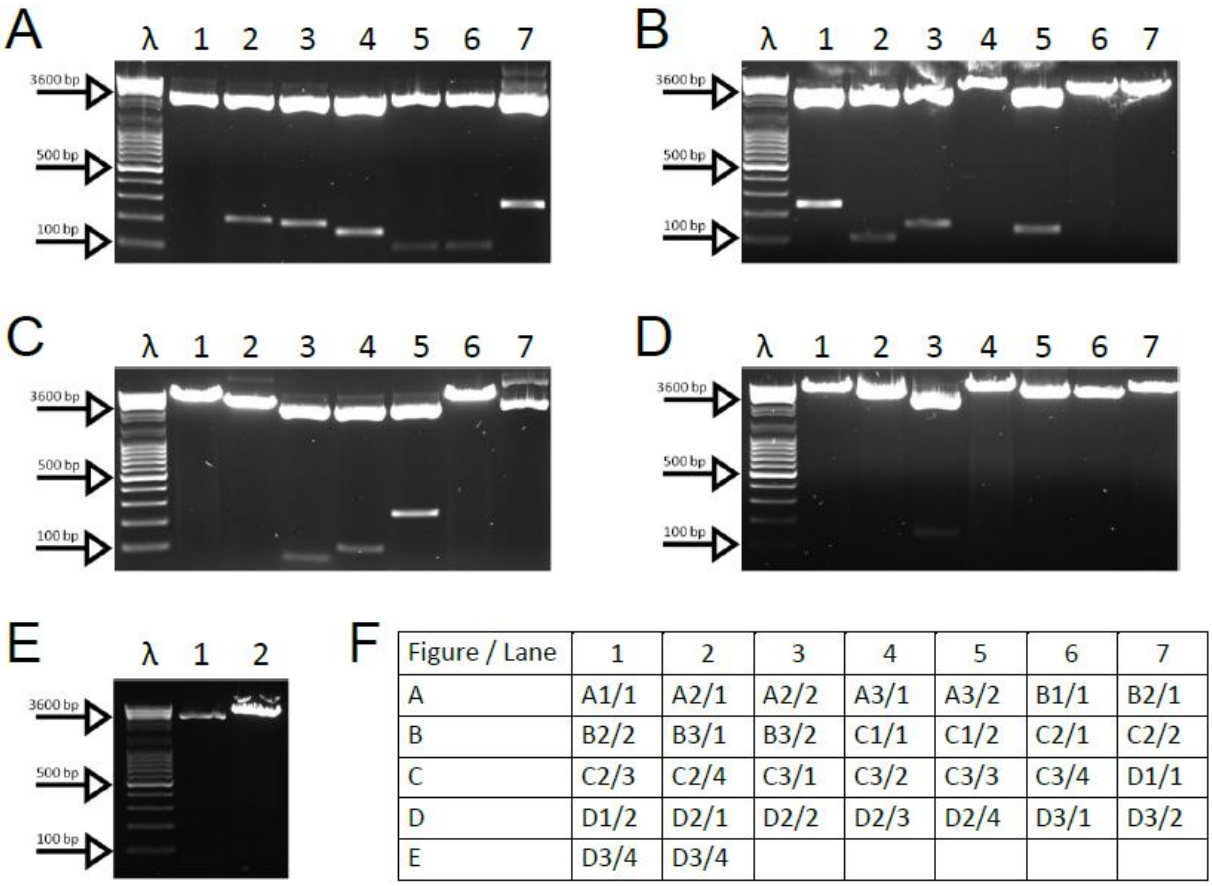
Figure 44: Agarose gel electrophoresis of the PCR amplified concatemers. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis). Figure A: Lane1: PCR product of A2; Lane2: purified PCR product of A2; Lane3: PCR product of A3; Lane4: PCR product of B2; Lane5: purified PCR product of A2; Lane6: PCR product of B3; Figure B: Lane1: PCR product of C2; Lane2: purified PCR product of C2; Lane3: PCR product of C3; Lane4: PCR product of D2; Lane5: purified PCR product of D2; Lane6: PCR product of D3

The amplified concatemers were then again spin-column purified, A-tailed and cloned in the pGemT®Easy vector. All of these vectors containing the products of the different groups of the experiments were transformed into E. coli TOP10F' and plated on 1.5% agar plates containing LB media, ampicillin, IPTG and X-gal. The number of grown colonies is shown in Table 12.

	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
blue	~300	93	~1000	~1000	108	~1000	~300	52	~1000	68	55	~1000
white	21	18	~100	12	15	10	64	11	~100	12	14	~100

Table 12: Numbers of the colonies grown on the plates. After the PCR amplification of the concatemers they were cloned into pGemT®Easy, then transformed into E. coli TOP10F' and plated onto 1.5% agar plates containing LB media, ampicillin and IPTG and X-gal for alpha complementation. The grown blue and white colonies were counted, or their number was estimated if the overall colony count was above 300.

As expected on the plates of group 1 and 2 were fewer colonies than on the plate of group 3. The ratio between blue and white colonies stays nearly the same, which could indicate that these 10-15% are false positive colonies. Nonetheless, for the experiment A and B were one colony picked for group 1, two for group 2 and two for group 3. And for the experiment C and D were one colony for part 1, four for part 2 and also four for part three selected for miniprep and restriction digestion analysis. The enzymes used for the restriction digestion analysis are shown in Table 10. The results of the restriction digestion analysis are shown in Figure 45.



F

Figure / Lane	1	2	3	4	5	6	7
A	A1/1	A2/1	A2/2	A3/1	A3/2	B1/1	B2/1
B	B2/2	B3/1	B3/2	C1/1	C1/2	C2/1	C2/2
C	C2/3	C2/4	C3/1	C3/2	C3/3	C3/4	D1/1
D	D1/2	D2/1	D2/2	D2/3	D2/4	D3/1	D3/2
E	D3/4	D3/4					

Figure 45: Agarose gel electrophoresis of digested plasmids of the different parts of the experiment in Figure 43. Figure 45A-E: Restriction digestion analysis of minipreps of positive colonies. Figure 45F: Loading scheme of the agarose gels. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis).

Five out of six digested plasmids from the first experimental group didn't show a band, which could indicate a low concatemer concentration. Therefore nearly all selected colonies were false positives. Digested plasmids from the other groups showed fragments of various sizes. Table 13 shows the sequenced miniprep DNA and the result of the sequencing.

Sequenced Mini	Result
A2/1 (A2)	5 binding sites
A2/2 (A3)	4 binding sites
B2/1 (A7)	1 binding site
B2/2 (B1)	6 binding sites
C3/3 (C5)	6 binding sites

Table 13: Sequenced miniprep DNA and the result given in the number of incorporated, multimerized binding sites. The sequenced miniprep is given in the first column with the corresponding gel and lane from Figure 45 in braces. The results column shows the number of binding sites found in the sequenced plasmid regardless of its integrity and correctness.

The complete sequences are shown in the appendix 7.4.2. The sequencing results shown in Table 13 show all found binding site sequences regardless if they were complete and correct or not. Nearly on all multimerized binding sites one end part was attached. One fourth of the obtained binding site sequences could not be used, because they were incomplete. All these missing sequence parts and additional end pieces which are also the primers could indicate that these multimerized binding sites are artefacts from the PCR amplification process. A

strong evidence for this theory was the sequence of miniprep DNA B2/1 (Figure 45 A7), were only one binding site was found. The following sequence is the repetition of exactly one half of the same S2 oligonucleotide. These findings were enough to omit this PCR amplification approach to increase concatemer length and cloning efficiency.

4.2.1.5 Limitation of multimerization through competition of excess end parts

A new strategy for the multimerization was needed, were the end parts are removed, so that they could not interfere with the formation of long concatemers. Figure 46 shows a scheme to this next approach.

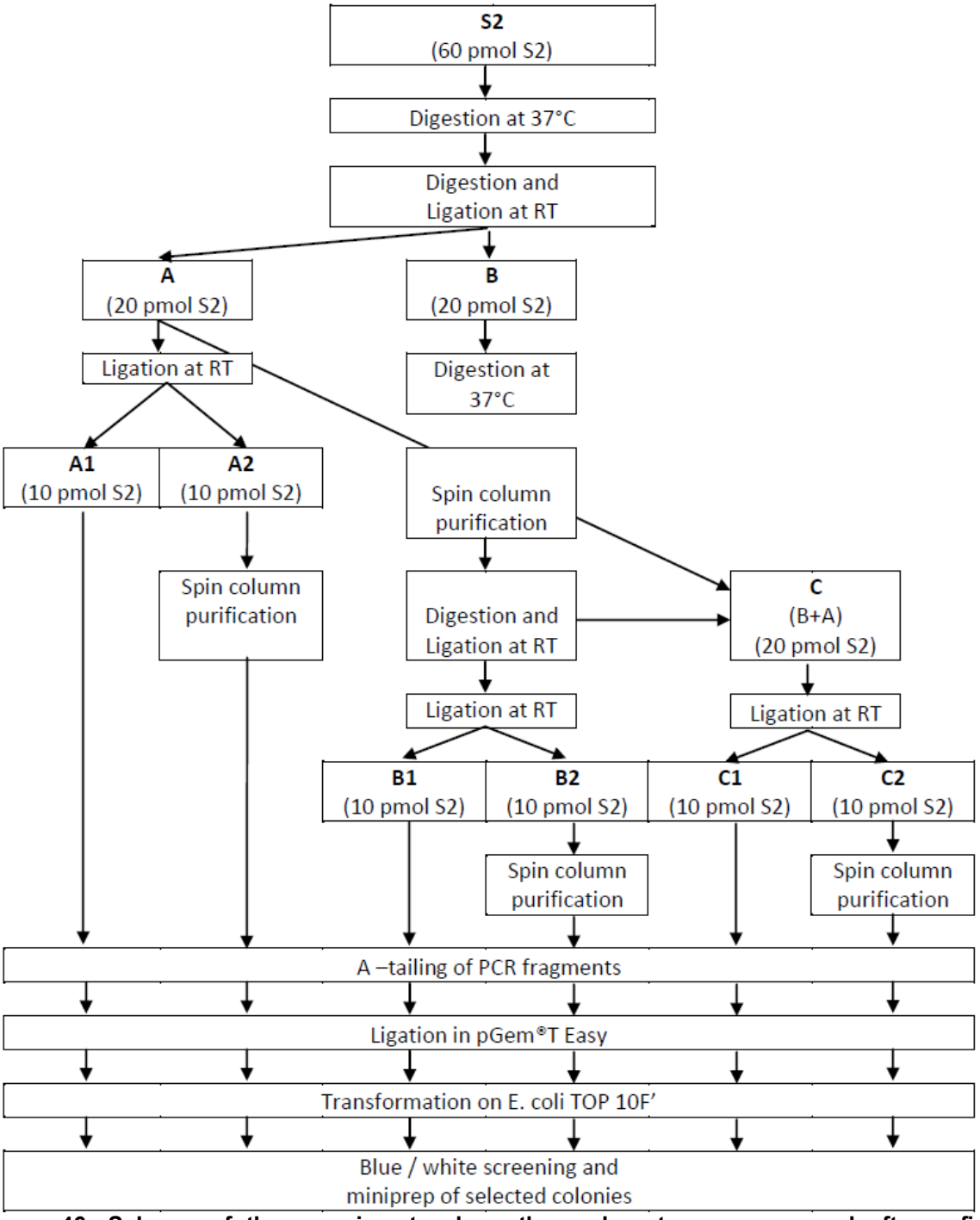
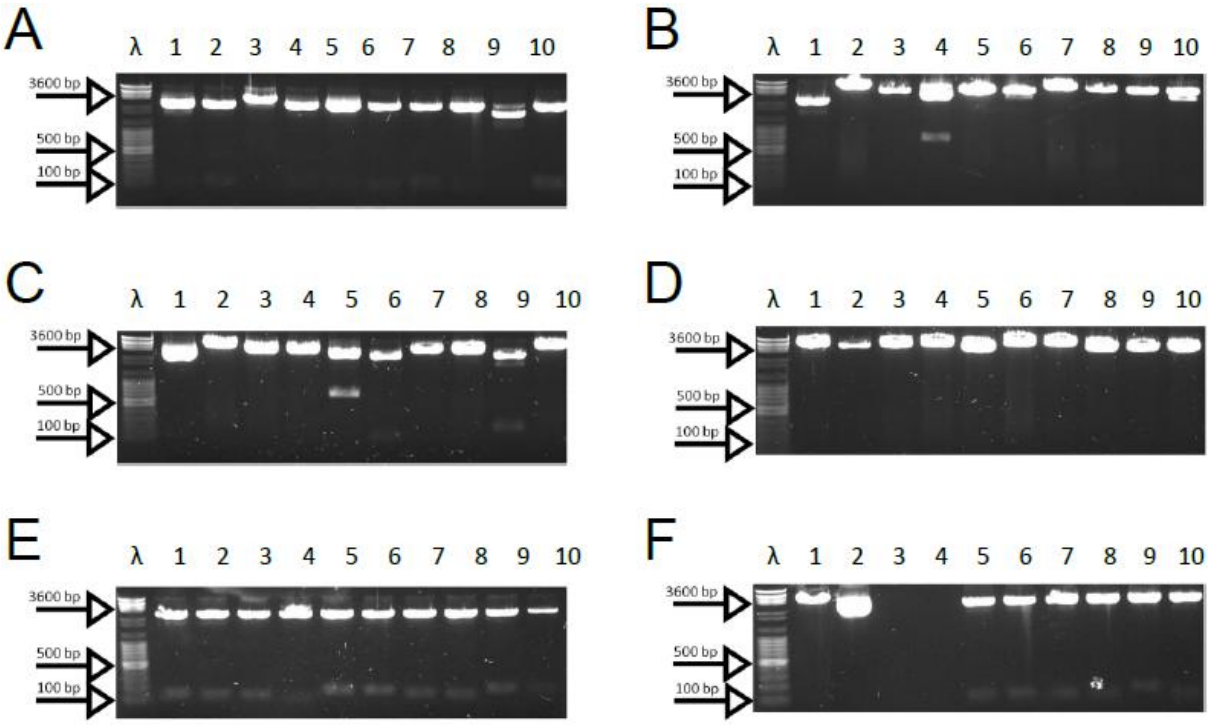


Figure 46: Scheme of the experiment, where the end parts were removed after a first multimerization step to get concatemers with cohesive ends, which can then be easily ligated to form even longer concatemers.

A predigestion of S2 at 37°C was performed to ensure that all binding sites are digested at the beginning of the multimerization step. Therefore the simultaneously digestion ligation step was performed at room temperature to favour the ligase. The formed concatemers were then subdivided into two groups. From group A, the control group, one third of the concatemers were saved for later use, and to the rest T4 DNA Ligase was added to ligate end parts to all concatemers. Group A was then again subdivided into two groups A1 and A2, whereas A2 was spin-column purified before A-tailing and cloning into pGemT®Easy. In group B the concatemers were digested again by BpiI at 37°C to ensure that no end parts are attached to the concatemers, so that all concatemers have cohesive ends. After the end parts were cut off they were removed by spin-column purification. The result of the purification was concatemers with cohesive ends without any end parts to interfere with the formation of very long concatemers. The ligation of this smaller concatemers should increase the length of every single concatemer. If there were, in spite of the purification, some end parts still residing, the ligation was simultaneous digest by BpiI. After this concatenation step a third group, group C, was created, consisting of one third of the concatenated group B and the untreated third of group A to provide enough end parts. Group B was then treated the same way as group A, it was ligated, so that the remaining end parts attach to the concatemers, subsequently it was divided in two subgroups with group B2 being spin-column purified before A-tailing and cloning into pGemT®Easy. Group B was second control group to see if there are still end parts remaining after the first purification step. Group C, with the long concatemers provided from group B and the end parts from untreated group A. After one third of each group was mixed together, they were ligated at room temperature and divided into two subgroups. Group C2 was spin-column purified before A-tailing and cloning into the pGemT®Easy vector. The first subgroup of every group was A-tailed as they were and cloned in pGemT®Easy. All vectors were then transformed into E. coli TOP10F' and plated on 1.5% agar plates containing LB media, ampicillin, IPTG and X-gal. All plates contained nearly the same number of colonies with the same ratio of positive to negative (data not shown).

Therefore no conclusion could be made about the effectiveness of the purification step. Seven white colonies from each group were selected for miniprep and restriction digestion analysis (Figure 47). The enzymes used for the restriction digestion analysis are shown in Table 10.



G

Figure	A	B	C	D	E	F
Group of experiment	A1	A2	B1	B2	C1	C2

Figure 47: Agarose gel electrophoresis of restriction digestion fragments of the selected minipreps form each group. From each group, seven colonies were selected for miniprep. The obtained plasmids were then restriction digested to verify the insert length. Figure 47A-F: The loading scheme of the gels is shown in Figure 47G. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrow. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis).

Figure 47 shows that, only the control groups A and B, have each one band of reasonable size. Both of the A subgroups showed (Figure 47A and B), as expected, smaller inserts in nearly all analysed plasmids. There were no large inserts in group C, only small fragments could be found in nearly every plasmid. But the overall concatemer lengths were not big enough. The average size of the bands (~200 bp) indicates only five to six multimerized binding sites. Therefore no miniprep DNA was sequenced.

4.2.1.6 Reduction of circularization by PEG and blunt end ligation

It was not yet clear, if the small concatemer length resulted from DNA circularization or just from a low multimerization degree. In the following experiment the digestion-ligation step at different temperatures was picked up from a previous experiment (4.2.1.2). In addition polyethylene glycol (PEG) was added to the reactions to increase intermolecular ligation and to reduce intramolecular ligation, the circularization of concatemers[122,123,124,125,126,127]. In previous experiments, were the end parts were removed, no long concatemers could be obtained. Therefore a new strategy was tested, no simultaneously digestion and ligation reaction to multimerize the binding sites, instead the oligonucleotides were not digested but ligated together blunt ended. Because PEG enhances the effectiveness of blunt end ligation, it was also done in the presence of 10% PEG[128,129]. A scheme of the whole experiment is shown in Figure 48.

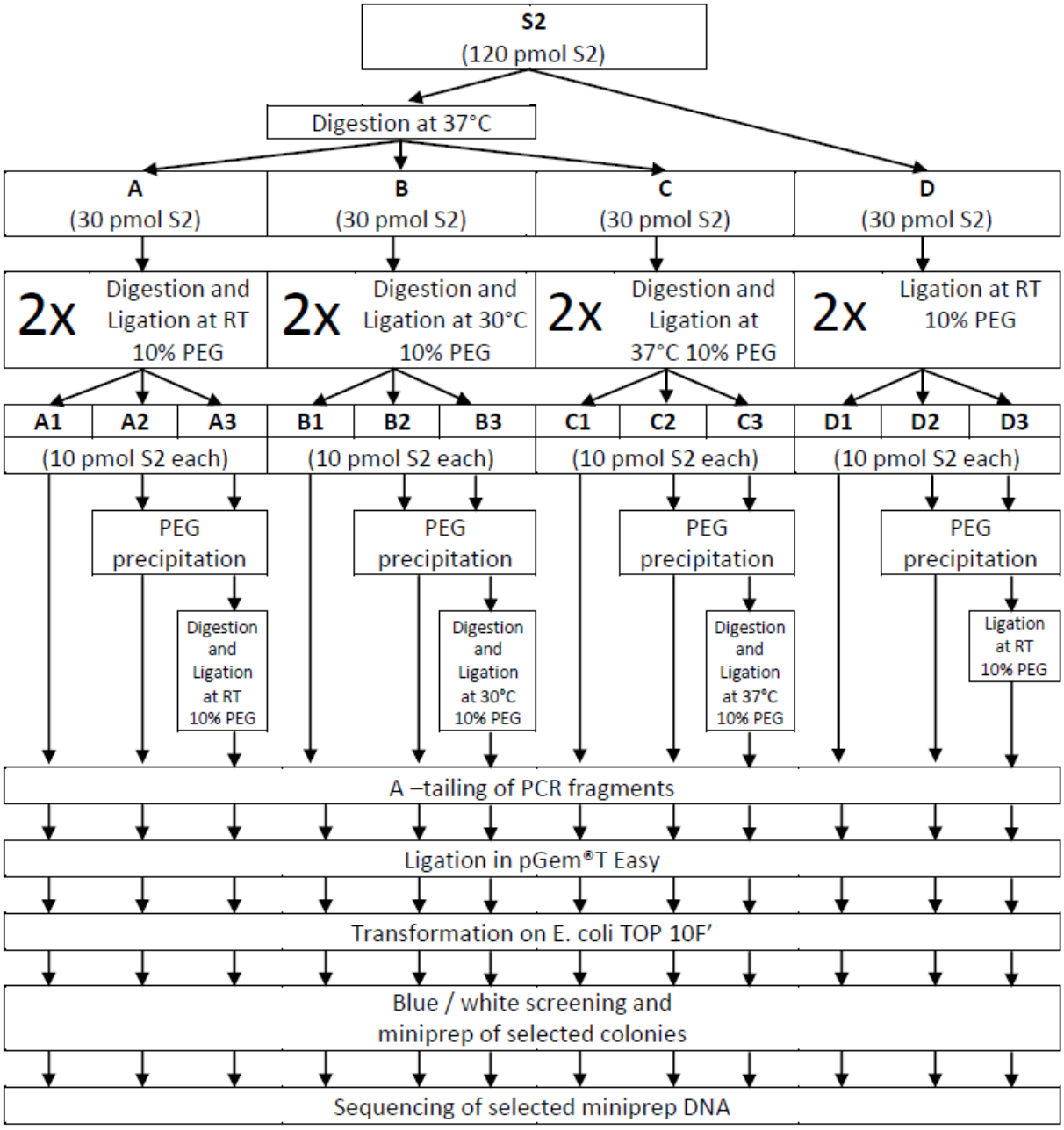
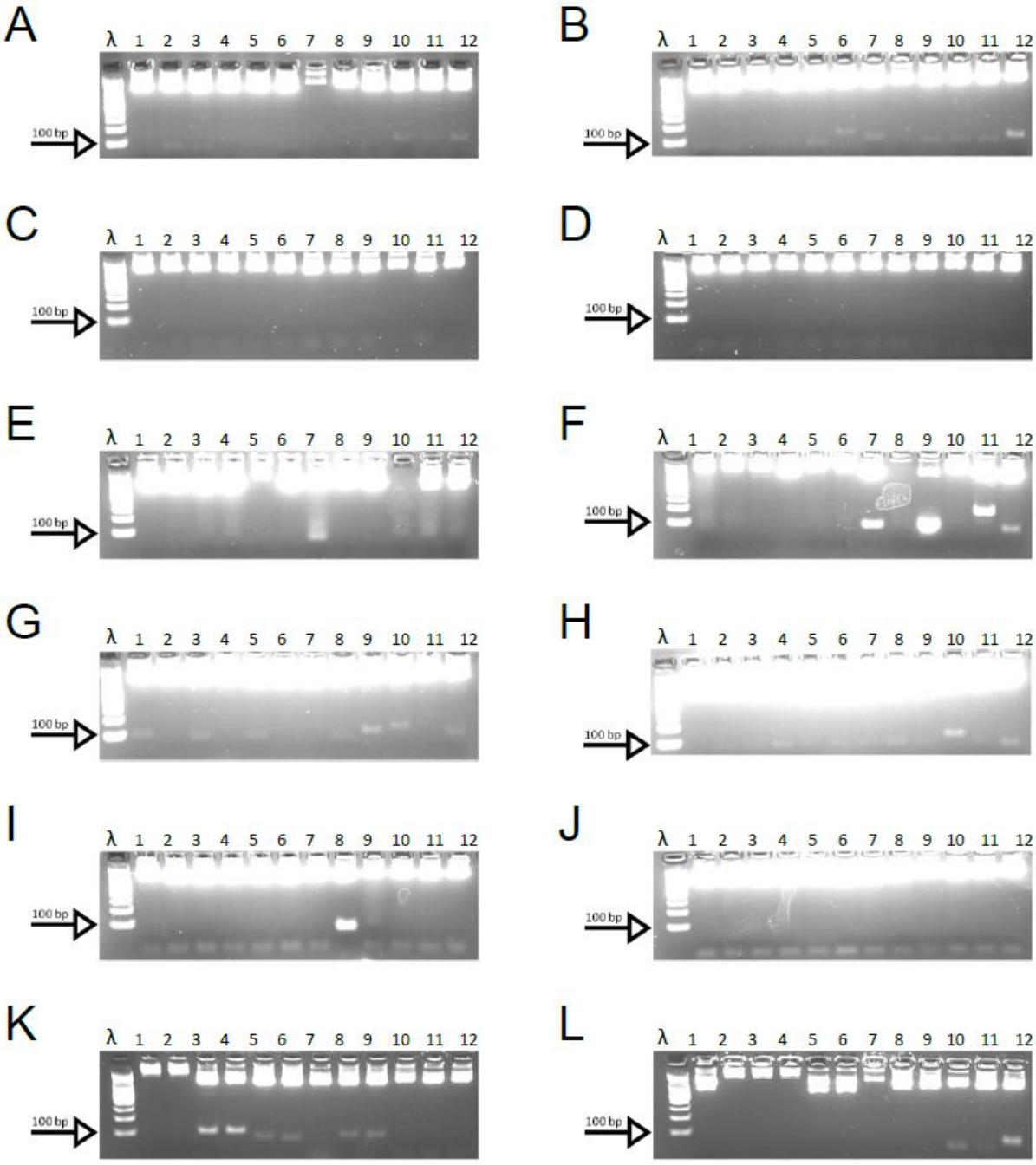


Figure 48: Experimental design to increase the concatemer length by stimulation of intermolecular and reduction of intramolecular ligation, circularization, by polyethylene glycol. And an experiment where the S2 oligonucleotides were blunt end ligated.

