

INAUGRALDISSERTATION  
zur Erlangung der Doktorwürde der  
Naturwissenschaftlich-Mathematischen Gesamtfakultät der  
Ruprecht-Karls-Universität Heidelberg

**Vorgelegt von:**

Diplom Chemiker  
Christian Krauß

aus Heidelberg

**eingereicht am:** 12.12.2008  
**Tag der mündlichen Prüfung:** 20.03.2009



Thema

MICROARRAY-TECHNOLOGIE ZUR BESTIMMUNG DER DNA-BINDUNGSSPEZIFITÄTEN  
KREBSRELEVANTER TRANSKRIPTIONSFAKTOREN

**Gutachter:** Prof. Dr. Jürgen Wolfrum  
Prof. Dr. Roland Krämer

Die vorliegende Arbeit wurde in der Zeit vom 01.11.2005 - 15.12.2008 am Deutschen Krebsforschungszentrum in Heidelberg in der Abteilung Funktionelle Genomanalyse unter der Leitung von Dr. Jörg Hoheisel angefertigt.



Meinen Freunden, meiner Familie  
und besonders Julia

## Publikationen

*Nature Methods, 2008, submitted*

"A HIGH THROUGHPUT METHOD FOR THE DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C.Krau, M. Fugazza, Marita Schrenk, N. Tonisson, J. Taipale, J. Hoheisel

*Journal of Biological Chemistry, 2008, in preparation*

DETERMINATION OF CANCER ASSOCIATED TRANSCRIPTION FACTOR BINDING SPECIFICITIES USING MICROARRAY TECHNOLOGY

C.Krau, Marita Schrenk, J. Taipale, J. Hoheisel

## Vortrge

*10. Statusseminar fr Chiptechnologien, Dechema Frankfurt, 2008*

"DNA-CHIP BASED DETERMINATION OF BINDING SPECIFICITIES OF TRANSCRIPTION FACTORS LINKED TO COLORECTAL CANCER"

C.Krau, M. Fugazza, N. Tonisson, J. Taipale, J. Hoheisel

## Posterprsentationen

*DKFZ PHD-Retreat 2006*

"TECHNOLOGY AND METHODS FOR DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C. Krau, N. Tnisson, M. Schrenk, J. Taipale, J.D. Hoheisel

*DKFZ FSB Retreat 2006*

"TECHNOLOGY AND METHODS FOR DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C. Krau, M. Fugazza, N. Tnisson, M. Schrenk, J. Taipale, J.D. Hoheisel

*2nd HBGS Student Council Symposium, Helsinki Biomedical Graduate School, Transcription, 20 07*

"DNA-CHIP BASED DETERMINATION OF BINDING SPECIFICITIES OF TRANSCRIPTION FACTORS LINKED TO COLORECTAL CANCER"

C.Krau, M. Fugazza, N. Tonisson, J. Taipale, J. Hoheisel

*DKFZ PHD Retreat 2007*

"TECHNOLOGY AND METHODS FOR DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C.Krauß, M. Fugazza, N. Tônisson, M. Schrenk, J. Taipale, J.D. Hoheisel

*9. Statusseminar Chiptechnologien, Dechema Frankfurt, 2007*

"DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C.Krauß, N. Tonisson, J. Taipale, J. Hoheisel

*DKFZ FSB Retreat 2007*

"TECHNOLOGY AND METHODS FOR DETERMINATION OF TRANSCRIPTION FACTOR BINDING SPECIFICITIES"

C. Krauß, M. Fugazza, N. Tônisson, M. Schrenk, J. Taipale, J.D. Hoheisel

## Zusammenfassung

Die vorliegende Arbeit war ein Teil des “*Regulatory Genomics*” EU-Projektes “*Advanced genomics instruments, technology and methods for determination of transcription factor binding specificities; applications for identification of genes predisposing to colorectal cancer*”. Ziel des Projektes ist die Entwicklung verschiedenster neuartiger Methoden zur Bestimmung der Bindungsspezifitäten von Transkriptionsfaktoren, die im Verdacht stehen Krebs zu verursachen.