The experiments A, B and C are a mere repetition of the experiment in 4.2.1.2 with slight modifications. The time period for the simultaneous digestion and ligation step was set at 2.5 hours to ensure complete concatenation. In addition to the previous experiment the S2 oligonucleotide was predigested by BpiI at 37°C over night before the multimerization reaction. In each experiment the multimerization reactions were carried out at a different temperature: in A at 23, in B at 30 and in experiment C at 37°C. The concatenations were performed in the presence of 10% PEG6000, polyethylene glycol with an average molecular weight of 6000 g / mol. In experiment D S2 was blunt end ligated to circumvent the problems that may be associated with the simultaneous digestion and ligation. The ligation was carried out the same way as the digestion ligation steps at the optimal ligation temperature, room temperature. To sufficiently multimerize the binding sites the simultaneous digestion ligations reaction and the blunt end ligation was done twice. Each multimerization experiment was then divided into three subgroups (f. Ex. A1, A2, and A3). The first group of each experiment was A-tailed as it was and cloned into pGemT®Easy. In the second and third subgroup the concatemers were precipitated by PEG to remove the spurious end parts and small non-multimerized binding sites. After that, the second group was A-tailed and cloned into pGemT®Easy. The third groups of A, B and C were again simultaneous digested and ligated, because with the end parts removed, the odds of ligating two concatemers together are increased. The third group of experiment D was again blunt end ligated. The idea behind this was the same as before: joining of two concatemers to form even larger concatemers. With the single S2 oligonucleotides removed through the PEG precipitation, the odds of joining two concatemers are higher. This multimerizations were again carried out in the presence of 10% PEG6000. After the multimerization step the third group was A-tailed and also cloned into pGemT®Easy. The vectors were then transformed into E. coli TOP10F' and plated on 1.5% agar plates containing LB media, ampicillin, IPTG and X-gal. Twelve positive colonies from each group were selected for miniprep and restriction digestion analysis (Figure 49). The enzymes used for the restriction digestion analysis are shown in Table 10.



M

Figure	A	B	C	D	E	F	G	H	I	J	K	L
Group of experiment	A1	B1	A2	B2	A3	B3	C1	D1	C2	D2	C3	D3

Figure 49: Agarose gel electrophoresis of restriction digestion of plasmids form the PEG assisted ligation experiment. 10% PEG was added to the simultaneous digestion ligation reaction to improve intermolecular ligation and to reduce circularization. A blunt end ligation of the whole S2 oligonucleotide was also tested. Twelve positive colonies from each group were selected for miniprep. The obtained plasmids were then restriction digested to verify the insert length. The agarose gels belonging to the different experiments are annotated in Figure 49M. The arrows point at the bands of the DNA ladder (λ) corresponding to the base pairs annotated over the arrows. The DNA ladder is described in Material and Methods (3.2.5. Agarose gel electrophoresis).

The agarose gel electrophoresis presented the same results as all the other experiments before: a very low grade of multimerization if the oligonucleotides were multimerized at all. Some plasmids showed indeed a large insert, and a few of them were selected for sequencing (Table 14). The whole sequences can be seen in appendix 7.4.3.

Sequenced Mini	Result
A3_4 (E4)	Negative
A3_11 (E11)	Negative
A3_12 (E12)	Negative
B3_11 (F11)	5 binding sites
D3_1 (L1)	Negative
D3_5 (L5)	Negative
D3_6 (L6)	Negative

Table 14: Sequenced minipreps and their result. Anything other than multimerized binding sites is negative.

Even in the one plasmid that contained multimerized binding sites they were not multimerized as intended. All other sequences were negative in that way that they did not contain any multimerized binding sites. The analysis of these six sequences was rather surprising. All of these sequences were the same and belong to the HSP70A1A gene, a plasmid which was intensively used by another lab member at the same time the multimerization reactions were carried out. It therefore seemed that it is easier to contaminate multiple multimerization reactions with airborne DNA (the work benches were at different sides of the lab) than to multimerize the binding site to reasonable length. Therefore this approach to form concatemers was omitted and an alternative method to the simultaneous digestion and ligation was sought.

4.2.2 Second Approach: SAGE-like multimerization of binding sites

SAGE, serial analysis of gene expression, is a method for the quantitative and simultaneous analysis of a large number of transcripts, first published by Viktor E. Velculescu in 1995.[130] In SAGE a RNA library is hybridized to two linker molecules attached to magnetic beads, and converted to cDNA. The linkers contain a restriction site for a restriction enzyme that cuts the cDNA generating a tag and each one a different site for primer annealing. Two tags, with different primer sites, are ligated together, forming a ditag. The ditags are then amplified by PCR, the linkers are cut off and the ditags are concatenated. The concatemers are cloned into backbone vectors, transformed into E. coli and then sequenced. The obtained sequences are then compared with a sequence database and the sequenced tags are then assigned to genes. This gives not only an insight on which genes are expressed, but also the relative expression levels of these genes, determined by the quantity of tags obtained per gene.

The similarities between the SAGE method and the method to multimerize the binding sites can be reviewed as: library of short DNA fragments are ligated together to form a concatemer, which are then sequenced. So the method to multimerize the binding sites was adapted according to SAGE protocols and is therefore called SAGE-like. This method does not use the simultaneous restriction digestion and ligation to concatenate the fragments, but it separates all steps by a purification of the fragments by PAGE.

A new oligonucleotide, S3 (Table 8), was ordered for this new method. The overall design of S3 is very similar to S2, two BpiI restriction sites and a random nucleotide site in the middle. This time the random nucleotide stretch was modified to carry the HMG box typical TTTG

high affinity binding site[131], flanked by two and four random. This modification should ensure the binding of Lef1 in the correct orientation wit high affinity and the selection for even higher affinity would only take place at the adjacent random nucleotides. S3 was designed in a way that it could be used for the SAGE-like multimerization as well as for the simultaneous restriction and ligation method.

For the first experiment a simple step by step procedure was chosen (schematic shown in Figure 50). To simulate the outcome of an actual binding site selection experiment small amounts of S3 DNA were amplified by PCR. This step was also important to generate a high amount of starting material because during all of this different purification steps DNA gets lost.

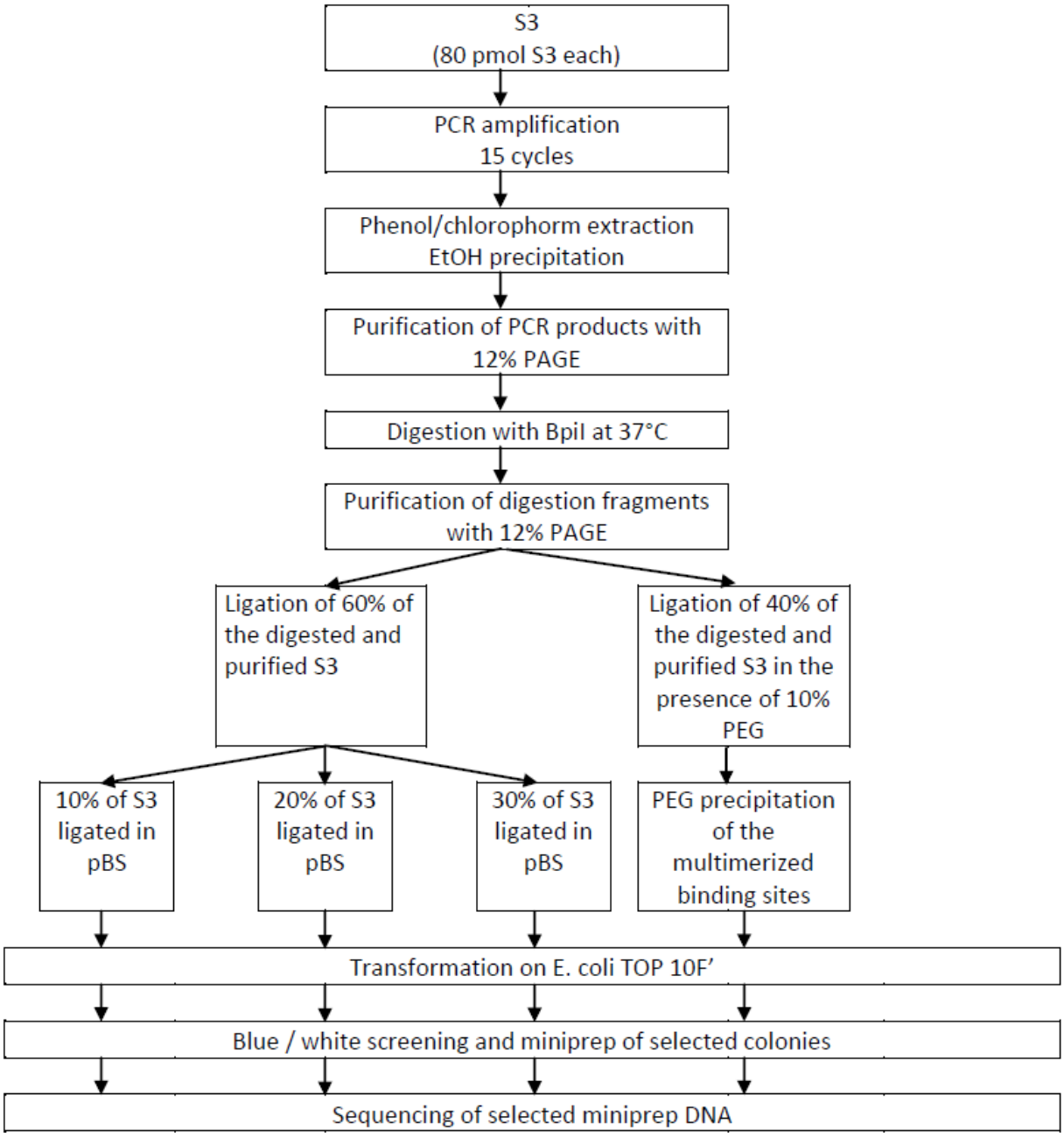


Figure 50: Experimental design of the first SAGE-like approach to multimerize binding sites.

Eight reactions, with 333 fmol S3 each, were amplified by 15 cycles of PCR. The PCR products were pooled and split into three equal volumes for phenol/chloroform extraction and ethanol precipitation. The purified PCR product was then separated by PAGE on a 12% polyacrylamide gel (Figure 51).

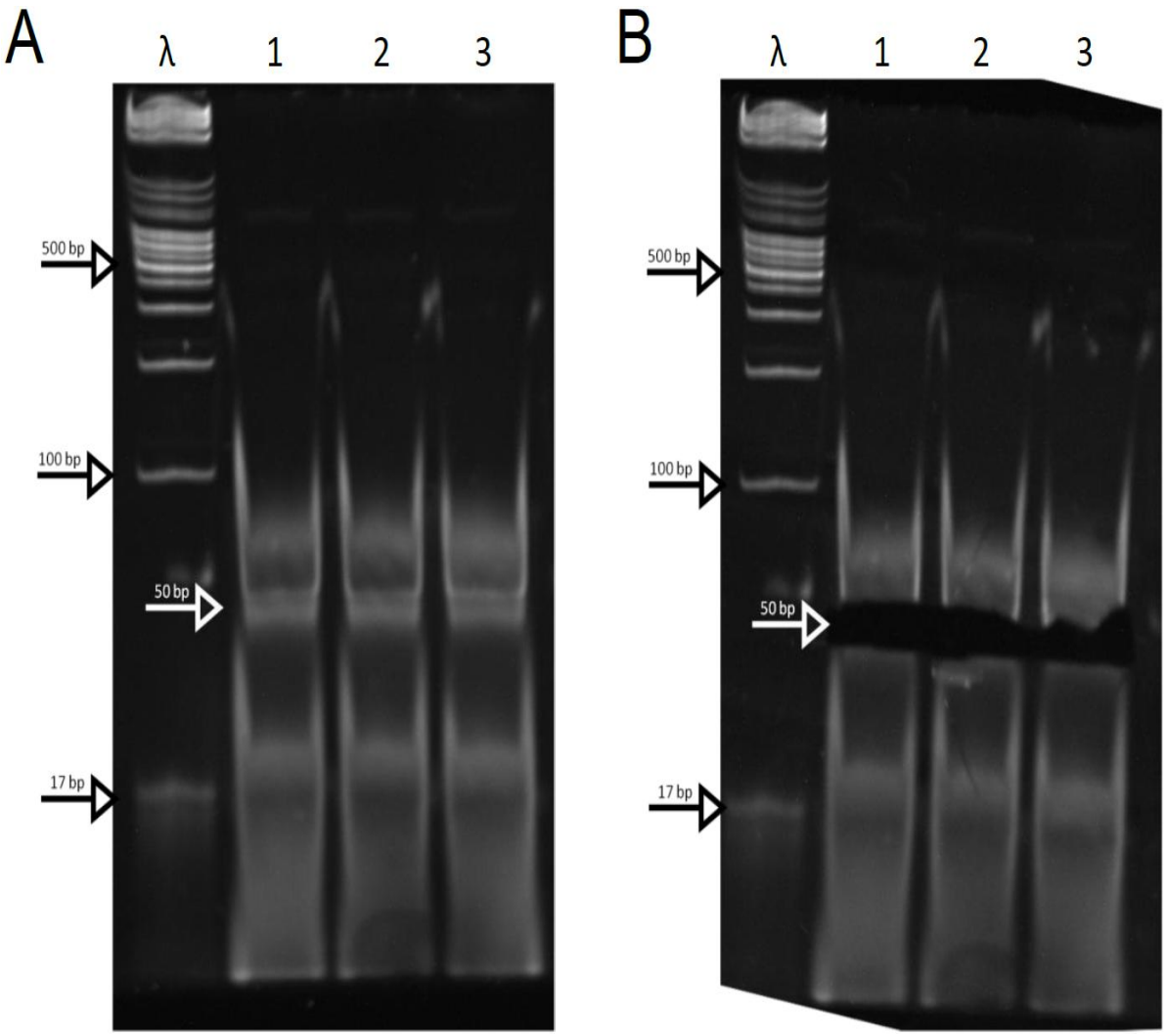


Figure 51: 12% non-denaturing PAGE of the phenol/chloroform extracted and ethanol precipitated PCR product of S3. A: The purified PCR product of S3 was loaded in slot 1, 2 and 3 of a 12% non-denaturing PAGE and run at 20V / cm until the bromophenole blue band reached the end of the gel. S3, the 50 bp band, was excised and S3 was eluted in LoTE. B: The same gel after excision of the 50 bp band. The DNA was stained by incubating the gel in 0.5 µg / mL ethidium bromide in 0.25x TBE for 15 min and destaining in ultrapure water for another 15 min. The DNA was visualized by an Alphaimager® Mini (Cell Biosciences). The DNA ladder (λ) is the same as the one used for the agarose gel electrophoresis and is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

The 50 bp bands, comprising S3, were excised and S3 was eluted in LoTE and subsequent ethanol precipitated. S3 from each band was treated separately. S3 was then digested by BpiI and the digestion fragments were purified by PAGE (Figure 52).

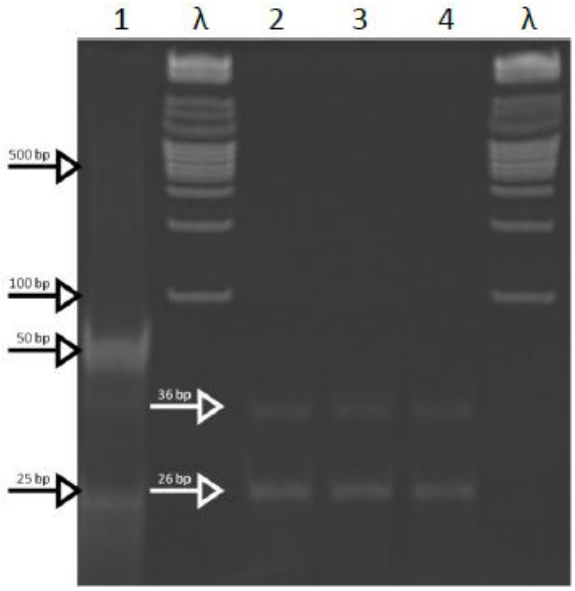


Figure 52: Non-denaturing PAGE of the digestion fragments of S3 digested with BpiI. The digested S3 was loaded on a 12% non-denaturing PAGE, in lanes 2-4, and run at 20V / cm. The 26 bp fragment was excised and eluted in LoTE. The 36 bp band indicates semi-digested S3. The DNA was stained by incubating the gel in 0.5 µg / mL ethidium bromide in 0.25x TBE for 15 min and destaining in ultrapure water for another 15 min. The DNA was visualized by an Alphaimager® Mini (Cell Biosciences) The DNA ladder (λ) is the same as the one used for the agarose gel electrophoresis and is described in Material and Methods (3.2.5. Agarose gel electrophoresis). In addition to the DNA ladder 25 ng of two oligonucleotides with the length of 25 and 50 bp were loaded in lane 1.

In the lanes 2-4 in Figure 52 two bands can be seen. The 36 bp bands indicate only semi-digested S3, with one end still attached and the 26 bp fragment is the fully digested S3. The 26 bp fragment, consisting of the binding site, was excised and eluted in LoTE, and again precipitated by ethanol. The S3 binding sites were then multimerized by ligation with T4 DNA ligase at room temperature over night. Forty percent of the digested S3 was separated and 10% PEG was added to improve intermolecular ligation and to reduce intramolecular ligations[123,122,127]. The multimerized binding sides from the ligation with 10% PEG was precipitated in 30% PEG and then ligated into pBSII KS(+). The remaining 60% S3 ligation reaction was ligated in different amounts (10, 20 and 30% of total S3) directly into pBSII KS(+). The generated plasmids were then transformed in E. coli TOP10F'. Bacteria were grown on LB agar plates containing ampicillin, IPTG and X-Gal. Positive clones were selected and subjected to miniprep and restriction digestion analysis (Figure 53).

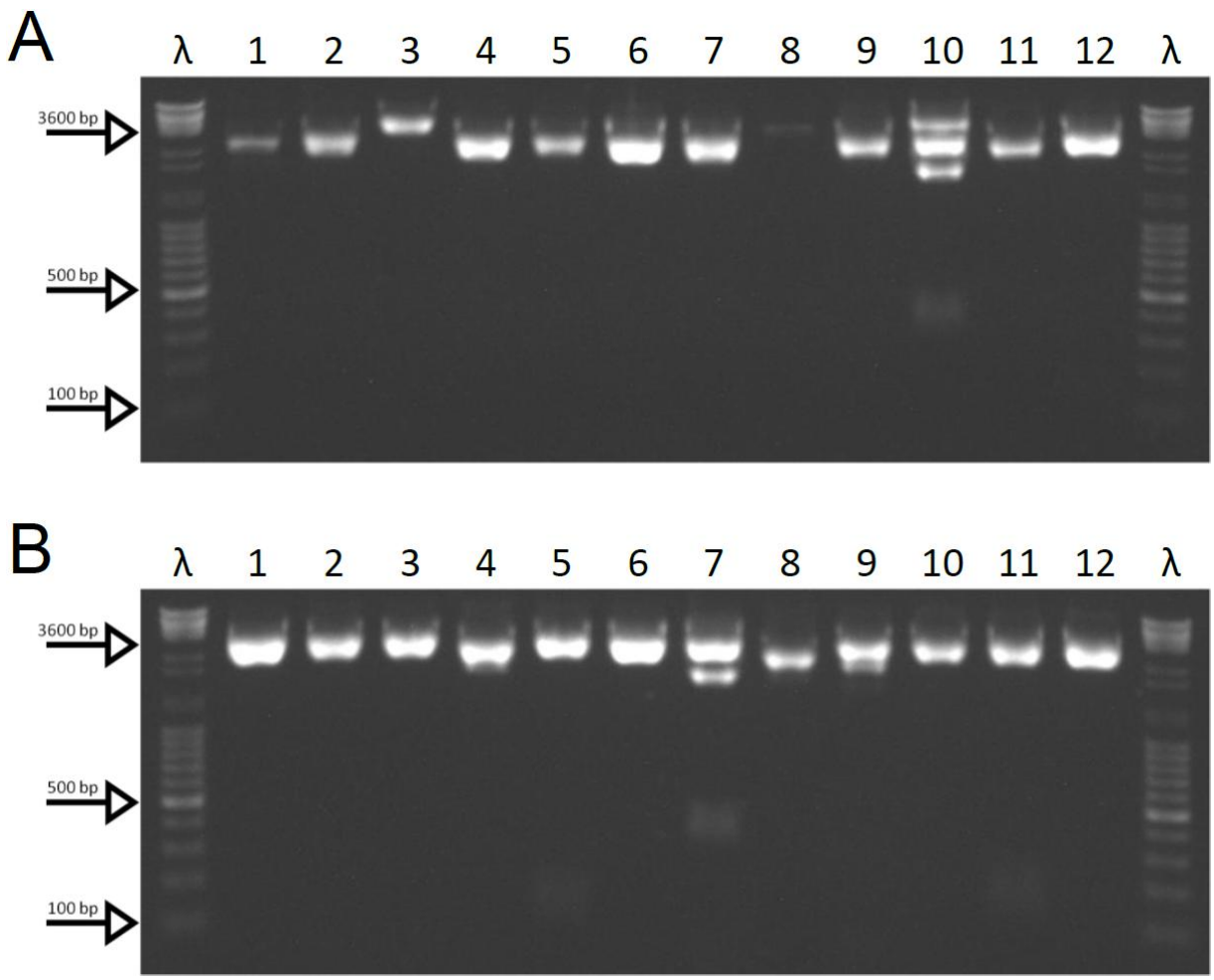


Figure 53: Agarose gel electrophoresis of the restriction digested minipreps the positive clones containing pBSII KS(+) with the multimerized binding sites as insert. In slot A1-6 10%, in A7-12 20% and in B1-6 30% of the total ligated S3 DNA was used for ligation in pBSII KS(+). In Slot B7-12 the remaining 40% S3 DNA were ligated in presence of 10% PEG and subsequently cloned in pBSII KS(+). The DNA ladder (λ) is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

Isolated plasmids from all four minipreps that showed bands indicating inserted DNA (Figure 53A11, 36B5, 36B7 and 36B11) were sequenced. The miniprep of Figure 53B7 could not be sequenced and sequencing of Figure 53A11 and 36B11 gave useless sequences. Only the miniprep of Figure 53B5 showed useable results. A small number of three multimerized binding sites could be found in the insert. This result showed that in principle this method is suited for the multimerization of the binding sites, but it needed to be improved to yield enough usable sequences. To achieve this, the number of positive clones, carrying plasmids with preferably high numbers of multimerized binding sites, has to be increased. There are several reasons why there are so few positive clones, the most obvious two are that there is to little DNA for efficient ligation, or the ligation works just fine, but the resulting product is not incorporable in an plasmid because it gets circularized. This circularization should be in part prevented by the addition of PEG, but this didn't seem to work reliably. In a different SAGE-method called "Robust-LongSAGE"[132] the problem of the circularized concatemers is solved by a partial digestion. This method also uses different concentrations of the polyacrylamide gels, better suited to separate the different small fragments. It also seperates the concatemers by PAGE and subdivides them into three size dependend groups.

4.2.2.1 Optimized SAGE-like experiment

Because S3, once ligated, has no internal restriction sites, it can not be used for partial digestion and a fourth oligonucleotide, S4; had to be designed. S4 carries two XhoI restriction sites with minimal distance to the ends to maximise the difference of the fragment length between the end parts and the selection part. It also contains a prolonged random nucleotide strech, again harboring the HMG high affinity consensus sequence flanked by 4 and 6 random nucleotides. The method was optimized and a new experiment was carried out (Figure 54)

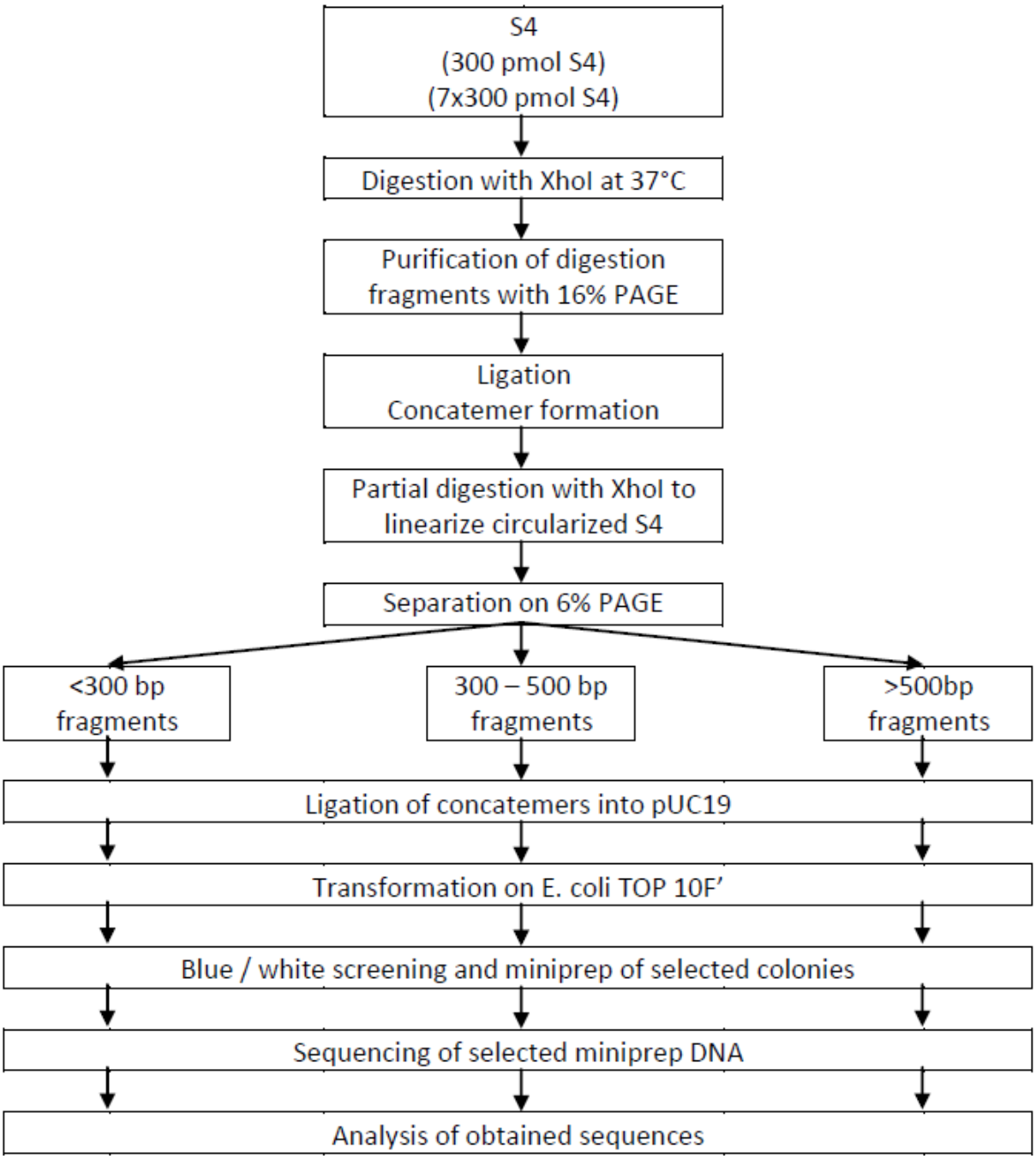


Figure 54: Experimental design of the improved SAGE-like multimerization method with partial digestion of the concatemers to reduce circularization. S4 was digested by XhoI purified and ligated. The formed concatemers were partially digested by XhoI to linearize circulated binding sites. After the partial digestion the concatemers are separated by PAGE and subdivided in three fragment length dependent groups, then cloned into the pUC19 backbone. The plasmid was then cloned into E. coli TOP10F'. Minipreps of the positive clones were screened for inserts and insert carrying plasmids were sequenced. The obtained sequences were analysed.

The second SAGE-like experiment started with the digestion of the S4 oligonucleotide, because the PCR amplification and purification step worked in the first experiment, as could be seen in Figure 51 and Figure 52. First 300 pmol of S4 were digested by XhoI and the digestion fragments were separated by a 16% polyacrylamide gel (Figure 55A).

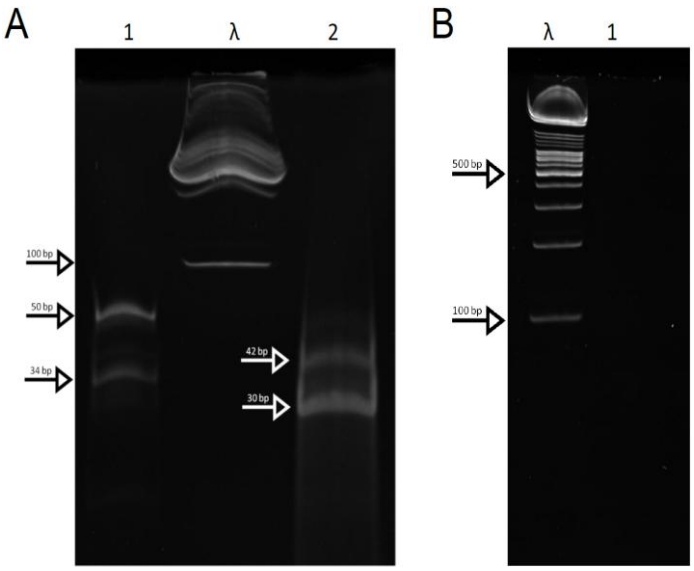


Figure 55: PAGEs of the second experiment of the SAGE-like approach. A: 16% PAGE of the XhoI digestion products. Lane 1: 25ng of a 50 bp and a 34 bp oligonucleotide. Lane 2: digestion fragments of S4. The 30bp band was excised and eluted in LoTE. The 42 bp band is semi-digested S4. B: 6% PAGE of the partially digested concatemers. Both PAGE's were run at 20 V/cm until the bromophenole blue band reached the end of the gel. The DNA was stained by soaking the gel in a 3x staining solution of GelRed™ (Biotium) in 0.25x TBE for 30 min, and no destaining is necessary. The DNA was visualized by an Alphaimager® Mini (Cell Biosciences). The DNA ladder (λ) is the same as the one used for the agarose gel electrophoresis and is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

As seen before (in Figure 52), there are semi and fully digested oligonucleotide bands visible. The fully digested oligonucleotide (30 bp band) was excised, eluted in LoTE and ethanol precipitated. The digested S4 were concatenated and subsequently partial digested by XhoI for one minute at 37°C. The partial digested concatemers were separated on a 6% PAGE (Figure 55B). No band or even a smear could be seen on the gel, so the experiment was repeated with an increased amount of S4 DNA at the beginning.

The sevenfold, in sum 2.1 nmol, amount of S4 DNA was digested by XhoI and separated on a 16% PAGE (Figure 56).

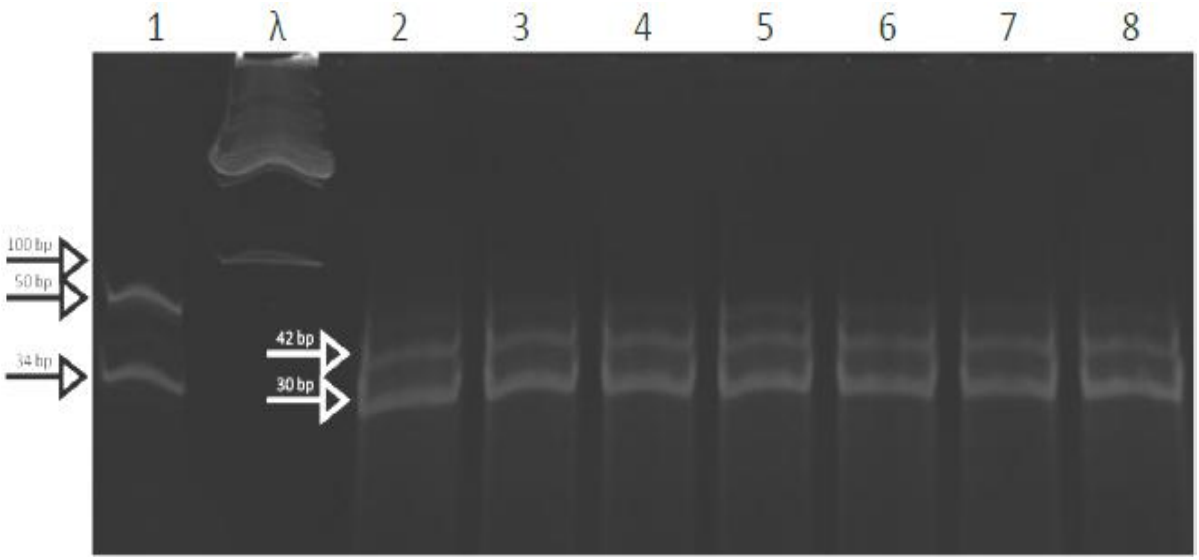


Figure 56: 16% PAGE of the XhoI digestion products of 7x300 pmol S4. Lane 1: 25ng of a 50 bp and 50 ng of a 34 bp oligonucleotide. Lane 2-8: XhoI digestion products of 7x300 pmol of S4. The 42 bp band indicates only semi-digested S4 with still on end attached. The 30bp band, the fully digested S4, was excised and eluted in LoTE. The gel was run at 20 V/cm until the bromophenole blue band reached the end of the gel. The DNA was stained by soaking the gel in a 3x staining solution of GelRed™ (Biotium) in 0.25x TBE for 30 min, with no destaining. The DNA was visualized by an Alphaimager® Mini (Cell Biosciences). The DNA ladder (λ) is the same as the one used for the agarose gel electrophoresis and is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

The 30 bp band, indicating fully digested S4 oligonucleotide was excised, eluted in LoTE, and ethanol precipitated. The digested S4 were then again ligated to form concatemers. To linearize the circulated concatemers they were partially digested by XhoI for 1 minute at 37°C and separated on a 6% polyacrylamide gel (Figure 57).

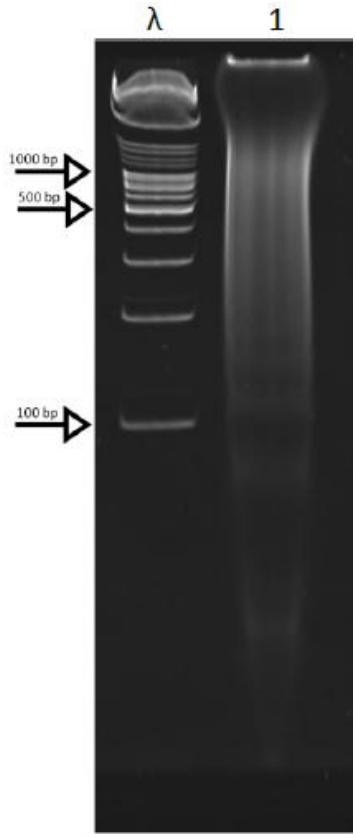


Figure 57: 6% PAGE of the partially digested concatemers. Lane1: The fully digested S4 binding sites were ligated and then partially digested by XhoI for one minute at 37°C. The concatemers were excised in three groups: smaller than 300 bps, 300-500 bps and all concatemers that are bigger than 500 bps. The concatemers were eluted in LoTE over night and subsequently ethanol precipitated. The gel was run at 20 V/cm until the bromophenole blue band reached the end of the gel. The DNA was stained by soaking the gel in a 3x staining solution of GelRed™ (Biotium) in 0.25x TBE for 30 min, with no destaining. The DNA was visualized by an Alphaimager® Mini (Cell Biosciences). The DNA ladder (λ) is the same as the one used for the agarose gel electrophoresis and is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

The PAGE showed a smear running over the whole length of the gel. Even slightly brighter regions in the smear are visible, especially around 100 bp, indicating concatemers with the lengths of multiples of 30 bps. There are concatemers visible with the length of 60, 90, 120, 150 and 180 bps. The remaining concatemers are not separated enough to actually see them individually. The concatemers were excised and subdivided into three groups. The first group was holding all the small concatemers up to a length of 300 bps. The second one held the 300-500 bp section and the third group held all the concatemers that were bigger than 500 bps. The groups of concatemers were then separately eluted in LoTE over night and subsequently ethanol precipitated. The three groups of concatemers were than ligated into a SalI digested pUC19 vector and transformed into E. coli TOP 10F'. The transformed E. coli were grown on 1.5% agar plates containing ampicillin, X-gal and IPTG. Eight positive colonies were taken for miniprep and restriction digestion analysis (Figure 58)

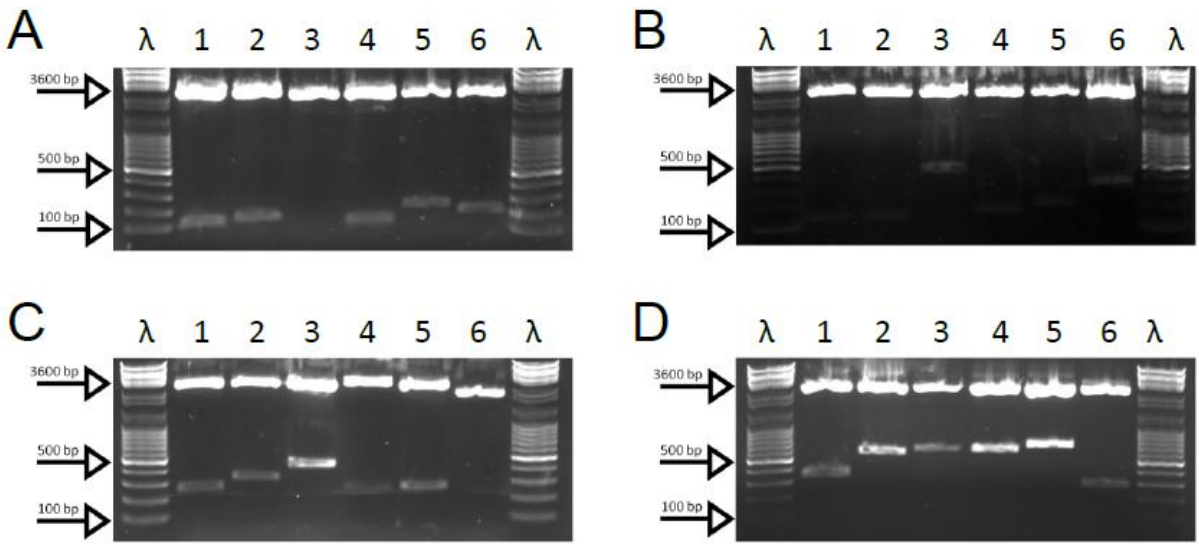


Figure 58: Agarose gel electrophoresis of the restriction digested minipreps the positive clones containing pUC19 with the multimerized binding sites as insert. The multimerized and partially digested binding sites were subdivided into three length dependent groups. These groups were cloned into SalI digested pUC19 and transformed into E. coli. Positive colonies were taken for miniprep and restriction digestion analysis. In lanes A1-6 and B1-2 the concatemers smaller than 300 bps were loaded. In lanes B3-6 and C1-4 the concatemers ranging from 300 to 500 bps and in lanes C5-6 and D1-6 the concatemers larger than 500 bps were loaded. The DNA ladder (λ) is described in Material and Methods (3.2.5. Agarose gel electrophoresis)

The restriction analysis in Figure 58 showed that only two of the 24 selected colonies (Figure 58A3 and Figure 58C6) didn't carry an insert. The isolated plasmids of seven of the positive colonies were sequenced using the sequencing primer (Table 10). Plasmids from all three length dependent groups were sequenced. The number of sequences found in each sequenced plasmid and its associated band in Figure 58 is depicted in Table 15.

Sequenced plasmids from minipreps	Number of binding sites
<300bp Mini 5 (A5)	7
<300bp Mini 6 (A6)	6
300-500bp Mini 1 (B3)	13
300-500bp Mini 7 (C3)	12
>500bp Mini 4 (D2)	16
>500bp Mini 6 (D4)	17
>500bp Mini 7(D5)	19
Total sequences	90
Total useful sequences	**83**

Table 15: Numbers of sequences obtained from plasmids. Large insert carrying plasmids (seen in Figure 58) were selected for sequencing. The sequences were analysed for multimerized binding sites. Only binding sites with correctly sequenced nucleotides in the random, binding site selection area were counted as useful sequences.

4.2.2.2 Analysis of the obtained sequences

All sequenced plasmids contained multimerized versions of the unselected binding site oligos. The number of binding sites from Table 15 and the length of the digestion fragments in Figure 58 corresponded in all cases. The sequences were then edited in the meaning that all binding sites had the correct orientation. The orientation was easily determined by the Lef/Tcf consensus binding site of TTTG. All sequences with the wrong orientation, with the complementary Lef/Tcf consensus binding site CAAA, were corrected. The corrected sequences were then aligned using MEGA5[119]. The obtained .FASTA file was used to generate position specific nucleotide tables (Table 16).

A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	9	24	22	20	0	0	0	0	11	14	17	16	18	14
C	22	21	14	17	0	0	0	0	21	28	23	16	18	27
G	29	27	25	25	0	0	0	83	27	26	25	24	24	27
T	22	11	21	21	83	83	83	0	24	15	18	27	21	15

B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	11	29	27	24	0	0	0	0	13	17	20	19	22	17
C	27	25	17	20	0	0	0	0	25	34	28	19	22	33
G	35	33	30	30	0	0	0	100	33	31	30	29	31	33
T	27	13	25	25	100	100	100	0	29	18	22	33	25	18

Table 16: Position specific nucleotide matrix of the unselected binding sites. The nucleotide sequences of the sequenced plasmids were analysed and the multimerized, unselected binding sites were divided into single sites. The sequences were then checked for their integrity and orientation. Sequences with too many nucleotides, missing some nucleotides, contained random mutations in the Lef/Tcf consensus sequence or could not entirely be sequence were omitted. If the binding site had the wrong orientation its consensus sequence was taken. A: Matrix that gives the numbers of nucleotides occurring at a specific position. B: Matrix that gives the probability in percent of the nucleotides occurring at that specific position. Only the 14 random bases long, binding site selection part of the oligonucleotide is shown. Position 5-8 is the Tcf/Lef consensus binding site consisting of TTTG.

Table 16 shows the occurrence of each nucleotide at a specific site in the 14 random bases long binding site selection part in Table 16A: in numbers and B: in percent.

To get a more visual representation of these results, the sequence alignment file was used to generate a logo by submitting it to WebLogo3 (http://weblogo.threeplusone.com). At first a sequence logo of the whole multimerized oligonucleotide was generated to visualize the stability of the given sequence parts and the pure random fashion of the binding site selection part, because no selection was carried out (Figure 59).

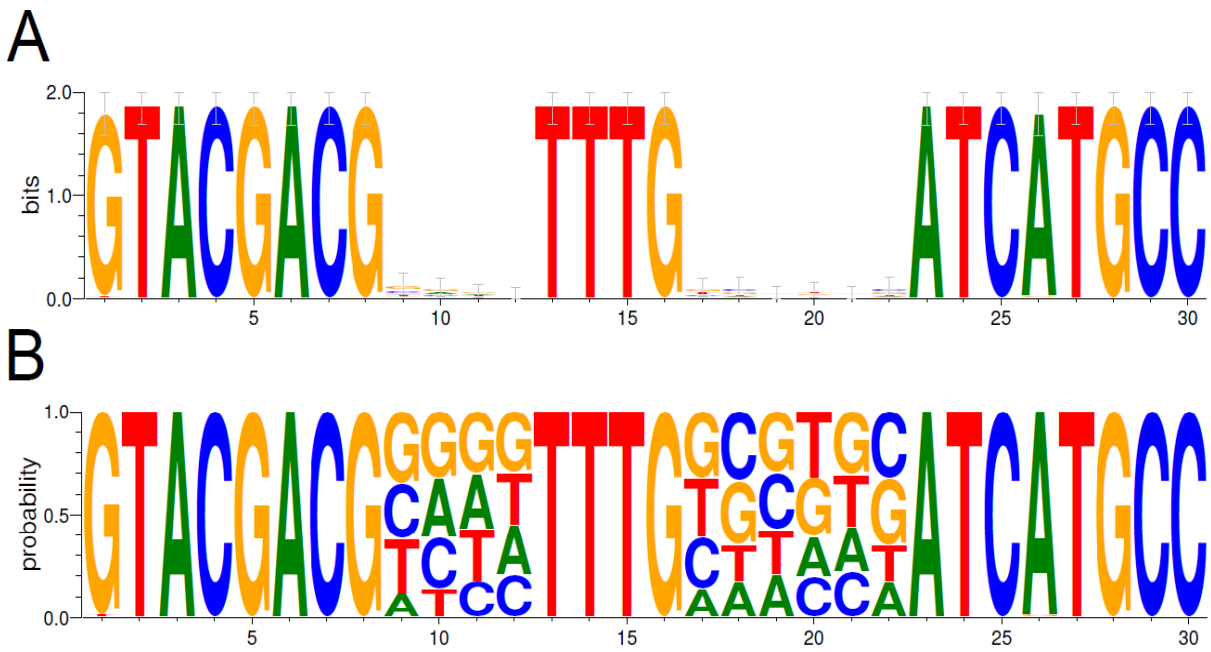


Figure 59: Sequence logos of the whole unselected, multimerized binding site selection oligo. The aligned sequences of the unselected, multimerized binding site oligos were submitted to WebLogo3 and two sequence logos were generated. A: Sequence logo that shows the information content in bits, a measure of entropy in information theory. B: A Sequence logo showing the probability of every nucleotide at each position. Both sequence logos show a high degree of randomness in the binding site selection part, and a high stability in all other parts.