Der Mechanismus, mit dem onkogene Transkriptionsfaktoren das Wachstum regulieren, ist bis heute noch nicht ganz verstanden. Nur wenige Ausnahmen, wie z. B. der Östrogen- oder der Androgenrezeptor wurden bis jetzt charakterisiert. Es sind zur Zeit etwa 40 Transkriptionsfaktoren bekannt, bei denen eine Veränderung der Genregulation mit Krebs in Verbindung gebracht werden kann. Darüber hinaus ist der Einfluss von Mutationen in der DNA Bindesequenz, welche das Binden von Transkriptionsfaktoren beeinflusst, noch nicht vollständig geklärt. Die Information über die spezifischen Bindestellen dieser Transkriptionsfaktoren sind meist unvollständig oder nur die Sequenz mit der höchsten Affinität ist bekannt. Viele dieser so genannten Konsensus Sequenzen werden als einzelne Sequenz angegeben. Diese Angabe führt jedoch oft zu falschen Ergebnissen. Besitzt eine Position der Bindeseite zu 75% ein Adenin und zu 25% Thymin, gehen bei der Angabe nur einer Sequenz Information verloren. Bezogen auf alle Positionen dieser Bindesequenz kann dies zu falschen Bindesequenzen führen. Desweiteren können theoretische Modelle diese Sequenzen nur unvollständig beschreiben oder vorhersagen. Von den im menschlichen Genom geschätzten 2000 Transkriptionsfaktoren sind bislang nur ein Bruchteil der Bindestellen bekannt, gut charakterisiert und in öffentlichen Datenbanken zugänglich.

Je nach Fragestellung werden zur Zeit verschiedene Methoden zur Untersuchung von Protein-DNA-Wechselwirkungen angewendet. Zu nennen wären hier vor allem SELEX (Systematic Evolution of Ligands by Exponential Enrichment), EMSA (Electrophoretic Mobility Shift Assay), ChIP-chip (chromatin immunoprecipitation-on-chip), DamID (DNA adenine methyltransferase identification) oder auch der von einem Kooperationspartner entwickelte Luziferase Assay. Mit diesen Methoden kann nach verschiedenen Bindesequenzen gesucht werden. Aber auch die Plasmonenresonanzspektroskopie oder die QCMD-Technologie können für diese Untersuchungen genutzt werden. Mit ihnen kann die Kinetik dieser Bindung bestimmt werden. Einen Nachteil haben jedoch alle diese Methoden: sie sind in Bezug auf die zu untersuchenden Sequenzen limitiert. Die Entwicklung einer Methode, die auf DNA Microarrays basiert, würde die Kosten senken und gleichzeitig die Anzahl der zu untersuchenden Proben erhöhen. Dadurch sollte es möglich sein, in naher Zukunft genomweit nach bestimmten Transkriptionsfaktorbindestellen zu suchen.

---

Die Microarray-Technologie hat sich aus dem in den 70er Jahren entwickelten Southern-Blot heraus zu einer Hochdurchsatz-Methode mit vielseitigen Einsatzgebieten entwickelt. Bei beiden Methoden ist die Grundidee die gleiche. Bestimmte Moleküle (DNA, DNA-Analoga, Proteine, Peptide, Antikörper) werden auf speziell modifizierten Oberflächen in einer definierten Anordnung fixiert und anschließend mit Proben hybridisiert. Durch hybridisieren der Proben ist es möglich, aus komplexen Gemischen das komplementäre Zielmoleküle nachzuweisen. febit Biomed GmbH stellt mit der Geniom®-Technologie eine Microarray-Plattform zur Verfügung, die sehr flexibel und vielfältig einsetzbar ist. Es ist mit diesem Gerät möglich, sowohl die Synthese der Oligonukleotide auf dem Chip, die Protein-Inkubation als auch die Detektion der Fluoreszenzsignale durchzuführen.

Ziel der hier vorliegenden Arbeit war Entwicklung und Etablierung einer auf Microarray basierenden Methode zur Bestimmung von Protein-DNA-Wechselwirkungen im *high throughput* Format. Hierzu sollte die Geniom-Plattform in Zusammenarbeit mit der febit biomed GmbH weiterentwickelt und für diese Experimente etabliert werden. Zunächst wurde anhand des Fusionsproteins MBP-Gli2 die Entwicklung und Etablierung der Methode durchgeführt. In zweiten Teil wurde diese Methode genutzt, um die Bindungseigenschaften von 10 Transkriptionsfaktoren zu untersuchen. Diese wurden allesamt in *E. coli* - Bakterien als Fusionsproteine exprimiert. Da der vollständige Transkriptionsfaktor aufgrund seiner Größe nicht als Fusionsprotein in *E. coli* exprimiert werden kann, wurde für die Generierung nur die für die DNA Bindung notwendige Bindedomäne verwendet. Die Fusionsproteine sind alle nach folgendem Schema aufgebaut: MBP-3xMyc-TF. Das Maltose bindende Protein (MBP) wurde für die Aufreinigung der Fusionsproteine benötigt. Dadurch wurde zusätzlich die Löslichkeit und Stabilität des Fusionsproteins erhöht, da in vielen Fällen die Bindedomäne des Transkriptionsfaktors alleine schlecht löslich und instabil ist. Mit Hilfe der 3-fach Myc-Markierung (3xMyc) konnte unter der Verwendung eines Anti-Myc-Antikörpers und des entsprechend fluoreszenz-markierten 2. Antikörpers das Fusionsprotein detektiert werden.