The sequence logo in Figure 59A shows the content of information contained in a message, in information theory, this is called entropy[133]. It shows the predictability of the nucleotides at each position, with a maximum of two bits. Two bits mean that all nucleotides in this message can be predicted, and in this case the nucleotides in one position of the oligonucleotide would all be the same throughout the alignment. In Figure 59B the probability of one nucleotide at one position is shown.

Finally, sequence logos were generated visualizing only the random, binding site selection part, like it would be done, if a binding site selection was carried out and the selected binding site was depicted (Figure 60)

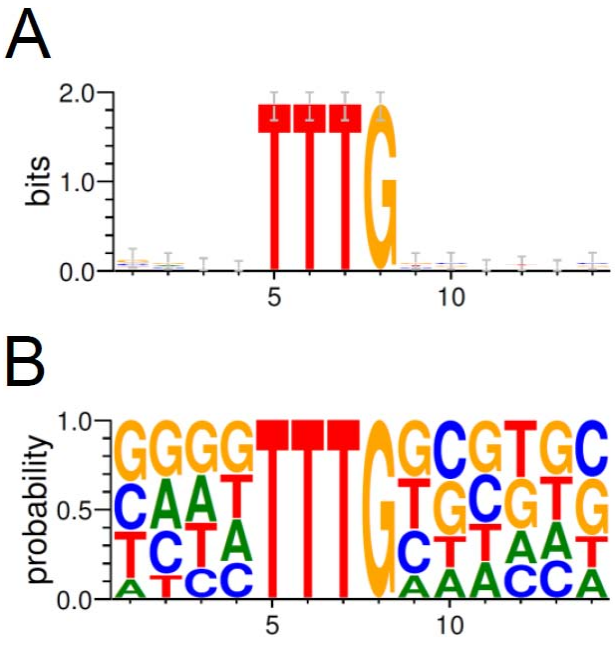


Figure 60: Sequence logos of the binding site. A: Sequence logo that shows the information content in bits, a measure of entropy in information theory. B: A Sequence logo showing the probability of every nucleotide at each position.

5 Discussion

The Wnt signalling pathway has an important role in development and disease [3,9–11]. In this signalling pathway, signals from the cell membrane are transduced to the nucleus where target gene transcription is activated. For this activation the interaction of the amino terminal β-catenin binding domain of the architectural transcription factor Lef1 with the armadillo repeats of β-catenin is necessary. Additional transcription activating factors are then recruited by the armadillo repeats to enable the CTD of β-catenin and target gene transcription.

In 2005 Daniels and Weis[54] discovered an additional β-catenin binding site on Lef1, which involves the HMG DNA-binding domain of Lef1. Based on their investigation the hypothesis of the present study was that the interaction of β-catenin with this newly discovered Lef1 binding site alters the DNA sequence to which Lef1 alternatively binds. To prove this hypothesis a binding site selection experiment was applied. The following binding site selection experiment was designed on the basis of the publications of Wright et al.[110] and Atcha et al[6]. To enable the planned binding site experiments different preparatory steps had to be performed and tested.

The aim of the present thesis was to express functional proteins for the binding site selection and to establish a method for the multimerization of the selected binding sites to increase the yield of sequenced binding sites.

5.1 Protein expression

The first preparative step for successful binding site selection is the expression and thereby the controlled production of the selected proteins.

Four proteins were chosen, two different truncated Lef1 proteins, containing only the HMG domain (LefHMG and Lef293), and two Lef-β-catenin fusion proteins based on these two truncated Lef1 proteins. The β-catenin fused to Lef was N–terminally truncated and consisted only of the armadillo repeats 9-12 and the C-terminal domain. To be able to express these proteins in E. coli, plasmid vectors for the expression of the Lef and Lef fusion proteins were constructed. These expression vectors contained a 6xhistidine tag for subsequent protein purification.

The proteins were expressed for a time course experiment in E. coli BL21(DE3). This experiment showed that none of the proteins were expressed or they were expressed at very low concentrations. Following this, an analysis of the codons used by E. coli and the needed proteins revealed major differences in codon demand and offered tRNAs. Therefore, a different E. coli strain, E. coli BL21(DE3)RIL, capable of satisfying this tRNA requirement, was chosen for protein expression and a second time course experiment was performed. Additionally to the new E. coli strain two different inducer concentrations were introduced. The results of this experiment showed strong expression of all proteins even at the lower inducer concentration.

Based on these results it was decided to express the proteins for purification in E. coli BL21(DE3)RIL in larger quantity. The HMG domain of Lef1 is unstructured if not bound to DNA[80]. Therefore, the purification under denaturing conditions with subsequent refolding should not be a problem, at least for the LefHMG and Lef293 proteins. The armadillo repeat domain of β-catenin is highly structured and it is possible that they cannot be refolded properly during dialysis[83]. Although the expressed Lef-β-catenin fusion proteins contain only a small part of these armadillo repeats, refolding was assumed to be possible. Based on these considerations, the proteins were purified under denaturing conditions. The functionality of the purified proteins was tested by an electrophoretic mobility shift assay (EMSA). The EMSA clearly showed that the proteins containing Lef293 were able to shift DNA. Contrary, purified LefHMG and its β-catenin fusion protein were not able to bind to and shift DNA.

With the EMSA confirming the expression and purification of functional proteins, at least for Lef293 and Lef293βcat, the purification procedure could be optimized.

Daniel et al.[134] and Newhouse et al.[112] used proteins bound to magnetic beads in their binding site selection experiments. Based on their experiments, native purification of the proteins by the aid of magnetic beads was considered. This has the advantages of being a quick and easy procedure and there is no need for refolding by dialysis or even elution.

This has the additional advantage that, once the amount of protein bound to the magnetic beads is determined, the same amount of protein used and recovered in all rounds of selection during the repetitive binding site selection procedure. This could not be ensured if the protein has to be eluted, dialyzed and then again get bound to the magnetic beads subsequently to each selection step.

For the purification of proteins under native conditions the solubility of the protein is very important and therefore different tests concerning the protein solubility were performed[135]. The solubility of the proteins was tested under different conditions (at two different time points during expression, at two different inducer concentrations as well as in different lysis buffers). These tests revealed that the composition of the lysis buffer had the highest influence on protein solubility. All of the tested lysis buffers contained detergents which may interfere with subsequent procedures. Further analysis showed that the detergents contained in these buffers did not significantly increase the solubility and were therefore excluded.

To further increase the solubility of the expressed proteins the rate of expression was reduced by lowering the temperature during expression. Although the presented experiments did not show an improvement in protein solubility, the lowered expression temperature was then used for further experiments. This decision was based on recommendations in literature[135].

Another important step for successful protein expression is the lysis of the harvested cells. If the lysis is not complete, proteins in the cytoplasm of still intact cells are lost for further processing. In the present study the chosen technique for cell lysis was sonication. To test the success rate of this technique, lysozyme was added to improve the results of lysis. As the addition of lysozyme showed no improvement on cell lysis, it was deducted that the lysis with sonication is enough to disrupt all cell walls and to release the expressed proteins in the crude extracts.

The parameters for preparative expression of the proteins, cell lysis and its buffer for high protein solubility were deducted from the experiments mentioned above. With these parameters, the proteins were expressed and purified under native conditions with the aid of magnetic beads. The concentrations of the expressed and purified proteins were determined by a Bradford assay. Protein concentrations were in a range well usable for the binding site selection in comparison with the amounts used for binding site selection experiments found in the literature[110,113,134].

To further improve the expression of proteins it could be considered to optimize the codon usage of the protein to fit the codon usage of the expression host. In addition the expression cassette could be integrated into the host genome to ensure the stable expression of the protein. Bacterial growth and expression could be improved by the optimization of the culture media or by improving the culturing conditions.

The solubility of the expressed proteins can be increased by further lowering the expression rate additionally to the reduced expression temperature. A decrease in the expression rate can be achieved by using weaker promoters, low copy number plasmids, or via the reduction of the inducer concentration. The solubility of the expressed proteins, their correct folding and disulfide formation can also be improved by secretion of the proteins into the periplasmic space. Other benefits of periplasmic production are the decrease proteolysis and a simplified protein release by osmotic shock, which reduces drastically the amount of contaminating proteins in the starting material for purification[136,137].

Insoluble proteins can be expressed as fusion proteins, fused to highly soluble proteins making them soluble. The fusion protein is subsequently removed by protease cleavage[138].

Furthermore, the solubility and stability of proteins can be improved by the coexpression of chaperons and/or by increasing the osmotic stress of the expression hosts. It is also possible

to treat the expression culture with a heat shock to stimulate the heat shock response which leads to the expression of cellular heat shock proteins, mostly chaperones[139,140].

Native DNA binding proteins attach to DNA upon cell disruption and are therefore not available for purification. A possible solution for this problem is the treatment of the crude cell extract with DNAse during sonication or the addition of high salt concentration in the lysis buffer[141,142]. This high salt concentration leads to the dissociation of the proteins from the DNA and has to be maintained throughout the purification process.

Additional to the expression in E. coli the proteins were expressed in mammalian cell culture to ensure the correct folding of the proteins. Furthermore, it is not clear, if posttranslational modifications of the CTD of β-catenin contribute to this additional interaction with the HMG domain of Lef1. Proteins expressed in mammalian cells should be modified accordingly. If these modifications are necessary, differences in the binding site sequence between the proteins expressed in E. coli and cell culture would be detected.

New vectors, based on the E. coli expression vectors, were constructed for the expression in mammalian cells. In the first experiment HeLa cells, human cervix adenocarcinoma cells, were used as expression host. The proteins were obtained from nuclear extracts. An EMSA was performed to check if the desired proteins were expressed and functional, but none of the HeLa nuclear extracts contained proteins able to bind to and shift DNA.

These findings could indicate low transfection efficiency, low expression rate or that the proteins are misfolded and degradation by the proteasom. The transfection efficiency was checked by a transfectional control and was quite low, and thereby the introduction of another cell line was decided.

COP-8, a polyoma virus transformed mouse fibroblast cell line, which is capable of amplifying plasmids containing the polyoma origin of replication (ori) [143], was chosen. For this purpose a new expression vector containing the polyoma ori was constructed. The proteins were expressed in COP-8, extracted from the nucleus and checked for functionality by EMSA. This time the EMSA showed correct expression and functionality of Lef293 and its fusion protein and confirmed that the LefHMG proteins are not functional and can therefore not be used in the binding site selection. Because the proteins were not purified from the nuclear extracts it is difficult to determine the amount to use for binding site selection. One possible way is to estimate the amount of protein from the amount of shifted probe by EMSA.

5.2 Multimerization

At the beginning of binding site selection experiments, a pool of random oligonucleotides was offered to the previously expressed proteins. Oligonucleotides containing sequences with higher affinity to the protein get enriched during multiple rounds of selection. The outcome of such binding site selection experiments is a pool of enriched high affinity oligonucleotides. In binding site selection experiments found in literature, these oligonucleotides are then individually cloned into plasmids for sequencing[109,111,113,114,144–146].

In the present study a method for the multimerization of the selected oligonucleotides containing the desired binding sites was established. At first, an approach where the oligonucleotides are simultaneously digested and the binding sites ligated to concatemers was chosen. To achieve this, oligonucleotides which contain a random nucleotide patch for the actual binding site selection which can be easily multimerized were designed. To find optimal conditions for this simultaneous digestion and ligation reaction different variations of the duration and temperature were tested.

Because of the lack of positive results of these first experiments a further step was introduced: an amplification of the concatemers. After multimerization of the binding sites the concatemers were amplified by PCR. The idea was to thereby increase the probability of the concatemers being ligated into the plasmid. Although the PCR results looked promising, sequencing revealed, that only one half of the selection oligonucleotide was multimerized. A reason for this might be, that the selection oligo was designed to use only one primer for the

initial double strand formation and that the sequences flanking the random nucleotide patch were palindromic. Therefore, the single stranded oligonucleotides form a hairpin structure as likely as a primer would anneal. This fact made this oligonucleotide not suitable for PCR amplification, but gives an explanation for these results.

A second oligonucleotide was designed, carrying two different primer sites, and the multimerization experiment with PCR amplification was repeated with little success. Four out of five obtained sequences contained some desired, at small numbers multimerized oligonucleotides. This concatemers have been of poor quality with frequently attached primer sequences. There were also some nucleotides missing, indicating that the multimerization is rather due to the PCR amplification and not the simultaneous digestion and ligation procedure.

One of five sequenced plasmids showed, again, multimerization of just one half of the oligonucleotide attached to a single, entire oligonucleotide. This sequence was like the sequences obtained from the first experiment that used PCR amplified concatemers. But this time the observed phenomenon could not be explained by the design of the oligonucleotide. Because of this uncertainty concerning concatemer formation, the PCR amplification step of the concatemers was omitted and different strategies had to be tested.

Although the previous experiments showed, that PCR could not be used to increase the yield of multimerized binding sites, it gave an insight in the number of multimerized binding sites. This raised the question, if there is a natural limit for the number of binding sites that can be multimerized. The ligation of two binding sites produces an excess of two end parts which compete for further multimerization. Maybe this excess of end parts limits the multimerization beyond five to six binding sites, as it was obtained in the previous experiment.

To test if there is a limit of multimerization, based on the excess end parts, the simultaneous digestion and ligation procedure was divided into smaller steps. At first short concatemers were produced, the end parts were removed and the concatemers were then multimerized in the absence of end parts. This experiment showed that small number concatemers were present in nearly every analyzed plasmid, and therefore disproved the hypothesis of the limitation of multimerization through competition of excess end parts.

Another possibility for these results is the circularization of longer concatemers. This circular concatemers are not anymore available for cloning into a plasmid and would therefore be lost for sequencing. There are several publications where intramolecular ligations are reduced by the addition of polyethylene glycol (PEG) to the ligation reaction[122–126].

To test this, an experiment with simultaneous digestion and ligation in the presence of PEG was performed. Deducted from the restriction digestion analysis, the plasmids contained very short concatemers or even single binding sites. Seven plasmids containing larger inserts were sequenced and showed that all except one were contaminated with foreign DNA. The only sequence containing any binding sites was again of bad quality and low number multimerization. Additionally, blunt end ligation of undigested oligonucleotides was tested, but gave only negative results.

Because none of the previous experiments showed a trend in the right direction, the simultaneous digestion and ligation procedure was omitted and a new approach was introduced. A method, called “serial analysis of gene expression (SAGE)” including a technique for the multimerization of small DNA fragments to long concatemers, was chosen[130]. This technique was adopted and was called the “SAGE like approach”. In this approach the multimerization steps of digestion and ligation are separated by purification to remove the end parts and to increase concatemer length. For this experiment a new oligonucleotide was designed, which included four specified highly conserved nucleotides in the random nucleotide part to aid further analysis. The SAGE like approach was tested and resulted in a few positive colonies. They were sequenced, whereas two sequences were negative, but one contained a small number concatemer of three multimerized binding sites in good quality. This, again, raised the question, why there is only one positive colony containing a small number concatemer.

The only reasonable explanation was the circularization of concatemers. Because the addition of PEG had no effect in previous experiments other solutions to solve the circularization problem were searched for. A publication, presenting improvements of the SAGE method, showed that the circularization of concatemers was indeed a problem and was solved by a partial digestion of the concatemers before ligation into the plasmid[132].

The oligonucleotide used in the previous SAGE like multimerization experiment contained no internal restriction site, and therefore a new oligonucleotide meeting these requirements was designed. The SAGE like multimerization experiment was repeated with a partial digestion of the concatemers. The concatemers were divided into three groups according to their length and were subsequently ligated into plasmids. Nearly all selected colonies showed fragments with lengths according to their predefined group. Seven plasmids were sequenced from the different groups and a total of 90 binding site sequences were obtained. Seven of these 90 binding site sequences contained errors, but 83 were still useful. These sequences were aligned and a sequence logo was created to visualize the probability of each nucleotide at each position of the random nucleotide binding site selection part.

The partial digestion of the concatemers finally solved the circularization problem and the sequencing yield could be increased more than tenfold. This was considered a success and the method was ready to be used to multimerize actually selected binding sites.

Circularization of the concatemers seems to be the major problem because a simple partial digestion step increased the efficiency considerably. In retrospective, with this problem in mind, the approach to simultaneously digest and ligate the oligonucleotides is worth reconsidering. Additionally, in the publication of Shore et al. it was shown that DNA fragments with lengths over approximately 250 bp are circularized with high probability and that there is a sharp reduction in this probability below that fragment size[147]. These findings correspond to the results of the previous experiments presented in this study. In the simultaneous digestion and ligation approach nearly all obtained plasmids contained fragments of less than 250 bp in length. This could indicate that all fragments over 250 bp length are circularized and therefore lost for further processing.

Because the SAGE like approach is quite laborious and the circularization problem was solved, the simultaneous digestion and ligation approach could be modified or these two methods could be combined to create a new one. This combined method should benefit from both approaches and therefore shorten the duration of the multimerization procedure. To achieve this, a combined method, should start with a high amount of DNA, like in the SAGE like approach, which is then simultaneously digested and ligated. Subsequently, the circularized concatemers are linearized by partial digestion, purified and ligated into a plasmid. The size of the concatemers is determined by restriction digestion and selected plasmids are sequenced.

5.3 Binding site selection

One of the most sensitive steps in the binding site selection procedure is the recovery of protein-bound DNA fragments. This is usually achieved by purifying the proteins with magnetic beads using affinity tags, in this case a histidine tag. In the literature, it has also been shown that the selected oligonucleotides can be recovered by immunoprecipitation or EMSA[113,114]. This technique was also used to purify the expressed proteins from the crude cell extracts in the present study. During this procedure the magnetic beads formed agglomerates most likely because of the DNA in the extract. For purification of proteins expressed in E. coli this was not a problem because the cell lysis by sonication reduced the DNA strand length by shear force. But this effect can become a big problem for the binding site selection using nuclear extracts of mammalian cells, where chromosomal DNA is present.

This agglomeration can be of random nature by just mechanically raveling DNA around beads and protein. It is also possible, that this is caused by specific interaction of the beads binding to the His-tag of the expressed proteins which are in turn bound to native,

chromosomal DNA. In consequence, cross-linking of the beads would lead to agglomeration. During the preparation of the nuclear extracts, the concentration of monovalent ions is high to inhibit these protein-DNA interactions. Obviously, the salt concentration must be drastically lowered to allow specific interactions for the binding site selection. On the other hand, lowering the salt concentration enables again the proteins to bind to native DNA. Therefore, this problem must be solved in a different way.

A simple possibility is breaking down the DNA mechanically through shear forces, like by sonication. The DNA could be removed by enzymatic digestion by the addition of DNAse I, which has to be inactivated before the binding site selection oligonucleotides are added. Common techniques for this inactivation are the addition of specific inhibitors or by heat inactivation. In the present experiment heat inactivation would lead to protein denaturation and is thereby inadequate. A further way to remove the DNA from crude extracts is the precipitation with polyethyleneimine (PEI)[148]. To remove only DNA but not DNA-bound proteins, the correct precipitation conditions must be determined empirically. The removal of DNA is going to be a major task in the optimization of the binding site selection, if it is going to be used with nuclear extracts of mammalian cells. The best way to accomplish this has to be determined in future studies.

6 References

1. Burg, M., Kwon, E. & Kültz, D. Osmotic regulation of gene expression. *FASEB J.* 1598–1606 (1996). at <http://www.fasebj.org/content/10/14/1598.short>

2. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11,** 3286–3305 (1997).

3. Wodarz, A. & Nusse, R. Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* **14,** 59–88 (1998).

4. Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423,** 448–52 (2003).

5. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423,** 409–14 (2003).

6. Yoshikawa, S., McKinnon, R. D., Kokel, M. & Thomas, J. B. Wnt-mediated axon guidance via the Drosophila Derailed receptor. *Nature* **422,** 583–8 (2003).

7. Schmitt, A. M. *et al.* Wnt-Ryk signalling mediates medial-lateral retinotectal topographic mapping. *Nature* **439,** 31–7 (2006).

8. Liu, Y. *et al.* Ryk-mediated Wnt repulsion regulates posterior-directed growth of corticospinal tract. *Nat. Neurosci.* **8,** 1151–9 (2005).

9. Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19,** 877–90 (2005).

10. Polakis, P. Wnt signaling and cancer. *Genes Dev.* 1837–1851 (2000). doi:10.1101/gad.14.15.1837

11. Van Es, J. H., Barker, N. & Clevers, H. You Wnt some, you lose some: oncogenes in the Wnt signaling pathway. *Curr. Opin. Genet. Dev.* **13,** 28–33 (2003).

12. Welters, H. J. & Kulkarni, R. N. Wnt signaling: relevance to beta-cell biology and diabetes. *Trends Endocrinol. Metab.* **19,** 349–55 (2008).

13. Grant, S. F. a *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38,** 320–3 (2006).

14. Koay, M. A. & Brown, M. a. Genetic disorders of the LRP5-Wnt signalling pathway affecting the skeleton. *Trends Mol. Med.* **11,** 129–37 (2005).

15. Prud'homme, B., Lartillot, N., Balavoine, G., Adoutte, A. & Vervoort, M. Phylogenetic analysis of the Wnt gene family. Insights from lophotrochozoan members. *Curr. Biol.* **12,** 1395 (2002).

16. Nusse, R. An ancient cluster of Wnt paralogues. *Trends Genet.* **17,** 9525 (2001).

17. Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433,** 156–60 (2005).

18. Nusse, R. & Varmus, H. E. Wnt Genes. **69,** (1992).

19. Miller, J. The Wnts. *Genome Biol* 1–15 (2002).

20. Mikels, a J. & Nusse, R. Wnts as ligands: processing, secretion and reception. *Oncogene* **25,** 7461–8 (2006).

21. Franch-Marro, X., Wendler, F., Griffith, J., Maurice, M. M. & Vincent, J.-P. In vivo role of lipid adducts on Wingless. *J. Cell Sci.* **121,** 1587–92 (2008).

22. Kurayoshi, M., Yamamoto, H., Izumi, S. & Kikuchi, A. Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem. J.* **402,** 515–23 (2007).

23. Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337,** 59–64 (2012).

24. Angers, S. & Moon, R. T. Proximal events in Wnt signal transduction. *Nat. Rev. Mol. Cell Biol.* **10,** 468–77 (2009).

25. Minami, Y., Oishi, I., Endo, M. & Nishita, M. Ror-family receptor tyrosine kinases in noncanonical Wnt signaling: their implications in developmental morphogenesis and human diseases. *Dev. Dyn.* **239,** 1–15 (2010).

26. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407,** 530–5 (2000).

27. Wehrli, M. *et al.* arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407,** 527–30 (2000).

28. Hao, H.-X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485,** 195–200 (2012).

29. Koo, B.-K. *et al.* Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* **488,** 665–9 (2012).

30. Mlodzik, M. Planar polarity in the Drosophila eye: a multifaceted view of signaling specificity and cross-talk. *EMBO J.* **18,** 6873–6879 (1999).

31. McEwen, D. G. & Peifer, M. Wnt signaling: Moving in a new direction. *Curr. Biol.* **10,** R562–4 (2000).

32. Wu, J., Klein, T. J. & Mlodzik, M. Subcellular localization of frizzled receptors, mediated by their cytoplasmic tails, regulates signaling pathway specificity. *PLoS Biol.* **2,** E158 (2004).

33. Boutros, M., Paricio, N., Strutt, D. I. & Mlodzik, M. Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and wingless signaling. *Cell* **94,** 109–18 (1998).

34. Strutt, D. Frizzled signalling and cell polarisation in Drosophila and vertebrates. *Development* **130,** 4501–13 (2003).

35. Miller, J. R., Hocking, a M., Brown, J. D. & Moon, R. T. Mechanism and function of signal transduction by the Wnt/beta-catenin and Wnt/Ca2+ pathways. *Oncogene* **18,** 7860–72 (1999).

36. Kühl, M. Non-canonical Wnt signaling in Xenopus: regulation of axis formation and gastrulation. *Semin. Cell Dev. Biol.* **13,** 243–249 (2002).

37. Pandur, P., Läsche, M., Eisenberg, L. & Kühl, M. Wnt-11 activation of a non-canonical Wnt signalling pathway is required for cardiogenesis. *Nature* **418,** 17–19 (2002).

38. Sheldahl, L. C., Park, M., Malbon, C. C. & Moon, R. T. Protein kinase C is differentially stimulated by Wnt and Frizzled homologs in a G-protein-dependent manner. *Curr. Biol.* **9,** 695–8 (1999).

39. Hikasa, H., Shibata, M., Hiratani, I. & Taira, M. The Xenopus receptor tyrosine kinase Xror2 modulates morphogenetic movements of the axial mesoderm and neuroectoderm via Wnt signaling. *Development* **129,** 5227–39 (2002).

40. Slusarski, D., Corces, V. & Moon, R. Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **390,** 410–413 (1997).

41. Wang, H.-Y. & Malbon, C. C. Wnt signaling, Ca2+, and cyclic GMP: visualizing Frizzled functions. *Science* **300,** 1529–30 (2003).

42. Kohn, A. D. & Moon, R. T. Wnt and calcium signaling: beta-catenin-independent pathways. *Cell Calcium* **38,** 439–46 (2005).

43. Slusarski, D. C. & Pelegri, F. Calcium signaling in vertebrate embryonic patterning and morphogenesis. *Dev. Biol.* **307,** 1–13 (2007).

44. Liu, C. *et al.* Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108,** 837–47 (2002).

45. Yost, C., Torres, M. & Miller, J. The axis-inducing activity, stability, and subcellular distribution of beta-catenin is regulated in Xenopus embryos by glycogen synthase kinase 3. *Genes …* 1443–1454 (1996).

46. Polakis, P. Casein kinase 1: a Wnt'er of disconnect. *Curr. Biol.* **12,** 499–501 (2002).

47. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127,** 469–80 (2006).

48. Mao, J., Wang, J., Liu, B., Pan, W. & III, G. F. Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol. Cell* **7,** 801–809 (2001).

49. Umbhauer, M. *et al.* The C-terminal cytoplasmic Lys-thr-X-X-X-Trp motif in frizzled receptors mediates Wnt/beta-catenin signalling. *EMBO J.* **19,** 4944–54 (2000).

50. Hino, S., Michiue, T., Asashima, M. & Kikuchi, A. Casein kinase I epsilon enhances the binding of Dvl-1 to Frat-1 and is essential for Wnt-3a-induced accumulation of beta-catenin. *J. Biol. Chem.* **278,** 14066–73 (2003).

51. Tolwinski, N., Wehrli, M. & Rives, A. Wg/Wnt signal can be transmitted through arrow/LRP5, 6 and Axin independently of Zw3/Gsk3β activity. *Dev. Cell* **4,** 407–418 (2003).

52. Cavallo, R., Cox, R., Moline, M. & Roose, J. Drosophila Tcf and Groucho interact to repress Wingless signalling activity. *Nature* **395,** (1998).

53. Chen, G. & Fernandez, J. A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development. *… Dev.* 2218–2230 (1999). at <http://genesdev.cshlp.org/content/13/17/2218.short>

54. Daniels, D. L. & Weis, W. I. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12,** 364–71 (2005).

55. Mosimann, C., Hausmann, G. & Basler, K. Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nat. Rev. Mol. Cell Biol.* **10,** 276–86 (2009).

56. Nusse, R. The WNT Homepage / Target Genes. at <www.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes>

57. McNeill, H. & Woodgett, J. R. When pathways collide: collaboration and connivance among signalling proteins in development. *Nat. Rev. Mol. Cell Biol.* **11,** 404–13 (2010).

58. Baarsma, H. a, Königshoff, M. & Gosens, R. The WNT signaling pathway from ligand secretion to gene transcription: molecular mechanisms and pharmacological targets. *Pharmacol. Ther.* **138,** 66–83 (2013).

59. Bauer, M. & Willert, K. Wnt signaling: the β-cat (enin)'s meow. *Genes Dev.* (2012). at <http://genesdev.cshlp.org/content/26/2/105.short>

60. Atcha, F. a, Munguia, J. E., Li, T. W. H., Hovanes, K. & Waterman, M. L. A new beta-catenin-dependent activation domain in T cell factor. *J. Biol. Chem.* **278,** 16169–75 (2003).

61. Levanon, D. *et al.* Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11590–5 (1998).

62. Brunner, E., Peter, O., Schweizer, L. & Basler, K. pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. (1997). at <http://www.nature.com/nature/journal/v385/n6619/abs/385829a0.html>

63. Riese, J. *et al.* LEF-1 , a Nuclear Factor Coordinating Signaling Inputs from wingless and decapentaplegic. **88,** 777–787 (1997).

64. Van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88,** 789–99 (1997).

65. Lin, R., Priess, J. F. & Hutchinson, F. pop-l Encodes an HMG Box Protein Required for the Specification of a Mesoderm Precursor in Early C . elegans Embryos POP-1. **83,** (1995).

66. Thorpe, C. J., Schlesinger, A., Carter, J. C. & Bowerman, B. Wnt Signaling Polarizes an Early C . elegans Blastomere to Distinguish Endoderm from Mesoderm. **90,** 695–705 (1997).

67. Hobmayer, B. *et al.* WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. *Nature* **407,** 186–9 (2000).

68. Arce, L., Yokoyama, N. N. & Waterman, M. L. Diversity of LEF/TCF action in development and disease. *Oncogene* **25,** 7492–504 (2006).

69. Cordray, P. & Satterwhite, D. J. TGF-beta induces novel Lef-1 splice variants through a Smad-independent signaling pathway. *Dev. Dyn.* **232,** 969–78 (2005).

70. Carlsson, P., Waterman, M. L. & Jones, K. a. The hLEF/TCF-1 alpha HMG protein contains a context-dependent transcriptional activation domain that induces the TCR alpha enhancer in T cells. *Genes Dev.* **7,** 2418–2430 (1993).

71. Giese, K. & Grosschedl, R. LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. *EMBO J.* **12,** 4667–4676 (1993).

72. Travis, a, Amsterdam, a, Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev.* **5,** 880–894 (1991).

73. Waterman, M. L., Fischer, W. H. & Jones, K. a. A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev.* **5,** 656–669 (1991).

74. Giesel, K., Grosschedll, R. & Wright, P. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* (1995).

75. Giese, K., Pagel, J. & Grosschedl, R. Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12845–50 (1997).

76. Bryan T. MacDonald, Keiko Tamai, and X. H., MacDonald, B., Tamai, K. & He, X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17,** 9–26 (2009).

77. Giese, K., Amsterdam, a & Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5,** 2567–2578 (1991).

78. Eastman, Q. & Grosschedl, R. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr. Opin. Cell Biol.* **11,** 233–40 (1999).

79. Atcha, F. a *et al.* A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol. Cell. Biol.* **27,** 8352–63 (2007).

80. Love, J. J., Li, X., Chung, J., Dyson, H. J. & Wright, P. E. The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry* **43,** 8725–34 (2004).

81. Kikuchi, A., Kishida, S. & Yamamoto, H. Regulation of Wnt signaling by protein-protein interaction and post-translational modifications. *Exp. Mol. Med.* **38,** 1–10 (2006).

82. Hülsken, J., Birchmeier, W. & Behrens, J. E-cadherin and APC compete for the interaction with beta-catenin and the cytoskeleton. *J. Cell Biol.* **127,** 2061–9 (1994).

83. Xing, Y., Takemaru, K., Liu, J. & Berndt, J. Crystal structure of a full-length β-catenin. *Structure* (2008).

84. Gottardi, C. J. & Gumbiner, B. M. Distinct molecular forms of beta-catenin are targeted to adhesive or transcriptional complexes. *J. Cell Biol.* **167,** 339–49 (2004).

85. Mo, R. *et al.* The terminal region of beta-catenin promotes stability by shielding the Armadillo repeats from the axin-scaffold destruction complex. *J. Biol. Chem.* **284,** 28222–31 (2009).

86. Hecht, A., Litterst, C., Huber, O. & Kemler, R. Functional Characterization of Multiple Transactivating Elements in β-Catenin, Some of Which Interact with the TATA-binding Proteinin Vitro. *J. Biol. Chem.* **274,** 18017–25 (1999).

87. Vleminckx, K., Kemler, R. & Hecht, A. The C - terminal transactivation domain of β-catenin is necessary and sufficient for signaling by the Lef-1 / β-catenin complex in Xenopus laevis. *Mech. Dev.* **81,** 65–74 (1999).

88. Takemaru, K. & Moon, R. The transcriptional coactivator CBP interacts with β-catenin to activate gene expression. *J. Cell Biol.* (2000).

89. Barker, N. *et al.* The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation. *EMBO J.* **20,** 4935–43 (2001).

90. Parker, D. S., Jemison, J. & Cadigan, K. M. Pygopus, a nuclear PHD-finger protein required for Wingless signaling in Drosophila. *Development* **129,** 2565–76 (2002).

91. Thompson, B., Townsley, F., Rosin-Arbesfeld, R., Musisi, H. & Bienz, M. A new nuclear component of the Wnt signalling pathway. *Nat. Cell Biol.* **4,** 367–73 (2002).

92. Kramps, T. *et al.* Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. *Cell* **109,** 47–60 (2002).

93. Takemaru, K.-I. *et al.* Chibby, a nuclear beta-catenin-associated antagonist of the Wnt/Wingless pathway. *Nature* **422,** 905–9 (2003).

94. Takemaru, K., Fischer, V. & Li, F. Fine-tuning of nuclear β-catenin by Chibby and 14-3-3. *Cell Cycle* 210–213 (2009).

95. Gottardi, C. J. & Gumbiner, B. M. Role for ICAT in beta-catenin-dependent nuclear signaling and cadherin functions. *Am. J. Physiol. Cell Physiol.* **286,** C747–56 (2004).

96. Daniels, D. L. & Weis, W. I. ICAT inhibits beta-catenin binding to Tcf/Lef-family transcription factors and the general coactivator p300 using independent structural modules. *Mol. Cell* **10,** 573–84 (2002).

97. Kimelman, D. & Xu, W. Beta-Catenin Destruction Complex: Insights and Questions From a Structural Perspective. *Oncogene* **25,** 7482–91 (2006).

98. Maher, M. T., Mo, R., Flozak, A. S., Peled, O. N. & Gottardi, C. J. Beta-catenin phosphorylated at serine 45 is spatially uncoupled from beta-catenin phosphorylated in the GSK3 domain: implications for signaling. *PLoS One* **5,** e10184 (2010).

99. Valenta, T., Hausmann, G. & Basler, K. The many faces and functions of β-catenin. *EMBO J.* **31,** 2714–36 (2012).

100. Graham, T. a, Weaver, C., Mao, F., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/Tcf complex. *Cell* **103,** 885–96 (2000).

101. Graham, T. a, Ferkey, D. M., Mao, F., Kimelman, D. & Xu, W. Tcf4 can specifically recognize beta-catenin using alternative conformations. *Nat. Struct. Biol.* **8,** 1048–52 (2001).

102. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316,** 1497–502 (2007).

103. Park, P. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10,** 669–680 (2009).

104. Oliphant, A., Brandl, C. & Struhl, K. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell. Biol.* (1989). doi:10.1128/MCB.9.7.2944.Updated

105. Ellington, A. & Szostak, J. In vitro selection of RNA molecules that bind specific ligands. *Nature* (1990). at <http://www.ff.ul.pt/FCT/PTDC/BIM-MEC/0444/2012/24.pdf>

106. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-.).* (1990). at <http://www.sciencemag.org/content/249/4968/505.short>

107. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **24,** 381–403 (2007).

108. Pollock, R. & Treisman, R. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18,** 6197–204 (1990).

109. Funk, W. & Wright, W. Cyclic amplification and selection of targets for multicomponent complexes: Myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor. *… Natl. Acad. Sci.* **89,** 9484–9488 (1992).

110. Wright, W. E., Binder, M. & Funk, W. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.* **11,** 4104–10 (1991).

111. Gee, M. S., Sarkisian, C. J. & el-Deiry, W. S. Identification of a novel AP-2 consensus DNA binding site. *Biochem. Biophys. Res. Commun.* **243,** 307–16 (1998).

112. Newhouse, C. D. & Silverstein, S. Orientation of a Novel DNA Binding Site Affects Human Papillomavirus-Mediated Transcription and Replication. *J. Virol.* **75,** 1722–1735 (2001).

113. Choi, Y. S. & Sinha, S. Determination of the consensus DNA-binding sequence and a transcriptional activation domain for ESE-2. *Biochem. J.* **398,** 497–507 (2006).

114. Li, G., Li, W., Angelastro, J. M., Greene, L. a & Liu, D. X. Identification of a novel DNA binding site and a transcriptional target for activating transcription factor 5 in c6 glioma and mcf-7 breast cancer cells. *Mol. Cancer Res.* **7,** 933–43 (2009).

115. Donaldson, N. S. *et al.* Kaiso regulates Znf131-mediated transcriptional activation. *Exp. Cell Res.* **316,** 1692–705 (2010).

116. Larouche, K., Bergeron, M.-J., Leclerc, S. & Guerin, S. Optimization of Competitor Poly(dI-dC).Poly(dI- dC) Levels is Advised in DNA-Protein Interaction Studies Involving En- riched Nuclear Proteins. *Biotechniques* **20,** 439–444 (1996).

117. Scientfic, Thermo (IL, U. SDS-PAGE band profile of the Thermo Scientific PageRuler Prestained Protein Ladder. at <http://www.piercenet.com/browse.cfm?fldID=717EAB22-C50E-319F-D227-C1EB41C4343C>

118. Kumar, S., Tamura, K. & Nei, M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. …* (1994). at <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:MEGA+::+Molecular+Evolutionary+Genetics+Analysis#5>

119. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–9 (2011).

120. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **260,** 649–63 (1996).

121. Holz, M. & Mayele, M. 8 Influence of Additives on Hydrophobic Association in Polynary Aqueous Mixtures. An NMR Relaxation and Self-Diffusion Study. *researchgate.net* 150–183

122. Hayashi, K. & Nakazawa, M. Regulation of inter-and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic acids …* **14,** 7617–7631 (1986).

123. Harrison, B. & Zimmerman, S. Polymer-stimulated ligation: enhanced ligation of oligo- and polynucleotides by T4 RNA ligase in polymer solutions. *Nucleic Acids Res.* **12,** 8235–8251 (1984).

124. Pheiffer, B. H. & Zimmerman, S. B. Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. *Nucleic Acids Res.* **1,** 7853–7871 (1983).

125. Sobczak, J. & Duguet, M. Effect of histone H1, poly(ethyleneglycol) and DNA concentration on intermolecular and intramolecular ligation by T4 DNA ligase. *Eur. J. Biochem.* **175,** 379–85 (1988).

126. Hayashi, K. & Nakazawa, M. Influence of monovalent cations on the activity of T4 DNA ligase in the presence of polyethylene glycol. *Nucleic acids …* **13,** (1985).

127. Ken'ichiro Hayashi, Masako Nakazawa, Yukuo Ishizaki, N. H. and A. O. Stimulation of intermolecular ligation with E. coli DNA ligase by high concentrations of monovalent cations in polyethylene glycol solutions. **13,** 7979–7992 (1985).

128. TERAOKA, H. & TSUKADA, K. Influence of Polyethylene Glycol on the Ligation Reaction with Calf Thyinus DNA Ligases I and II. *J. Biochem.* **101,** 225–231 (1987).

129. Thermo Fisher Scientific Inc. T4 DNA Ligase - Product Manual. at <http://www.thermoscientificbio.com/dna-and-rna-modifying-enzymes/t4-dna-ligase/>

130. Velculescu, V. & Zhang, L. Serial analysis of gene expression. *Sci. …* (1995).

131. Wetering, M. Van De & Clevers, H. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson -Crick double helix. *EMBO J.* **1,** 3039–3044 (1992).

132. Gowda, M. & Jantasuriyarat, C. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiol.* **134,** 890–897 (2004).

133. RRNYI, A. On measures of entropy and information. *Fourth Berkeley Symp. Math. …* **547,** (1961).

134. Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B. & Baig, A. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res.* **30,** 2911–9 (2002).

135. EMBL. Protein Expression - E.coli - Improving Protein Solubility. at <http://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/improving_protein_solubility/>

136. Nakamoto, H. & Bardwell, J. C. a. Catalysis of disulfide bond formation and isomerization in the Escherichia coli periplasm. *Biochim. Biophys. Acta* **1694,** 111–9 (2004).

137. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol.* **115,** 113–28 (2005).

138. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17,** 353–8 (2006).

139. Oganesyan, N., Ankoudinova, I., Kim, S. & Kim, R. Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein Expr. …* **52,** 280–285 (2007).

140. Chen, J., Acton, T. B., Basu, S. K., Montelione, G. T. & Inouye, M. Enhancement of the solubility of proteins overexpressed in Escherichia coli by heat shock. *J. Mol. Microbiol. Biotechnol.* **4,** 519–24 (2002).

141. EMBL. Protein Purification Extraction and Clarification. at <http://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/removal_dna/>

142. Dignam, J. D., Lebovitz, R. M., Roeder, R. G. & Limited, I. R. L. P. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. **11,** 1475–1489 (1983).

143. Tyndall, C., Mantia, G. La & Robert, K. the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication. *Nucleic acids …* **9,** 6231–6250 (1981).

144. Blackwell, T. & Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science (80-.).* **250,** 1104–1110 (1990).

145. Denissova, N. G., Pouponnot, C., Long, J., He, D. & Liu, F. Transforming growth factor beta -inducible independent binding of SMAD to the Smad7 promoter. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6397–402 (2000).

146. Hewetson, a. Identification of the RUSH Consensus-Binding Site by Cyclic Amplification and Selection of Targets: Demonstration that RUSH Mediates the Ability of Prolactin to Augment Progesterone-Dependent Gene Expression. *Mol. Endocrinol.* **16,** 2101–2112 (2002).

147. Shore, D., Langiwski, J. & Baldwin, R. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. …* **78,** 4833–4837 (1981).

148. Burgess, R. R. Use of polyethyleneimine in purification of DNA-binding proteins. *Methods Enzymol.* **208,** 3–10 (1991).

7 Appendix

7.1 Abbreviations

APC	Adenomatous polyposis coli
APS	Ammonium persulfate
AR	Androgen receptor
ATP	Adenosine triphosphate
βCBD	β-catenin binding domain
βTrCP	β-transducin repeat-containing protein
Bcl9	B-cell CLL/Lymphoma 9 protein
bp	Base pair
Brg-1	Brahma/Brahma-related gene-1
BSA	Bovine serum albumin
CASTing	Cyclic amplification and selection of targets
CBP	CREB binding protein
CDRD	Contend-dependent regulatory domain
ChIP	Chromatin immunoprecipitation
CK1	Casein kinase 1
CMV	Cytomegalovirus
CREB	cAMP-responsive element-binding protein
CTD	Carboxy-terminal domain
Da	Dalton [g/mol]
DMEM	Dulbecco’s modified Eagle's medium
dn	Dominant negative
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
ds	Double stranded
DTT	Dithiothreitol
Dvl	Dishevelled
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic mobility shift assay
EtOH	Ethanol
FCS	Fetal calf serum
FOXO	Forkhead box protein O
FRAT	Frequently rearranged in advanced T-cell lymphomas
GCN4	General control protein 4
GFP	Green fluorescent protein

GSK3	Glycogen synthase kinase-3
GTP	Guanosine triphosphate
GTPase	Guanosine triphosphate hydrolase
HAT	Histone acetyltransferase
HDAC	Histone deacetylases
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICAT	Inhibitor of β-catenin and Tcf4
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
JNK	c-Jun N-terminal kinases
LB	Lysogeny broth
Lef	Lymphoid enhancer factor
LoTE	Low concentration Tris/EDTA
LRH-1	Liver receptor homologue 1
LRP5/6	Low density lipoprotein receptor related proteins 5 and 6
M	Molarity [mol/L]
MB	Magnetic bead binding buffer
MED 12	Mediator of RNA polymerase II transcription subunit 12
MLL-1	Mixed lineage leukaemia
NLS	Nuclear localization signal
NTD	Amino-terminal domain
OD	Optical density
ORF	Open reading frame
ori	Origin of replication
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCP	Planar cell polarity
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
PHD	Plant Homeo Domain finger protein
PKC	Protein kinase C
PMSF	Phenylmethylsulfonyl fluoride
poly(dI-dC)	Poly(deoxyinosinic-deoxycytidylic) acid
RING	Really interesting new gene 1 protein
RNA	Ribonucleic acid
RNF43	RING finger protein 43
ROR	Receptor tyrosine kinase-like orphan receptors
rpm	Revolutions per minute
RT	Room temperature

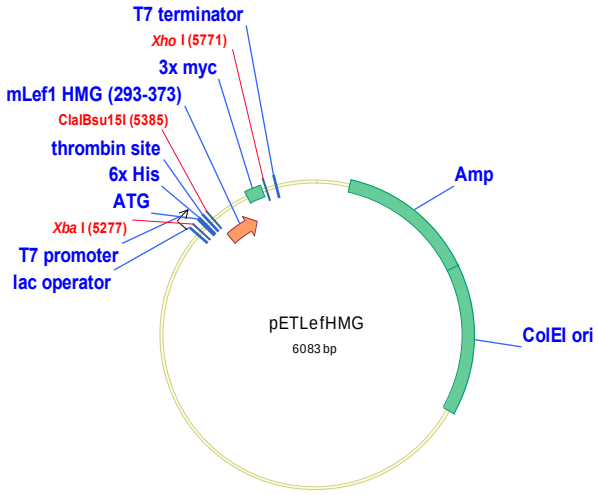
rxn	Reaction
RYK	Receptor tyrosine kinase
SDS	Sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment
Seq	Sequencing
SMARCA4	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 4
SNF	Sucrose NonFermentable
ss	Single stranded
SWI	SWItch
TBE	Tris/Borate/EDTA
TBE	Tris/Borate/EDTA
TBP	TATA-binding protein
Tcf	T-cell factor
TEMED	Tetramethylethylenediamine
TLE	Transducin-like enhancer protein
TN	Tris/NaCl
TN-NP40	Tris/NaCl + NP-40
TN-T	Tris/NaCl + Tween20
TN-TX	Tris/NaCl + Triton X-100
TRRAP	Transformation/transcription domain-associated protein
UV	Ultraviolet
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ZNRF3	Zinc/RING finger protein 3