Im ersten Teil der Arbeit stand die Entwicklung der Methode anhand des Transkriptionsfaktors Gli2 sowie ein Vergleich der Ergebnisse mit anderen Methoden im Vordergrund. Die Wahl von Gli2 zur Etablierung der Methode hatte vor allem zwei Gründe. Zum einen gehört Gli2 zu den bisher etwa 40 mit Krebs in Verbindung gebrachten Transkriptionsfaktoren. Zum anderen hatte bereits ein Kooperationspartner ergänzende Experimente gemacht und stellte auch einen Teil der verwendeten Plasmide der Fusionsproteine zur Verfügung. Als erstes musste der Aufbau der Methode geplant werden. Da die zu untersuchenden Transkriptionsfaktoren nur an doppelsträngige, etwa 9 bp lange DNA-Sequenzen binden, musste dies bei dem Sondendesign berücksichtigt werden. Die doppelsträngige DNA konnte durch die Synthese von selbstkomplementären Haarnadelstrukturen erreicht werden. Der hierfür verwendete Chip ist mit einer speziellen Oberfläche beschichtet und in 8 Kanäle unterteilt, die jeweils mit verschiedenen Proben hybridisiert werden können. In jedem dieser Kanäle lassen sich durch Licht-aktivierte Synthese auf 6776 Spots beliebige Sequenzen synthetisieren. Für den ganzen Chip ergibt dies die Möglichkeit, 54.208 verschiedene DNA-Sequenzen herzustellen. Da dieser Chip nicht über ausreichend Kapazität für alle vorstellbaren Nukleotidsequenzen eines 9 mers ( $4^9 = 262144$ )

---

verfügt, wurde zuerst nur ein Teil der Sequenzen untersucht. Ausgehend von der Konsensus Bindesequenz der Proteine galt es alle möglichen Permutationen von einem beziehungsweise zwei Nukleotiden auf ihre Bindungseigenschaften zu untersuchen.

Für die ersten Microarray Versuche wurde das etwas ältere Geniom 1X System benutzt. Dieses System ist auf 4 Arrays mit weniger Spots limitiert, bietet jedoch für die Entwicklung der Methode entscheidende Vorteile. Der wohl Wichtigste war die manuelle Unterbrechung der Synthese. Somit konnten die Chemikalien nachgefüllt und Oligonukleotide bis zu einer Länge von 48 Basenpaaren synthetisiert werden. Mit dem neueren Geniom 2X System konnten zu diesem Zeitpunkt nur Oligonukleotide mit bis zu 30 Basen synthetisiert werden. Ein weiterer Vorteil war die manuelle Bedienung der Waschprogramme sowie die manuelle Einstellung der Flussrate während der Waschprozesse. Bei dem Geniom 2X System werden diese Einstellungen in vorgefertigten Profilen zur Verfügung gestellt.

Mit den Versuchen am Geniom 1X konnte gezeigt werden, dass es mit dieser Methode möglich ist Protein-DNA-Wechselwirkungen zu untersuchen. Die Synthese eines 48 bp langen Oligonukleotids konnte durch das Stoppen der Synthese und das Auffüllen der Chemikalien optimiert werden. Das in *E. coli* - Bakterien exprimierte Fusionsprotein MBP-Gli2 wurde aufgereinigt und an die verschiedenen doppelsträngigen DNA Sequenzen gebunden. Dies konnte durch die Markierung mit einem biotinylierten Anti-cMyc-Antikörper und eines Streptavidin-Phycoerythrin-Konjugates nachgewiesen werden. Bei der Auswertung der Ergebnisse wurden die Signalintensitäten aller möglichen Sequenzen gegen die Konsensus Sequenz normalisiert. Verwendet wurden dabei mehrere Replikate innerhalb eines Arrays. Die Waschprozesse während der verschiedenen Inkubationen wurden ebenfalls optimiert. In Kooperation mit febit wurde danach ein Hard- und Softwareupdate an dem Geniom 2x System durchgeführt. Es war nun auch hier möglich Oligonukleotide bis zu einer Länge von 60 bp zu synthetisieren. Mit den Erkenntnissen aus den Versuchen mit dem 1X System konnten die notwendigen Inkubations- und Waschprofile auf dieses System übertragen werden. Die Reproduzierbarkeit dieser Experimente war jedoch anfänglich nicht immer gewährleistet. Dies konnte auch durch die Optimierung der Proteinaufreinigung nicht verbessert werden. Es wurde deshalb ein Glasobjektträger mit verschiedenen Konzentrationen der Konsensus Sequenz von Gli2 (GACCACCA) hergestellt und mit dem Fusionsprotein MBP-3xMyc-Gli2 inkubiert. Hiermit sollte der Einfluss der synthetisierten DNA hinsichtlich der Konzentration auf dem Chip überprüft werden. Da dies bei den kommerziell erhältlichen Chips der Firma febit nicht ohne Weiteres möglich war, wurde dieser Versuchsaufbau gewählt. Es konnte gezeigt werden, dass die Konzentration der DNA auf der Oberfläche einen erheblichen Einfluss auf das Binden des Transkriptionsfaktors hat. Ist die DNA-Konzentration zu hoch, kann aufgrund sterischer Wechselwirkungen keine Protein-DNA-Bindung erfolgen. In Zusammenarbeit mit der febit biomed GmbH wurde ein Chip mit nur 10% der herkömmlichen DNA-Konzentration hergestellt und getestet. Erreicht wurde dies durch eine Modifizierung der Aminosilan-Oberfläche. Mit diesem Chip konnten die Probleme der mangelnden Reproduzierbarkeit der Experimente behoben werden. Der Erfolg bei den folgenden Versuchen lässt sich im Nachhinein recht einfach erklären. Bei zu hoher DNA-Dichte ist nicht gewährleistet, dass die Einzelstränge die zum Gelingen der Experimente