7.2 Vector maps

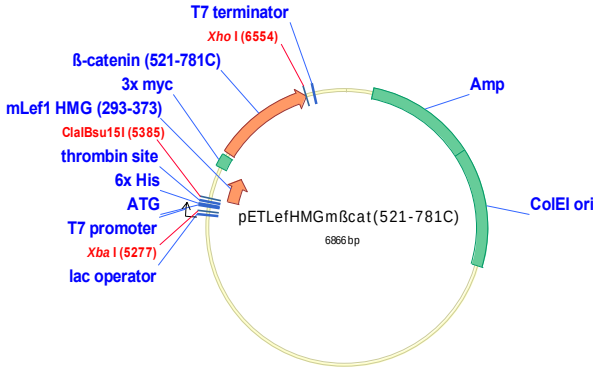
Vector maps were generated by VectorNTI

7.2.1 Vectors for protein expression in E. coli

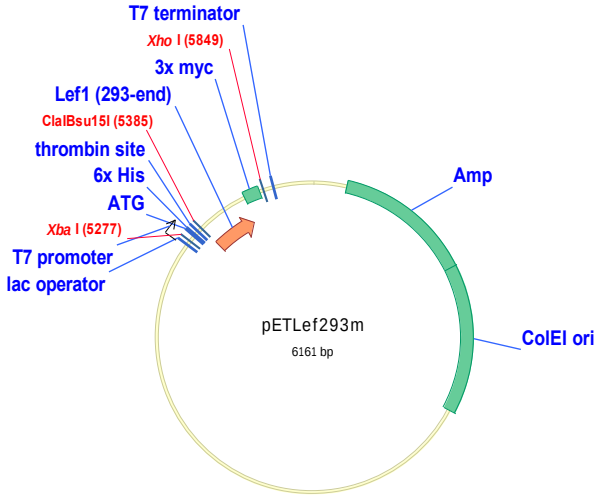
LefHMG



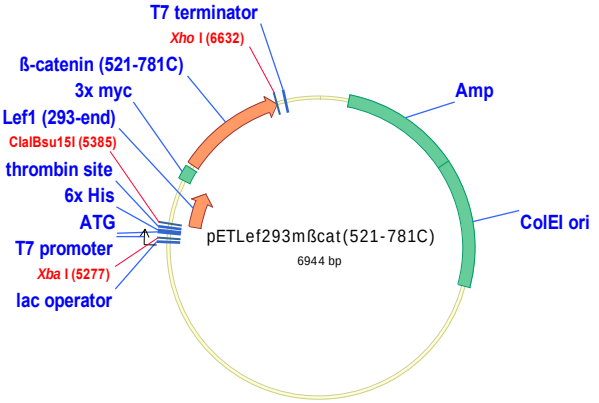
LefHMGβcat



Lef293

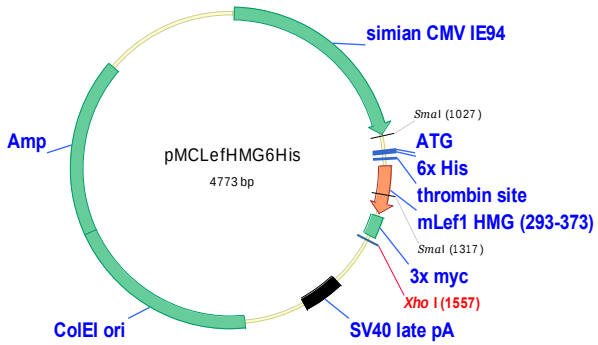


Lef293βcat

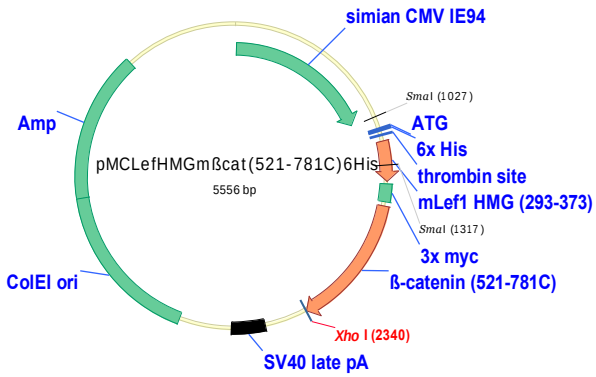


7.2.2 Vectors for protein expression in HeLa

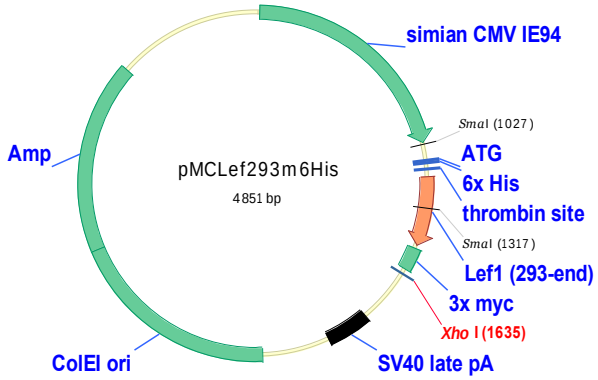
LefHMG



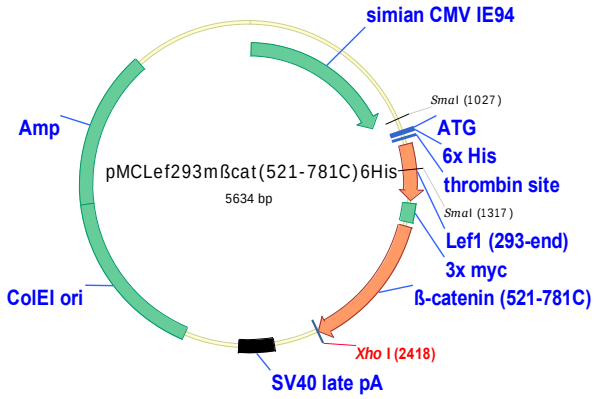
LefHMGβcat



Lef293

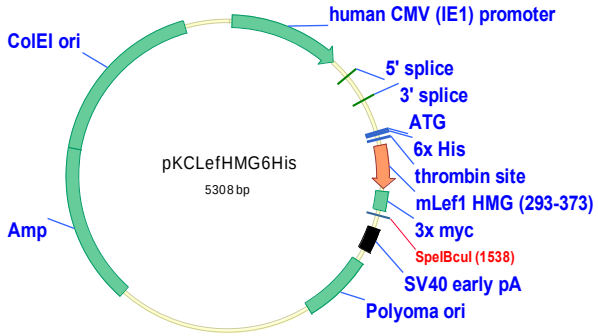


Lef293βcat

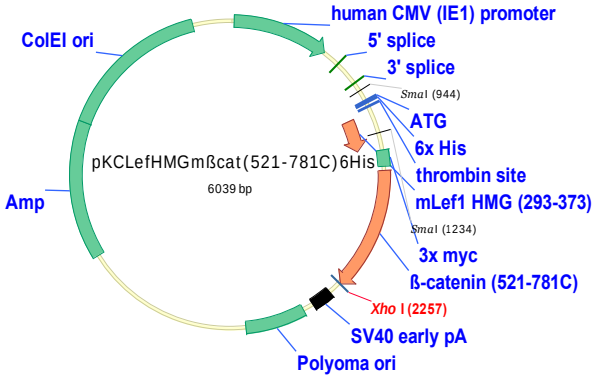


7.2.3 Vectors for protein expression in COP-8

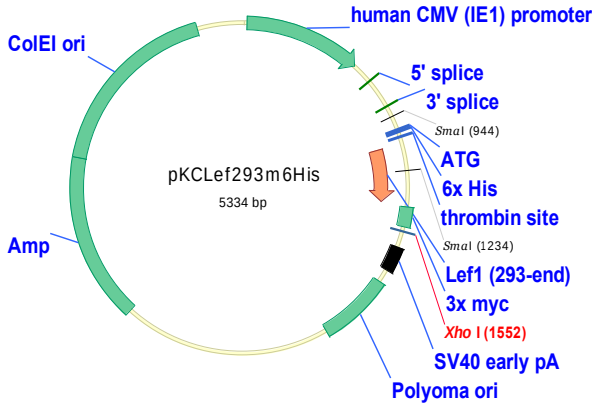
LefHMG



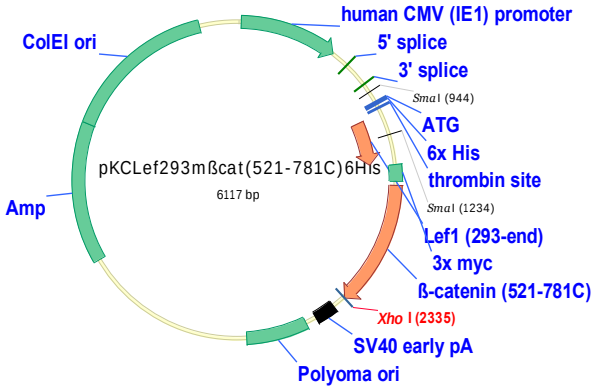
LefHMGβcat



Lef293



Lef293βcat



7.3 Protein properties

All protein sequences and amino acid compositions were calculated by VectorNTI.

7.3.1 LefHMG

7.3.1.1 DNA sequence of the ORF

```
ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG CGC
GGC AGC CAT ATG GCT ATC GAT AAA AGA CCT CAT ATT AAG AAG CCT CTG AAT
GCT TTC ATG TTA TAT ATG AAA GAA ATG AGA GCG AAT GTC GTA GCT GAG TGC
ACG CTA AAG GAG AGT GCA GCT ATC AAC CAG ATC CTG GGC AGA AGA TGG CAC
GCC CTC TCC CGG GAA GAG CAG GCC AAA TAC TAT GAA CTA GCA CGG AAA GAG
AGA CAG CTA CAC ATG CAG CTT TAT CCA GGC TGG TCA GCG CGA GAC AAT TAT
GGC AAG CTT ACT AGA GGT GAA CAA AAG TTG ATT TCT GAA GAA GAT TTG AAC
GGT GAA CAA AAG CTA ATC TCC GAG GAA GAC TTG AAC GGT GAA CAA AAA TTA
ATC TCA GAA GAA GAC TTG AAC GGA TCT ACT AGA GAA TTC GGA TCC CTC GAG
ACT AGT
```

7.3.1.2 Amino acid sequence

```
MGSSHHHHHH SSGLVPRGSH MAIDKRPHIK KPLNAFMLYM KEMRANVVAE CTLKESAAIN
QILGRRWHAL SREEQAKYYE LARKERQLHM QLYPGWSARD NYGKLTRGEQ KLISEEDLNG
EQKLISEEDL NGEQKLISEE DLNGSTREFG SLETS
```

7.3.1.3 Protein properties and Amino acid composition

Analysis	Entire Protein
Length	155 aa
Molecular Weight	17847.95
1 µg =	56,029 pmol
Molar Extinction coefficient	17900
1 A[280] corr. to	1,00 mg/mL
A[280] of 1 mg/mL	1,00 AU
Isoelectric Point	6.87
Charge at pH 7	-0.34

Amino Acids	Number count	% by weight	% by frequency
Charged (RKHYCDE)	61	45.66	39.35
Acidic (DE)	23	16.07	14.84
Basic (KR)	22	17.09	14.19
Polar (NCQSTY)	38	23.87	24.52
Hydrophobic (AILFWV)	41	24.87	26.45
A Ala	10	4.32	6.45
C Cys	1	0.59	0.65
D Asp	5	3.23	3.23
E Glu	18	12.84	11.61
F Phe	2	1.60	1.29
G Gly	11	4.00	7.10
H His	10	7.52	6.45
I Ile	7	4.45	4.52
K Lys	11	7.80	7.10
L Leu	17	10.81	10.97
M Met	6	4.34	3.87
N Asn	7	4.48	4.52
P Pro	4	2.23	2.58
Q Gln	7	4.96	4.52
R Arg	11	9.29	7.10
S Ser	14	7.13	9.03
T Thr	4	2.31	2.58
V Val	3	1.70	1.94
W Trp	2	1.98	1.29
Y Tyr	5	4.39	3.23
B Asx	12	7.71	7.74
Z Glx	25	17.80	16.13

Amino Acids	Codons Used
A Ala	GCA(2), GCC(2), GCG(2), GCT(4)
C Cys	TGC(1)
D Asp	GAC(3), GAT(2)
E Glu	GAA(12), GAG(6)
F Phe	TTC(2)
G Gly	GGA(2), GGC(6), GGT(3)
H His	CAC(3), CAT(7)
I Ile	ATC(5), ATT(2)
K Lys	AAA(5), AAG(6)
L Leu	CTA(4), CTC(2), CTG(3), CTT(2),TTA(2), TTG(4)
M Met	ATG(6)
N Asn	AAC(4), AAT(3)
P Pro	CCA(1), CCG(1), CCT(2)
Q Gln	CAA(3), CAG(4)
R Arg	AGA(7), CGA(1), CGC(1), CGG(2)
S Ser	AGC(5), AGT(2), TCA(2), TCC(3), TCT(2)
T Thr	ACG(1), ACT(3)
V Val	GTA(1), GTC(1), GTG(1)
W Trp	TGG(2)
Y Tyr	TAC(1), TAT(4)

7.3.2 LefHMGβcat

7.3.2.1 DNA sequence of the ORF

```
ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG CGC
GGC AGC CAT ATG GCT ATC GAT AAA AGA CCT CAT ATT AAG AAG CCT CTG AAT
GCT TTC ATG TTA TAT ATG AAA GAA ATG AGA GCG AAT GTC GTA GCT GAG TGC
ACG CTA AAG GAG AGT GCA GCT ATC AAC CAG ATC CTG GGC AGA AGA TGG CAC
GCC CTC TCC CGG GAA GAG CAG GCC AAA TAC TAT GAA CTA GCA CGG AAA GAG
AGA CAG CTA CAC ATG CAG CTT TAT CCA GGC TGG TCA GCG CGA GAC AAT TAT
GGC AAG CTT ACT AGA GGT GAA CAA AAG TTG ATT TCT GAA GAA GAT TTG AAC
GGT GAA CAA AAG CTA ATC TCC GAG GAA GAC TTG AAC GGT GAA CAA AAA TTA
ATC TCA GAA GAA GAC TTG AAC GGA TCT ACT AGA GAA TTC GGA TCC CCC GCA
AAT CAT GCA CCT TTG CGT GAG CAG GGT GCC ATT CCA CGA CTA GTT CAG TTG
CTT GTT CGT GCA CAT CAG GAT ACC CAG CGC CGT ACG TCC ATG GGT GGG ACA
CAG CAG CAA TTT GTG GAG GGG GTC CGC ATG GAA GAA ATA GTT GAA GGT TGT
ACC GGA GCC CTT CAC ATC CTA GCT CGG GAT GTT CAC AAC CGA ATT GTT ATC
AGA GGA CTA AAT ACC ATT CCA TTG TTT GTG CAG CTG CTT TAT TCT CCC ATT
GAA AAC ATC CAA AGA GTA GCT GCA GGG GTC CTC TGT GAA CTT GCT CAG GAC
AAG GAA GCT GCA GAA GCT ATT GAA GCT GAG GGA GCC ACA GCT CCT CTG ACA
GAG TTA CTT CAC TCT AGG AAT GAA GGT GTG GCG ACA TAT GCA GCT GCT GTT
TTG TTC CGA ATG TCT GAG GAC AAG CCA CAA GAT TAC AAG AAA CGG CTT TCA
GTT GAG CTG ACC AGC TCT CTC TTC AGA ACA GAG CCA ATG GCT TGG AAT GAG
ACT GCT GAT CTT GGA CTT GAT ATT GGT GCC CAG GGA GAA CCC CTT GGA TAT
CGC CAG GAT GAT CCT AGC TAT CGT TCT TTT CAC TCT GGT GGA TAT GGC CAG
GAT GCC TTG GGT ATG GAC CCC ATG ATG GAA CAT GAG ATG GGT GGC CAC CAC
CCT GGT GCT GAC TAT CCA GTT GAT GGG CTG CCA GAT CTG GGG CAT GCC CAG
GAC CTC ATG GAT GGG CTG CCT CCA GGT GAC AGC AAT CAG CTG GCC TGG TTT
GAT ACT GAC CTG CTC GAG ACT AGT
```

7.3.2.2 Amino acid sequence

```
MGSSHHHHHH SSGLVPRGSH MAIDKRPHIK KPLNAFMLYM KEMRANVVAE CTLKESAAIN
QILGRRWHAL SREEQAKYYE LARKERQLHM QLYPGWSARD NYGKLTRGEQ KLISEEDLNG
EQKLISEEDL NGEQKLISEE DLNGSTREFG SPANHAPLRE QGAIPRLVQL LVRAHQDTQR
RTSMGGTQQQ FVEGVRMEEI VEGCTGALHI LARDVHNRIV IRGLNTIPLF VQLLYSPIEN
IQRVAAGVLC ELAQDKEAAE AIEAEGATAP LTELLHSRNE GVATYAAAVL FRMSEDKPQD
YKKRLSVELT SSLFRTEPMA WNETADLGLD IGAQGEPLGY RQDDPSYRSF HSGGYGQDAL
GMDPMMEHEM GGHHPGADYP VDGLPDLGHA QDLMDGLPPG DSNQLAWFDT DLLETS
```

7.3.2.3 Protein properties and Amino acid composition

Analysis	Entire Protein
Length	416 aa
Molecular Weight	46432.19
1 µg =	21,537 pmol
Molar Extinction coefficient	38480
1 A[280] corr. to	1,21 mg/mL
A[280] of 1 mg/mL	0,83 AU
Isoelectric Point	5.44
Charge at pH 7	-18.42

Amino Acids	Number count	% by weight	% by frequency
Charged (RKHYCDE)	139	39.55	33.41
Acidic (DE)	62	16.30	14.90
Basic (KR)	42	12.79	10.10
Polar (NCQSTY)	93	22.78	22.36
Hydrophobic (AILFWV)	128	28.94	30.77
A Ala	36	5.95	8.65
C Cys	3	0.67	0.72
D Asp	24	5.93	5.77
E Glu	38	10.37	9.13
F Phe	8	2.45	1.92
G Gly	36	5.01	8.65
H His	20	5.76	4.81
I Ile	17	4.14	4.09
K Lys	15	4.07	3.61
L Leu	46	11.19	11.06
M Met	15	4.15	3.61
N Asn	14	3.43	3.37
P Pro	20	4.27	4.81
Q Gln	23	6.24	5.53
R Arg	27	8.72	6.49
S Ser	25	4.87	6.01
T Thr	16	3.54	3.85
V Val	17	3.69	4.09
W Trp	4	1.52	0.96
Y Tyr	12	4.03	2.88
B Asx	38	9.34	9.13
Z Glx	61	16.61	14.66

Amino Acids	Codons Used
A Ala	GCA(8), GCC(9), GCG(3), GCT(16)
C Cys	TGC(1), TGT(2)
D Asp	GAC(10), GAT(14)
E Glu	GAA(23), GAG(15)
F Phe	TTC(4), TTT(4)
G Gly	GGA(9), GGC(8), GGG(6), GGT(13)
H His	CAC(9), CAT(11)
I Ile	ATA(1), ATC(8), ATT(8)
K Lys	AAA(6), AAG(9)
L Leu	CTA(7), CTC(5), CTG(11), CTT(11), TTA(3), TTG(9)
M Met	ATG(15)
N Asn	AAC(6), AAT(8)
P Pro	CCA(8), CCC(4), CCG(1), CCT(7)
Q Gln	CAA(6), CAG(17)
R Arg	AGA(10), AGG(1), CGA(4), CGC(4),CGG(4), CGT(4)
S Ser	AGC(8), AGT(2), TCA(3), TCC(4), TCT(8)
T Thr	ACA(5), ACC(4), ACG(2), ACT(5)
V Val	GTA(2), GTC(3), GTG(4), GTT(8)
W Trp	TGG(4)
Y Tyr	TAC(2), TAT(10)

7.3.3 Lef293

7.3.3.1 DNA sequence of the ORF

```
ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG CGC
GGC AGC CAT ATG GCT ATC GAT AAA AGA CCT CAT ATT AAG AAG CCT CTG AAT
GCT TTC ATG TTA TAT ATG AAA GAA ATG AGA GCG AAT GTC GTA GCT GAG TGC
ACG CTA AAG GAG AGT GCA GCT ATC AAC CAG ATC CTG GGC AGA AGA TGG CAC
GCC CTC TCC CGG GAA GAG CAG GCC AAA TAC TAT GAA CTA GCA CGG AAA GAG
AGA CAG CTA CAC ATG CAG CTT TAT CCA GGC TGG TCA GCG CGA GAC AAT TAT
GGC AAG AAG AAG AAG AGG AAG AGA GAG AAG CTA CAG GAG TCG ACT TCA GGT
ACA GGT CCC AGA ATG ACA GCT GCC TAC ATC AAG CTT ACT AGA GGT GAA CAA
AAG TTG ATT TCT GAA GAA GAT TTG AAC GGT GAA CAA AAG CTA ATC TCC GAG
GAA GAC TTG AAC GGT GAA CAA AAA TTA ATC TCA GAA GAA GAC TTG AAC GGA
TCT ACT AGA GAA TTC GGA TCC CTC GAG ACT AGT
```

7.3.3.2 Amino acid sequence

```
MGSSHHHHHH SSGLVPRGSH MAIDKRPHIK KPLNAFMLYM KEMRANVVAE CTLKESAAIN
QILGRRWHAL SREEQAKYYE LARKERQLHM QLYPGWSARD NYGKKKKRKR EKLQESTSGT
GPRMTAAYIK LTRGEQKLIS EEDLNGEQKL ISEEDLNGEQ KLISEEDLNG STREFGSLET
S
```

7.3.3.3 Protein properties and Amino acid composition

Analysis	Entire Protein
Length	181 aa
Molecular Weight	20823.23 Da
1 µg =	48,023 pmol
Molar Extinction coefficient	19180
1 A[280] corr. to	1,09 mg/mL
A[280] of 1 mg/mL	0,92 AU
Isoelectric Point	9.36
Charge at pH 7	6.66

Amino Acids	Number count	% by weight	% by frequency
Charged (RKHYCDE)	73	46.92	40.33
Acidic (DE)	25	14.99	13.81
Basic (KR)	31	20.46	17.13
Polar (NCQSTY)	45	24.17	24.86
Hydrophobic (AILFWV)	45	23.14	24.86
A Ala	12	4.44	6.63
C Cys	1	0.50	0.55
D Asp	5	2.77	2.76
E Glu	20	12.23	11.05
F Phe	2	1.37	1.1
G Gly	13	4.06	7.18
H His	10	6.45	5.52
I Ile	8	4.36	4.42
K Lys	17	10.33	9.39
L Leu	18	9.81	9.94
M Met	7	4.34	3.87
N Asn	7	3.84	3.87
P Pro	5	2.39	2.76
Q Gln	8	4.86	4.42
R Arg	14	10.13	7.73
S Ser	16	6.99	8.84
T Thr	7	3.46	3.87
V Val	3	1.46	1.66
W Trp	2	1.70	1.1
Y Tyr	6	4.51	3.31
B Asx	12	6.61	6.63
Z Glx	28	17.08	15.47

Amino Acids	Codons Used
A Ala	GCA(2), GCC(3), GCG(2), GCT(5)
C Cys	TGC(1)
D Asp	GAC(3), GAT(2)
E Glu	GAA(12), GAG(8)
F Phe	TTC(2)
G Gly	GGA(2), GGC(6), GGT(5)
H His	CAC(3), CAT(7)
I Ile	ATC(6), ATT(2)
K Lys	AAA(5), AAG(12)
L Leu	CTA(5), CTC(2), CTG(3), CTT(2), TTA(2), TTG(4)
M Met	ATG(7)
N Asn	AAC(4), AAT(3)
P Pro	CCA(1), CCC(1), CCG(1), CCT(2)
Q Gln	CAA(3), CAG(5)
R Arg	AGA(9), AGG(1), CGA(1), CGC(1), CGG(2)
S Ser	AGC(5), AGT(2), TCA(3), TCC(3), TCG(1), TCT(2)
T Thr	ACA(2), ACG(1), ACT(4)
V Val	GTA(1), GTC(1), GTG(1)
W Trp	TGG(2)
Y Tyr	TAC(2), TAT(4)

7.3.4 Lef293βcat

7.3.4.1 DNA sequence of the ORF

```
ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG CGC
GGC AGC CAT ATG GCT ATC GAT AAA AGA CCT CAT ATT AAG AAG CCT CTG AAT
GCT TTC ATG TTA TAT ATG AAA GAA ATG AGA GCG AAT GTC GTA GCT GAG TGC
ACG CTA AAG GAG AGT GCA GCT ATC AAC CAG ATC CTG GGC AGA AGA TGG CAC
GCC CTC TCC CGG GAA GAG CAG GCC AAA TAC TAT GAA CTA GCA CGG AAA GAG
AGA CAG CTA CAC ATG CAG CTT TAT CCA GGC TGG TCA GCG CGA GAC AAT TAT
GGC AAG AAG AAG AAG AGG AAG AGA GAG AAG CTA CAG GAG TCG ACT TCA GGT
ACA GGT CCC AGA ATG ACA GCT GCC TAC ATC AAG CTT ACT AGA GGT GAA CAA
AAG TTG ATT TCT GAA GAA GAT TTG AAC GGT GAA CAA AAG CTA ATC TCC GAG
GAA GAC TTG AAC GGT GAA CAA AAA TTA ATC TCA GAA GAA GAC TTG AAC GGA
TCT ACT AGA GAA TTC GGA TCC CCC GCA AAT CAT GCA CCT TTG CGT GAG CAG
GGT GCC ATT CCA CGA CTA GTT CAG TTG CTT GTT CGT GCA CAT CAG GAT ACC
CAG CGC CGT ACG TCC ATG GGT GGG ACA CAG CAG CAA TTT GTG GAG GGG GTC
CGC ATG GAA GAA ATA GTT GAA GGT TGT ACC GGA GCC CTT CAC ATC CTA GCT
CGG GAT GTT CAC AAC CGA ATT GTT ATC AGA GGA CTA AAT ACC ATT CCA TTG
TTT GTG CAG CTG CTT TAT TCT CCC ATT GAA AAC ATC CAA AGA GTA GCT GCA
GGG GTC CTC TGT GAA CTT GCT CAG GAC AAG GAA GCT GCA GAA GCT ATT GAA
GCT GAG GGA GCC ACA GCT CCT CTG ACA GAG TTA CTT CAC TCT AGG AAT GAA
GGT GTG GCG ACA TAT GCA GCT GCT GTT TTG TTC CGA ATG TCT GAG GAC AAG
CCA CAA GAT TAC AAG AAA CGG CTT TCA GTT GAG CTG ACC AGC TCT CTC TTC
AGA ACA GAG CCA ATG GCT TGG AAT GAG ACT GCT GAT CTT GGA CTT GAT ATT
GGT GCC CAG GGA GAA CCC CTT GGA TAT CGC CAG GAT GAT CCT AGC TAT CGT
TCT TTT CAC TCT GGT GGA TAT GGC CAG GAT GCC TTG GGT ATG GAC CCC ATG
ATG GAA CAT GAG ATG GGT GGC CAC CAC CCT GGT GCT GAC TAT CCA GTT GAT
GGG CTG CCA GAT CTG GGG CAT GCC CAG GAC CTC ATG GAT GGG CTG CCT CCA
GGT GAC AGC AAT CAG CTG GCC TGG TTT GAT ACT GAC CTG CTC GAG ACT AGT
```

7.3.4.2 Amino acid sequence

```
MGSSHHHHHH SSGLVPRGSH MAIDKRPHIK KPLNAFMLYM KEMRANVVAE CTLKESAAIN
QILGRRWHAL SREEQAKYYE LARKERQLHM QLYPGWSARD NYGKKKKRKR EKLQESTSGT
GPRMTAAYIK LTRGEQKLIS EEDLNGEQKL ISEEDLNGEQ KLISEEDLNG STREFGSPAN
HAPLREQGAI PRLVQLLVRA HQDTQRRTSM GGTQQQFVEG VRMEEIVEGC TGALHILARD
VHNRIVIRGL NTIPLFVQLL YSPIENIQRV AAGVLCELAQ DKEAAEAIEA EGATAPLTEL
LHSRNEGVAT YAAAVLFRMS EDKPQDYKKR LSVELTSSLF RTEPMAWNET ADLGLDIGAQ
GEPLGYRQDD PSYRSFHSGG YGQDALGMDP MMEHEMGGHH PGADYPVDGL PDLGHAQDLM
DGLPPGDSNQ LAWFDTDLLE TS
```

7.3.4.3 Protein properties and Amino acid composition

Analysis	Entire Protein
Length	442 aa
Molecular Weight	49407.47
1 µg =	20,240 pmol
Molar Extinction coefficient	39760
1 A[280] corr. to	1,24 mg/mL
A[280] of 1 mg/mL	0,80 AU
Isoelectric Point	5.87
Charge at pH 7	-11.42

Amino Acids	Number count	% by weight	% by frequency
Charged (RKHYCDE)	151	40.44	34.16
Acidic (DE)	64	15.83	14.48
Basic (KR)	51	14.46	11.54
Polar (NCQSTY)	100	22.97	22.62
Hydrophobic (AILFWV)	132	27.97	29.86
A Ala	38	5.90	8.60
C Cys	3	0.63	0.68
D Asp	24	5.57	5.43
E Glu	40	10.26	9.05
F Phe	8	2.30	1.81
G Gly	38	4.97	8.60
H His	20	5.41	4.52
I Ile	18	4.12	4.07
K Lys	21	5.35	4.75
L Leu	47	10.75	10.63
M Met	16	4.16	3.62
N Asn	14	3.23	3.17
P Pro	21	4.22	4.75
Q Gln	24	6.12	5.43
R Arg	30	9.11	6.79
S Ser	27	4.95	6.11
T Thr	19	3.95	4.30
V Val	17	3.47	3.85
W Trp	4	1.42	0.90
Y Tyr	13	4.10	2.94
B Asx	38	8.79	8.60
Z Glx	64	16.38	14.48

Amino Acids	Codons Used
A Ala	GCA(8), GCC(10), GCG(3), GCT(17)
C Cys	TGC(1), TGT(2)
D Asp	GAC(10), GAT(14)
E Glu	GAA(23), GAG(17)
F Phe	TTC(4), TTT(4)
G Gly	GGA(9), GGC(8), GGG(6), GGT(15)
H His	CAC(9), CAT(11)
I Ile	ATA(1), ATC(9), ATT(8)
K Lys	AAA(6), AAG(15)
L Leu	CTA(8), CTC(5), CTG(11), CTT(11), TTA(3), TTG(9)
M Met	ATG(16)
N Asn	AAC(6), AAT(8)
P Pro	CCA(8), CCC(5), CCG(1), CCT(7)
Q Gln	CAA(6), CAG(18)
R Arg	AGA(12), AGG(2), CGA(4), CGC(4), CGG(4), CGT(4)
S Ser	AGC(8), AGT(2), TCA(4), TCC(4), TCG(1), TCT(8)
T Thr	ACA(7), ACC(4), ACG(2), ACT(6)
V Val	GTA(2), GTC(3), GTG(4), GTT(8)
W Trp	TGG(4)
Y Tyr	TAC(3), TAT(10)

7.4 Sequences

7.4.1 PCR amplification of S1 concatemers

Sequencing of S1 PCR A5 (A5)

Oligonucleotide S1

```
     5'..GCGAAGACGG ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA GGCAGAAGCG..5'
```

Sequence

pBS NCTCCGGCCGNCNANGGCGGCCGCGGGNAATTCGAT

G

```
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGTCAG
GCGAAGACGGAGCTGCCTGACAGGCAGCTCCGATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACC
ATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTG
GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACG
AGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGC
GCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCG
GGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA
ACGCAGGAAANAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCNCGTTGCTG
GNNTTTTTCCNNAGGCTCCNCCCCCCTGAN
```

Sequencing of S1 PCR B2 (B2)

Oligonucleotide S1

```
     5'..GCGAAGACGG ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA GGCAGAAGCG..5'
```

Sequence

pBS CCNTGGCGGCCGCGGGAATTCGATA

```
GCGAAGACGGAGCTGCCTGTGCACN
GCGNAGACGGAGCTGCCTGTGCACC
GCGAAGACGGAGCTGCCTGTGCNCC
GCGAAGACGGAGCTGCCTGTGCACC
GCGAAGACGGAGCNGCCTGTGCANC
GCGAAGANGGAGCTGNCTGTGCACC
GCGAAGACNGAGCTGCCTGTGCACC
GCNAAGACGGAGCNGCCTGTGCNCC
GCGAAGACGGAGCTGCCTGTGCACC
GCGAAGACGGAGCTGCCTGTGNACC
GCGAAGACGGAGCTGCCTGTGCACC
GCGAAGACGGAGCTGCCTGTGCACC
GCGAAGACGGAGCTGCCTGTGCNCC
GCGAAGACGGANCTGCCTGTGAC
NCGAANACGGAGCTGCCGAGGTTATGCTCGACAGGCAGCT
GCGAAGACGGAGCTGCCGAGGTNATGCTCGACAGGCAGCT
GCGAAGACGGNGCTGCCGAGGTTATGCTCGACNGGCAGCTGCGNANACGGAGCTGCCGAGGTTATGCT
CGACNGGCAGCTGCNNANATNTTTCTGCAGGNTCCGCCCTCCTGAGGAGGATCNGAAANATCGACGCN
NNGGTCAGAGGTGGCGAGGCCNNACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGG
```

Sequencing of S1 PCR C1 (C1)

Oligonucleotide S1

```
     5'..GCGAAGACGG ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT CCGTCTTCGC..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA GGCAGAAGCG..5'
```

Sequence

pBS NNCTCCGGCCGCCNANGGCGGCCGCGGGAAATTCGAT

GCGAAGACGGAGCTGCCTGTTACTC
GCGAAGACGGAGCTGCCTGTCAAAC
GCGAAGACGGAGCTGCCTGTCAAAC
GCGAAGACGGAGCTGCCTGTCAAAC
GCGAAGACGGAGCTGCCTGTCAAAC
GCGAAGACGGAGCTGCCTGTGTGT
GCGAAGACGGAGCTGCCTGTGCT
GCGAAGACGGAGCTGCCTGTGCT
GCGAAGACGGAGCTGCCTGTGCT
GCGAAGACGGAGCTGCCTGTGCT
GCGAAGACGGAGCTGCCTGTCAGGC
GCGAAGACGGAGCTGCCTGTGAG
GCGAAGACGGAGCTGCCTGTGAG
GCGAAGACGGAGCTGCCTAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCT
CCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGT
CATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATA
AAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGANAGAA
CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCNCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGANACCTGNCCNCCTTTNNNCCTTCGGGAAGCGNGGNNNTTTNNNANAGCTCACGCTNNAGGT
ATCNCAGTTCGGNNTAGGTCGTTCNNTCCAAGCTGGGCNN

7.4.2 PCR amplification of S2 concatemers

Sequencing of S2 PCR A2_1 (A2)

Oligonucleotide S2

```
     5′..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3′
     3′..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5′
```

Flipped over

```
     5′..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3′
     3′..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5′
```

Sequence

```
pGemT GNGCCGNNANGGCGGCCCGCGGGAATTCGATT
```

```
    AAGACCCACGTCGCGAGGTCGAGTGTAACAGGCAGCT
          ACGTCGCGAGAGAGCGGCACACAGGCAGCT
CGGAAGACCCACGTCGCGANNNNNNNNNCNACAGGCAGCTCCGTCTTCGC
          ACGTCGCGAGCGGCGGTAGAACAGGCAGCT
 GGAAGACCCACGTCGCGAGCATTAGTGTAACAGGCAGCTA
```

```
pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
           CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
           CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
           TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
           GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
           CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
           AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
           CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
           TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
           GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
           GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
           CANANATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
           GATACCAGGCGTTTCCNCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
           ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
           GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
           TTCGCTCCAAGCTGGGCTNNNTGCACGAACCCCCCGTTCAGCCCGACCGC
           TGCNCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAANANNCGA
           CTTATCGCCNCTGGCAGCAGCCNCTGGTAACAGGATTAGCNNNNCGAGGT
```

Sequencing of S2 PCR A2_2 (A3)

Oligonucleotide S2

```
     5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT NNNGGCGGCCCGCGGGNATTCGATT

```
   GAAGACGGAGCTGCCTGTNNNCCATGACGCGAGGT
GCGAAGACGGAGCTGCCTGTTGATCCAATACTCGCGACGTGGGTCT
              GCCTGTCAAACCGGATCTCGCGACGTGGGTCTTCCG
                                TCTCGCGACGTGGGTCT
          AGCTGCCTGTTCCACTCCAGCTCGCGACGTGGGTCTTCCGA
```

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTANAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCNCCCTGACGAGCATCA
CANAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAANANNCGA
CTTATCNCCACTGGCAGCAGCCACTGGTAACAGGATTANCNNN

Sequencing of S2 PCR B2_1 (A7)

Oligonucleotide S2

```
     5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT NAGNTCCGTCTTCGCGTGCTACAGGC

```
CGGANNNCCACGTCGCGAGTCTCCGCTGGACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                       GTGCTACAGGCAGCTCCGTCTTCGC
                                CACGTGGGTCNNCNGA
```

pGEmT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
NTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGNGAGCTAACTCACATTANTTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCNTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGGANAGG

Sequencing of S2 PCR B2_2 (B1)

Oligonucleotide S2

```
     5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT CTCCCGGGCNNNANGGCGGCCGCGGGAATTCGATT

```
GCGAAGACGGAGCTGCCTGTTTCACACTAGCTCGCGACGTGGGTCTT
          AGCTGCCTGTTACACACTTACTCGCGACGTGGGTCTTCCG
                                CGACGTGGGTCT
          AGCTGCCTGTCTCCCGTTTCCTCGCG
          ACGTCGCGAGCTCTAGGGCAACAGGC
          AGCTGCCTGTCAGTCCGCGAAGACGG
          AGCTGCCTGTAGACCAACACCTCGCGACGTGGGTCTTCCA
```

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CANAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGNAACTATCGTCTTGAGTCCAACCCGGNAANNMN

Sequencing of S2 PCR C3_3 (C5)

Oligonucleotide S2

```
     5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
     3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
     5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
     3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT NTCGATT

```
 CGGAANNNCACGTCGCGAGCCTGTGGCCAACAGGCAGCTCCGTCTTCGC
                     GCTGCACCACAGGCAGCTCCGTCTTCGC
               GTCGCGAGCAGGTGTTCCACAGGCAGCT
CGGAAGACCCACGTCGCGAGTCAAAGCTTAACAGGCAGCT
GCGAAGACGGAGCTGCCTGTATAACCGATTCTCGCGAGGTTCTGGACTACAGGC
          AGCTGCCTGTAAAACTAAACCTCGCGACGTGGGTCTTCCGA
```

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GANAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCNCCCTGACGAGCATCA
CAAANATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCNCCTGNAAGCTCCCTCGNGCGCTCTCCTGTTCCG
ACCCTGCCGCTTACCGGANACCTGGNCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGNNTNN

7.4.3 Reduction of circularization by PEG and blunt end ligation

Sequencing of S2 PEG A3_4 (E4)

Oligonucleotide S2

```
     5′..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3′
     3′..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5′
```

Flipped over

```
     5′..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3′
     3′..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5′
```

Sequence

```
pGemT      GCTCCCGGNCCGNNANGGCGGCCGCGGGAATTCGATT
```

```
GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAGCTATTCTCCTGCCT
CAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTAG
TAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTGA
TCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAACC
CTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTCT
GGAGTGCCATGTCGTTTGAGAAAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACTG
GAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGGA
GGCTGATGTCGA
```

```
pGemT      ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
           CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTANATAG
           CTTGGCGTAATCATGGTCATAGCTGTTTCCTGNGTGAAATTGTTNTCCNC
           TN
```

Sequencing of S2 PEG A3_11 (E11)

Oligonucleotide S2

```
5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT NNNNGGCGGCCCGCGGGAATTCGATT

GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAAGCTATTCTCCTGCC
TCAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTA
GTAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTG
ATCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAAC
CCTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTC
TGGAGTGCCATGTCGTTTGAGAAAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACT
GGAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGG
AGGCTGATGTCGA

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGNGTGAAATTGTTATCCGC
TCNCAATTCCNNN

Sequencing of S2 PEG A3_12 (E12)

Oligonucleotide S2

```
     5′..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3′
     3′..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5′
```

Flipped over

```
     5′..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3′
     3′..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5′
```

Sequence

pGemT NCTCCGGCCGNNANGGCGGCCCGCGGGNAATTCGATT

GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAAGCTATTCTCCTGCC
TCAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTA
GTAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTG
ATCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAAC
CCTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTC
TGGAGTGCCATGTCGTTTGAGAAAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACT
GGAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGG
AGGCTGATGTCGA

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTNNNNGAAATTGTTNTCCNC
TNNCAATTCC

Sequencing of S2 PEG B3_11 (F11)

Oligonucleotide S2

```
5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

```
pGemT      NCTCCCGGGCCGNCNANGGCGGCCGCGGGAATTCGATT
```

```
CGGAAGNCCC
ACGT CGCGAG GCGCGGTTGA ACAGGC
AGCT GCCTGT CGAACATTAC CC
ACGT CGCGGC CTGAGCAGC ACAGGC
AGCT GCCTGT CACACATAGA CCGTCGCGAGATTAGGCTATACAGGC
AGCT GCCTGT GACTCCCCAT CTCGCGACGT GA
```

```
pGemT      ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
           CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
           CTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
           TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
           GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
           CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
           AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
           CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
           TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
           GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTANAAAG
           GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
           CAAANATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
           GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
           ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
           GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
           TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
           TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAANACACGA
           CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCNNAACCNNGG
           TA
```

Sequencing of S2 PEG D3_1 (L1)

Oligonucleotide S2

```
5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT CNANNNGCGGCCGCGGGATTCGATT

GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAAGCTATTCTCCTGCC
TCAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTA
GTAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTG
ATCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAAC
CCTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTC
TGGAGTGCCATGTCGTTTGAGANAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACT
GGAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGG
AGGCTGATGTCGA

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGNCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGNGNGAAATTGTTNNCC

Sequencing of S2 PEG D3_5 (L5)

Oligonucleotide S2

```
     5′..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3′
     3′..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5′
```

Flipped over

```
     5′..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3′
     3′..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5′
```

Sequence

pGemT NGCCGCCNANGGCGGCCGCGGGAATTCGATT

GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAAGCTATTCTCCTGCC
TCAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTA
GTAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTG
ATCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAAC
CCTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTC
TGGAGTGCCATGTCGTTTGAGAAAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACT
GGAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGG
AGGCTGATGTCGA

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGTCATAGCTGTTTCCTGNGTGAAATTGTTATCCGC
TCACAATTCCNNNCAACATACGNNCCGGAAGCNTAAAGTNNNAAGCCTG

Sequencing of S2 PEG D3_6 (L6)

Oligonucleotide S2

```
5'..CGGAAGACCC ACGTCGCGAG NNNNNGNNNN ACAGGCAGCT CCGTCTTCGC..3'
3'..GCCTTCTGGG TGCAGCGCTC NNNNNCNNNN TGTCCGTCGA GGCAGAAGCG..5'
```

Flipped over

```
5'..GCGAAGACGG AGCTGCCTGT NNNNCNNNNN CTCGCGACGT GGGTCTTCCG..3'
3'..CGCTTCTGCC TCGACGGACA NNNNGNNNNN GAGCGCTGCA CCCAGAAGGC..5'
```

Sequence

pGemT NGCTCCCGGCCGCCNNTGGCGGCCGCGGGATTCGATT

GCCTCGAGTTGGTTTAAGCCTGAGGTATGTGGGCGTTGGCTGATCCGGGTAGGAGGAGCACTTGGGGA
AAGCTAAGATGGCTTCAGATTGATTGGTCGGAGCAGCCAGAATGGGGGAGGCGGGGGAAGGGAGTGCG
GCCTAGGAATCTGCGGTGTGGAGAGGGATAGACACTAAACCCTTTATTTTATTTTTATTTATTTATTT
TTAATTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGGTGGAGTGCAGTGGCACGATC
TTGGCTCACTGCAAGCTCTGCCTCCCGAGTTCAGGCCATTCTCCTGCCTCAGCCTTCCGAGTAGCAGG
GACTGCAGGCGCCCGCCACCGCGCCCGGCTAATTGCTCCACCTCCCGGGTTCAAGCTATTCTCCTGCC
TCAGCCATCCGAGTAGCTGGAACTACTGGCGCCCGCCACCACGCCCAGCTAATTTTTTCTATTTTTTA
GTAGAGACGAGGTTTCACCGTGTTAGCCAGGATGGTAGCCAGGATGGTCTCGTTCTCCTGACCCCGTG
ATCCGCCCGCCTCGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCCCCGCCCTTAAAC
CCTTTATTGATCAGTTATTGGCTCCAGAAAACAGGAGTTTTATTTAGGAGAAGGATACAGATACTGTC
TGGAGTGCCATGTCGTTTGAGAAAACTGGAAGGCTTGTGGTTTTCTAGCAGTCATCTTCAGCCTGACT
GGAATTTACTGCCGGTGAACCCCTAATAGAGAATATTATCCCAGGAAGGCAGCAGAACTCAAGTGTGG
AGGCTGATGTCGA

pGemT ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAG
CTTGGCGTAATCATGGNCATAGCTNNTTCCTGN

7.4.4 First SAGE-like experiment

Sequencing of S3 2µL_4 (A10)

Oligonucleotide S3

```
     5'..CGGAAGACGC TGCAGGCGAG NNTTTGNNNN ACACCGTGCA CGGTCTTCGC..3'
     3'..GCCTTCTGCG ACGTCCGCTC NNAAACNNNN TGTGGCACGT GCCAGAAGCG..5'
```

Flipped over

```
     5'..CGCTTCTGGC ACGTGCCACA NNNNGTTTNN GAGCGGACGT CGCAGAAGGC..3'
     3'..GCGAAGACCG TGCACGGTGT NNNNCAAANN CTCGCCTGCA GCGTCTTCCG..5'
```

Sequence

```
GCGGCCGCGGGAATTCGATTGGTGGTCATATGGCTCCCCCCAAGTCAGTCATCAGATGGTGCACCATA
TCCTCTCCAGAAGAGAAGAAATGCAACAACCTCAGGGACCTCACACAACAAGAGAGGATTTCATTGAC
CTGTGTGCAGAAAGCAACGTACCTTGACTGCATTAAAGCCATTGCGAATAATGAGGCAGATGCCATTA
GCTTGGACGGCGGTCAAGTGTTTGAGGCAGGCCTTGCCCCATACAAGCTGAAGCCCATTGCTGCTGAG
GTTTATGAACATACTGAAGGCTCCACAACCAGCTACTATGCTGTGGCCGTTGTGAAAAAAGGAACCGA
ATTCACCGTGAACGACTTGCAGGGCAAGACCTCCTGCCACACGGGGCTGGGCAGGTCTGCGGGCTGGA
ACATCCCCATTGGGACACTCATCCACCGGGGGGCCATTGAGTGGGAAGGCATAGAATCGGGCTCAGTG
GAACAAGCGGTGGCCAAGTTTTTCTCTGCCAGCTGCGTGCCTGGTGCCACCATCGAGCAAAAACTGTG
CCGTCAGTGCAAAGGAGACCCCAAAACCAAGTGTGCCCGCAATGCACCTTATTCTGGATATTCTGGAG
CTTTCCATTGTTTGAAAGATGGAAAAGGAGACGTGGCTTTTGTGAAGCACACAACTGTTAATGAAAAT
GCCCCGGATCAGAAGGATGAGTACGAGCTGCTGTGCCTGGATGGCAGCCGGCAGCCTGTGGACAACTA
CAAGACCTGTAACTGGGCCAGAGTGGCTGCTCACGCCGTCGTGGCTCGGGATGACAACAAGGTTGAAG
ATATCTGGGAGCTTCCTCTCAAAAGCACAGAGTGACTTTGGCGTGGACACCAAGAGTGACTTCCACCT
CTTTGGGCCACCTGGAAAGAAGGACCCGGTCCTCAAAGACTTGCTTTTCAAGACTCTGCCATAATGCT
GAAGCGCGTCCCCTCGCTGATGGAT
```

Sequencing of S3 3µL_5 (B5)

Oligonucleotide S3

5′..CGGAAGACGC TGCAGGCGAG NNTTTGNNNN ACACCGTGCA CGGTCTTCGC..3′
3′..GCCTTCTGCG ACGTCCGCTC NNAAACNNNN TGTGGCACGT GCCAGAAGCG..5′

Flipped over

5′..CGCTTCTGGC ACGTGCCACA NNNNGTTTNN GAGCGGACGT CGCAGAAGGC..3′
3′..GCGAAGACCG TGCACGGTGT NNNNCAAANN CTCGCCTGCA GCGTCTTCCG..5′

Sequence

pBS CGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCA

AGCACGGTGTGAACCAAAGCCTCGCC
TGCAGGCGAGCCTTTGTTG-ACACCG
TGCACGGTGTCACACAAAGCCTCGCC
TGCA

pBS TATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCC
CTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTT
TCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCG
GGAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACA
TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG
GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
GCCTTTCTCCCTTCGGGAAGCGTGGCGCNNCTCATAGCTCACGCTGTAGG
TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTG
AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA

Sequencing of S3 4µL_5 (B11)

Oligonucleotide S3

5′..CGGAAGACGC TGCAGGCGAG NNTTTGNNNN ACACCGTGCA CGGTCTTCGC..3′
3′..GCCTTCTGCG ACGTCCGCTC NNAAACNNNN TGTGGCACGT GCCAGAAGCG..5′

Flipped over

5′..CGCTTCTGGC ACGTGCCACA NNNNGTTTNN GAGCGGACGT CGCAGAAGGC..3′
3′..GCGAAGACCG TGCACGGTGT NNNNCAAANN CTCGCCTGCA GCGTCTTCCG..5′

Sequence

pBS AGTGGATCCCCCGGGCTGCAGGAATTCGATATCA

AGGGAACCTAACAGTAAGTTATTTAAAATTACAATATCTCCTACTCCCAAAAAAGGAATTATAATATA
ACTAAATAACTCACATTCTCTATGAACTGGGTATTAG

pBS TATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCC
CTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGGAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGC
TCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGT
GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT
TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAG
AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTT
TCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA

7.4.5 Second SAGE-like experiment

Sequencing of S4_<300_5 (A5)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 GGNNNCTCTAGAG

```
TCGAGGCATGATTCCAGACAAAGCCCCGTCGTAC
TCGAGTACGACGCGCGTTTGGGTTGGATCATGCC
TCGAGGCATGATGTTGCCCAAACATCCGTCGTAC
TCGAGGCATGATCGGAACCAAACTCGCGTCGTAC
TCGAGTACGACGGAATTTTGTCGGTGATCATGCC
TCGAGGCATGATCAAGGCCAAAAATCCGTCGTAC
TCGAGGCATGATTGCCCGCAAATTAGCGTCGTAC
```

```
pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
           TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
           GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
           ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
           GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
           GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
           CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
           AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
           AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
           CGCCCCCCTGACGAGCATCACANAAATCGACGCTCAAGTCAGAGGTGGCG
           AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
           TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
           TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
           TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
           CCNCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
           TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
           CAGGATTAGCAGAGCGAGGTATGTAGGCGNNGCTACAGAGTTCTTGAAGT
           GNNNGCCTAACTN
```

Sequencing of S4_<300_6 (A6)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 NNCTCTAGNG

```
TCGAGGCANGATCCACAACAAAACCCCGTCGTAC
TCGAGGCATGATGCCGCACAAAATCCCGTCGTAC
TCGAGTACGACGGCGATTTGAAGCCGATCATGCC
TCGAGTACGACGTTCGTTTGTGCACGATCATGCC
TCGAGGCATGATCGATATCAAAGAAGCGTCGTAC
TCGAGGCATGATCGCCGTCAAAAAAACGTCGTAC
```

```
pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
           TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
           GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
           ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
           GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
           GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
           CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
           AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
           AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
           CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG
           AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
           TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
           TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
           TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
           CCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
           TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
           CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGT
           GGTGGCCTAACTACGGCTANNCTAGAANAACAGTATTTGGNNTCTGCNCT
           CTGCTGAAGCCAGTTACCTTCGGAAANNNN
```

Sequencing of S4_300-500_1 (B3)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 NTAGAG

```
TCGAGTACGACGCAAGTTTGCACGGTATCATGCC
TCGAGGCATGATGGAACCCAAATGTGCGTCGTAC
TCGAGGCATGATCTGGAAAACATGCGTCGTAC
TCGAGTACGACGGGGCTTTGGTGAGCATCATGCC
TCGAGGCATGATATATGGCAAATCGACGTCCGTAC
TCGAGTACGACGGCTCTTTGGCGTAAATCATGCC
TCGAGGCATGATCTACCGGCAAAGCGACGTGGTAC
TCGAGTACGACGCGACTTTGCGGTTCATCATGCC
TCGAGTACGACGAATGTTTGGCAGCGATCATGCC
TCGAGGCATGATATATAGCAAATGCCCGTCGTAC
TCGAGTACGACGACTCTTTGTGATAGATCATGCC
TCGAGGCATGATACGAGCCAAAACGACGTCGTAC
TCGAGGCATGATGCCCCGCAAAGCTACGTCGTAC
```

```
pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
            TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
            GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
            ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
            GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
            GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
            CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
            AGAATCAGGGGATAACGCAGGANAGAACATGTGAGCAAAAGGCCAGCANA
            AGGCCAGGAACCGTAANAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
            CGCCCCNCTGACGAGCATCACAAANATCGACGCTCAAGTCAGAGGTGGCG
            AAACCCCGACAGGACTATAAAGATACCAGGCGTTTCNCCCTGGAAGCTCC
            CTCGTGCGCTCTCCTGTTCCGACNTTGCCGCTTANNGGATACCTGTN
```

Sequencing of S4_300-500_1 (C3)

Oligonucleotide S4

```
     5′..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3′
     3′..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5′
```

Reverse

```
     5′..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3′
     3′..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5′
```

Sequence

pUC19

```
    NNNGACGAAANGTTTGACGCGTATCATGCC
TCGAGGCATGATCGTTGTCAAAGAGGCGTCGTAC
TCGAGTACGACGGAGGTTTGCCGTAAATCATGCC
TCGAGTACGACGCACCTTTGCAGCCGATCATGCC
TCGAGGCATGATATATCCCAAACTGACGTCGTAC
TCGAGGCATGATTACACCCAAAGCGTCGTCGTAC
TCGAGGCATGATTCACGTCAAACCACCGNNGTAC
TCGAGGCATGATACTCGACAAACCTCCGTCGTAC
TCGAGGCATGATCGGGCGCAAACTGGCGTCGTAC
TCGAGGCATGATAGCAGTCAAACATACGTCGTAC
TCGAGTACGACGTTAGTTTGTTTGTCATCATGCC
TCGAGTACGACGTGNGTTTGCCATATATCATGCC
```

```
pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
           TNTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
           GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
           ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
           GCTGCATTAATGAATCGRCCAACGCGCGGGGAGAGGCGGTTTGCGTANNG
           GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
           CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
           AGAATCAGGGGATAACGCAGGANNGAACATGNGAGCAAAAGGCCAGCNAA
           AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
           CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGNGGNGGCG
           AAACCCGACAGGACTATNAAGANNNCAGGCGTTNNNCCCTGGAAGCCNCC
           CCGGGCG
```

Sequencing of S4_>500_4 (D2)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 CCGGGNNNCTCTAGAG

```
TCGAGGGCATGATGACGTGCAAATGTCCGTCGTAC
TCGAGGCATGATGTAGAACAAACAAACGTCGTAC
TCGAGTACGACGTAGTTTTGGCCTGGATCATGCC
TCGAGTACGACGGGTATTTGCTGCTGATCATGCC
TCGAGTACGACGGCAGTTTGGTGATCATCATGCC
TCGAGGCATGATTCCGCCCAAAGCGGCGTCGTAC
TCGAGTACGACGTGGTTTTGTTCGTGATCATGCC
TCGAGTACGACGGAGATTTGCAAACAATCATGCC
TCGAGGCATGATACGCTACAAACAACCGTCGTAC
TCGAGTACGACGTTCATTTGGGCCCGATCATGCC
TCGAGGCATGATGCACCCCAAATGTCCGTCGTAC
TCGAGGCATGATTGCCTACAAACACACGTCGTAC
TCGAGTACGACGGACCTTTGCAACGGATCATGCC
TCGAGGCATGATGATAGCCAAACTGTCGTCGTAC
TCGAGGCATGATCGCACCCAAACTCGCGTCGTAC
TCGAGTACGACGGGGTTTTGTGGTTAATCGTGCC
```

pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGANACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCANAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGANAGAACATGTGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTANNNAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAANATCGACGCTCAAGTCAGAAGGTGGC
GAAANCCCGACNNGGACNNNTAAAGATACCAGGNNGTTTNCCCCCTGGAAAN

Sequencing of S4_>500_4 (D4)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 NANAG

```
TCGAGTACGANNTAAGTTTGGGGGNNTCATGCC
TCGATTACGACGAGGTTTTGTCCAGTATCATGCC
TCGAGTACGACGCGGGTTTGACGCAGATCATGCC
TCGAGTACGACGACATTTTGTGCATCATCATGCC
TCGAGGCATGATGAATGCCAAAAACGCGTCGTAC
TCGAGGCATGATCCGTTGCAAATTTCCGTCGTAC
TCGAGTACGACGGGGCTTTGGCGCACGATCATGCC
TCGAGTACGACGTGGATTTGTATGTAATCATGCC
TCGAGGCATGATATCGTTCAAATCGACGTCGTAC
TCGAGGCATGATGATGCCCAAATTTGCGTCGTAC
TCGAGTACGACGTGTATTTGACAGGGATCATGCC
TCGAGGCATGATTAAAGCCAACTACCGTCGTAC
TCGAGGCATGATGCTTCACAAATACCCGTCGTAC
TCGAGTACGACGAGGATTTGCCTTAGATCATGCC
TCGAGGCATGATGGCGAACAAATTTGCGTCGTAC
TCGAGGCATGATGTGAGTCAAAATTGCGTCGTAC
TCGAGTACGACGTGTCTTTGGGCGGGATCATGCC
```

```
pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
           TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
           GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
           ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
           GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
           GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
           CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
           AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
           AGGCCAGGAACCGTANAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
           CGCCCCCCTGACGAGCATCACAAAAANCGACGCTCAAGTCAGAAGGNGGC
           GAAACNNNGACAGGACTNNNAAAGATACCANNNNGTTTCCCNCTGGNAAG
           CTCCCTNGNG
```

Sequencing of S4_>500_6 (D4)

Oligonucleotide S4

```
     5'..AGCTGCTCGA GTACGACGNN NNTTTGNNNN NNATCATGCC TCGAGTTCGC..3'
     3'..TCGACGAGCT CATGCTGCNN NNAAACNNNN NNTAGTACGG AGCTCAAGCG..5'
```

Reverse

```
     5'..GCGAACTCGA GGCATGATNN NNNNCAAANN NNCGTCGTAC TCGAGCAGCT..3'
     3'..CGCTTGAGCT CCGTACTANN NNNNGTTTNN NNGCAGCATG AGCTCGTCGA..5'
```

Sequence

pUC19 GGGNNNCTCTAGAG

```
TCGAGTACGACGTATGTTTGTGAAGNNNATGCC
TCGAGGCATGAGCTGCGCAAAGTCCGTCGTAC
TCGAGGCATGATGCTCGACAAAATTGCGTCGTAC
TCGAGTACGACGTCTTTTTGGCGTACATCATGCC
TCGAGGCATGATTAAATGCAAAGAGCCGTCGTAC
TCGAGGCATGATATATAACAAACGGGCGTCGTAC
TCGAGGCATGNNNCTACCCAAAAGACCGTCGTAC
TCGAGGCATGATGCGGGCCAAATACACGTCGTAC
TCGAGTACGACGCCTTTTTGACGTTCATCATGCC
TCGAGGCATGATCACCCACAAAGTCACGTCGTAC
TCGAGTACGACGAACGTTTGGCTAGTATCATGCC
TCGAGTACGACGGCATTTTGCTTCTCATCATGCC
TCGAGTACGACGCAGGTTTGGGTCATATCATGCC
TCGAGGCATGATGGTGTACAAAGTAGCGTCGTAC
TCGAGTACGACGGGCGTTTGGGATTCATCATGCC
TCGAGGCATGATTCAAGACAAAACATCGTCGTAC
TCGAGTACGACGGCCTTTTGGGCGTGATCATGCC
TCGAGTACGACGGCGTTTTGCTCGGAATCATGCC
TCGAGGCATGATGTGCTACAAAACTACGTCGTAC
```

pUC19 TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCNNAAGGCCAGCAAA
AGGCCAGGAACCGTANNNAGGCCGCGTTGCTGGCGTTTTTCCNNNGGCNC
CNGCNCCCCTGACNAGCATCNNAANAATCG

7.5 Multimerization raw data

7.5.1 Obtained sequences

	#	*	
<300bp Fragments	1	1	TCGAGGCATGATTCCAGACAAAGCCCCGTCGTAC
	2	2	TCGAGTACGACGCGCGTTTGGGTTGGATCATGCC
	3	3	TCGAGGCATGATGTTGCCCAAACATCCGTCGTAC
	4	4	TCGAGGCATGATCGGAACCAAACTCGCGTCGTAC
	5	5	TCGAGTACGACGGAATTTTGTCGGTGATCATGCC
	6	6	TCGAGGCATGATCAAGGCCAAAAATCCGTCGTAC
	7	7	TCGAGGCATGATTGCCCGCAAATTAGCGTCGTAC
	8	8	TCGAGGCANGATCCACAACAAAACCCCGTCGTAC
	9	9	TCGAGGCATGATGCCGCACAAAATCCCGTCGTAC
	10	10	TCGAGTACGACGGCGATTTGAAGCCGATCATGCC
	11	11	TCGAGTACGACGTTCGTTTGTGCACGATCATGCC
	12	12	TCGAGGCATGATCGATATCAAAGAAGCGTCGTAC
	13	13	TCGAGGCATGATCGCCGTCAAAAAAACGTCGTAC
300-500bp Fragments	14	14	TCGAGTACGACGCAAGTTTGCACGGTATCATGCC
	15	15	TCGAGGCATGATGGAACCCAAATGTGCGTCGTAC
	16		TCGAGGCATGATCTGGAAAACATGCGTCGTAC
	17	16	TCGAGTACGACGGGGCTTTGGTGAGCATCATGCC
	18	17	TCGAGGCATGATATATGGCAAATCGACGTCCGTAC
	19	18	TCGAGTACGACGGCTCTTTGGCGTAAATCATGCC
	20		TCGAGGCATGATCTACCGGCAAAGCGACGTGGTAC
	21	19	TCGAGTACGACGCGACTTTGCGGTTCATCATGCC
	22	20	TCGAGTACGACGAATGTTTGGCAGCGATCATGCC
	23	21	TCGAGGCATGATATATAGCAAATGCCCGTCGTAC
	24	22	TCGAGTACGACGACTCTTTGTGATAGATCATGCC
	25	23	TCGAGGCATGATACGAGCCAAAACGACGTCGTAC
	26	24	TCGAGGCATGATGCCCCGCAAAGCTACGTCGTAC
	27	25	NNNGACGAAANGTTTGACGCGTATCATGCC
	28	26	TCGAGGCATGATCGTTGTCAAAGAGGCGTCGTAC
	29	27	TCGAGTACGACGGAGGTTTGCCGTAAATCATGCC
	30	28	TCGAGTACGACGCACCTTTGCAGCCGATCATGCC
	31	29	TCGAGGCATGATATATCCCAAACTGACGTCGTAC
	32	30	TCGAGGCATGATTACACCCAAAGCGTCGTCGTAC
	33	31	TCGAGGCATGATTCACGTCAAACCACCGNNGTAC
	34	32	TCGAGGCATGATACTCGACAAACCTCCGTCGTAC
	35	33	TCGAGGCATGATCGGGCGCAAACTGGCGTCGTAC
	36	34	TCGAGGCATGATAGCAGTCAAACATACGTCGTAC
	37	35	TCGAGTACGACGTTAGTTTGTTTGTCATCATGCC
	38	36	TCGAGTACGACGTGNGTTTGCCATATATCATGCC
>500bp Fragments	39	37	TCGAGGGCATGATGACGTGCAAATGTCCGTCGTAC
	40	38	TCGAGGCATGATGTAGAACAAACAAACGTCGTAC
	41	39	TCGAGTACGACGTAGTTTTGGCCTGGATCATGCC
	42	40	TCGAGTACGACGGGTATTTGCTGCTGATCATGCC

43	41	TCGAGTACGACGGCAGTTTGGTGATCATCATGCC
44	42	TCGAGGCATGATTCCGCCCAAAGCGGCGTCGTAC
45	43	TCGAGTACGACGTGGTTTTGTTCGTGATCATGCC
46	44	TCGAGTACGACGGAGATTTGCAAACAATCATGCC
47	45	TCGAGGCATGATACGCTACAAACAACCGTCGTAC
48	46	TCGAGTACGACGTTCATTTGGGCCCGATCATGCC
49	47	TCGAGGCATGATGCACCCCAAATGTCCGTCGTAC
50	48	TCGAGGCATGATTGCCTACAAACACACGTCGTAC
51	49	TCGAGTACGACGGACCTTTGCAACGGATCATGCC
52	50	TCGAGGCATGATGATAGCCAAACTGTCGTCGTAC
53	51	TCGAGGCATGATCGCACCCAAACTCGCGTCGTAC
54	52	TCGAGTACGACGGGGTTTTGTGGTTAATCGTGCC
55		TCGAGTACGANNTAAGTTTGGGGGNNTCATGCC
56	53	TCGATTACGACGAGGTTTTGTCCAGTATCATGCC
57	54	TCGAGTACGACGCGGGTTTGACGCAGATCATGCC
58	55	TCGAGTACGACGACATTTTGTGCATCATCATGCC
59	56	TCGAGGCATGATGAATGCCAAAAACGCGTCGTAC
60	57	TCGAGGCATGATCCGTTGCAAATTTCCGTCGTAC
61	58	TCGAGTACGACGGGGCTTTGGCGCACGATCATGCC
62	59	TCGAGTACGACGTGGATTTGTATGTAATCATGCC
63	60	TCGAGGCATGATATCGTTCAAATCGACGTCGTAC
64	61	TCGAGGCATGATGATGCCCAAATTTGCGTCGTAC
65	62	TCGAGTACGACGTGTATTTGACAGGGATCATGCC
66		TCGAGGCATGATTAAAGCCAACTACCGTCGTAC
67	63	TCGAGGCATGATGCTTCACAAATACCCGTCGTAC
68	64	TCGAGTACGACGAGGATTTGCCTTAGATCATGCC
69	65	TCGAGGCATGATGGCGAACAAATTTGCGTCGTAC
70	66	TCGAGGCATGATGTGAGTCAAAATTGCGTCGTAC
71	67	TCGAGTACGACGTGTCTTTGGGCGGGATCATGCC
72		TCGAGTACGACGTATGTTTGTGAAGNNNATGCC
73	68	TCGAGGCATGAGCTGCGCAAAGTCCGTCGTAC
74	69	TCGAGGCATGATGCTCGACAAAATTGCGTCGTAC
75	70	TCGAGTACGACGTCTTTTTGGCGTACATCATGCC
76	71	TCGAGGCATGATTAAATGCAAAGAGCCGTCGTAC
77	72	TCGAGGCATGATATATAACAAACGGGCGTCGTAC
78		TCGAGGCATGNNNCTACCCAAAAGACCGTCGTAC
79	72	TCGAGGCATGATGCGGGCCAAATACACGTCGTAC
80	73	TCGAGTACGACGCCTTTTTGACGTTCATCATGCC
81	74	TCGAGGCATGATCACCCACAAAGTCACGTCGTAC
82	75	TCGAGTACGACGAACGTTTGGCTAGTATCATGCC
83	76	TCGAGTACGACGGCATTTTGCTTCTCATCATGCC
84	77	TCGAGTACGACGCAGGTTTGGGTCATATCATGCC
85	78	TCGAGGCATGATGGTGTACAAAGTAGCGTCGTAC
86	79	TCGAGTACGACGGGCGTTTGGGATTCATCATGCC
87	80	TCGAGGCATGATTCAAGACAAAACATCGTCGTAC

88	81	TCGAGTACGACGGCCTTTTGGGCGTGATCATGCC
89	82	TCGAGTACGACGGCGTTTTGCTCGGAATCATGCC
90	83	TCGAGGCATGATGTGCTACAAAACTACGTCGTAC

… number of fragment

* …number of useful fragment

7.5.2 Raw data of the sequence logo

LogoData
First column is position number, counting from zero
Subsequent columns are raw symbol counts
Entropy is mean entropy measured in nats.
Low and High are the 95% confidence limits.
Weight is the fraction of non-gap symbols in the column.
#

#	A	C	G	T	Entropy	Low	High	Weight
1	0	0	82	1	1.2395	1.0954	1.3835	1
2	0	0	0	83	1.291	1.1675	1.4146	1
3	83	0	0	0	1.291	1.1675	1.4146	1
4	0	83	0	0	1.291	1.1675	1.4146	1
5	0	0	82	0	1.2901	1.1654	1.4148	0.988
6	82	0	0	0	1.2901	1.1654	1.4148	0.988
7	0	83	0	0	1.291	1.1675	1.4146	1
8	0	0	83	0	1.291	1.1675	1.4146	1
9	9	22	29	22	0.0843	0.0274	0.1724	0.988
10	24	21	27	11	0.0618	0.0157	0.1389	1
11	22	14	25	21	0.0372	0.0055	0.0986	0.988
12	20	17	25	21	0.0265	0.0026	0.0774	1
13	0	0	0	83	1.291	1.1675	1.4146	1
14	0	0	0	83	1.291	1.1675	1.4146	1
15	0	0	0	83	1.291	1.1675	1.4146	1
16	0	0	83	0	1.291	1.1675	1.4146	1
17	11	21	27	24	0.0618	0.0157	0.1389	1
18	14	28	26	15	0.062	0.0144	0.1435	1
19	17	23	25	18	0.0299	0.0033	0.0848	1
20	16	16	24	27	0.0437	0.0072	0.1123	1
21	18	18	26	21	0.0289	0.003	0.0834	1
22	14	27	27	15	0.0616	0.0143	0.1427	1
23	82	0	0	0	1.2901	1.1654	1.4148	0.988
24	0	0	0	83	1.291	1.1675	1.4146	1
25	0	83	0	0	1.291	1.1675	1.4146	1
26	81	0	1	0	1.2381	1.0928	1.3834	0.988
27	0	0	0	83	1.291	1.1675	1.4146	1
28	0	0	83	0	1.291	1.1675	1.4146	1
29	0	83	0	0	1.291	1.1675	1.4146	1
30	0	83	0	0	1.291	1.1675	1.4146	1

End LogoData