---

benötigte Haarnadelstruktur korrekt ausbilden. Wegen der räumlichen Nähe der Moleküle zueinander ist die Bildung von netzwerkartigen Strukturen vorstellbar. Dass diese dann eine Bindung des Proteins verhindern, ist verständlich. Auch sterische Wechselwirkungen des Fusionsproteins können trotz einer korrekt ausgebildeten Haarnadel ein Binden verhindern. Ein Gelingen der Experimente mit dem alten Chip ist beispielsweise durch eine nicht optimal verlaufende Synthese zu erklären.

Um den Einfluss des 40 kDa großen MBP-Tags in Bezug auf das Bindeverhalten des Fusionsproteins zu untersuchen, wurde ein His-Gli2 Fusionsprotein mit einem sechsfachen His-Tag anstatt eines MBP-Tags kloniert. Das His-Gli2 Fusionsprotein wurde in *E. coli* - Bakterien exprimiert und über einen Ni<sup>2+</sup>-NTA-Komplex aufgereinigt. Die Microarray-Experimente wurden unter den selben Bedingungen mit identischen DNA-Sonden durchgeführt. Nach Normalisierung gegen die Konsensus-Sequenz ergaben sich nahezu identische Werte für beide Tags. Somit konnte ein Einfluss des Tags auf das Bindeverhalten ausgeschlossen werden.

Um die aus den Microarray Experimenten gewonnenen Daten zu validieren und zusätzlich das Bindungsverhalten der in *E. coli* - Bakterien exprimierten Fusionsproteine genauer zu charakterisieren, wurden zwei weitere Methoden zur Bestimmung der Protein-DNA-Wechselwirkung angewandt. Die Experimente mit der Quarzkristallmikrowaage unterstützten die oben gezeigten Ergebnisse. Mit Hilfe der Plasmonenresonanzspektroskopie konnte die Gleichgewichtskonstanten der Fusionsproteine bestimmt werden. Diese lagen mit 10<sup>-8</sup> M in dem aus der Literatur bekannten Bereich (10<sup>-9</sup> M). Der um den Faktor 10 abweichende Wert dieser Ergebnisse kann durch die Verwendung des Fusionsproteins anstatt eines vollständigen Transkriptionsfaktors und die dadurch fehlenden kooperativen Effekte erklärt werden. Durch die Verwendung dieser Methoden konnte die Funktionalität des Proteins sowie eine intakte Protein-DNA-Wechselwirkung auch für die Microarray-Experimente angenommen werden.

Im zweiten Teil der Arbeit wurden mit der auf Microarray-Technologie basierenden Methode die Bindungsspezifitäten von 10 krebisrelevanten Transkriptionsfaktoren bestimmt. Von einem Kooperationspartner wurden freundlicherweise die für die Proteinexpression notwendigen Plasmide zur Verfügung gestellt. Diese Fusionsproteine konnten hinsichtlich ihrer Bindungsspezifitäten erfolgreich untersucht werden. Ausgehend von der aus der Literatur bekannten Konsensus Sequenz wurden alle möglichen Ein- und Zweibasenaarmutationen untersucht.

Die in dieser Arbeit erzielten Ergebnisse zeigen, dass die hier entwickelte Methode für die Messung von Protein-DNA-Wechselwirkungen im *“high throughput”* Format angewendet werden kann. Dies wurde als *“proof of principle”* anhand des Transkriptionsfaktors Gli2 mit zwei verschiedenen Tags sowie für 10 weitere Transkriptionsfaktoren gezeigt. Es konnten hierbei innerhalb der Bindemotive Gemeinsamkeiten aufgezeigt werden. Bei den C2H2-Zink-Fingerproteinen Gli2 und WT1 ist hauptsächlich die Base Cytosin für die DNA Bindung verantwortlich. Auch ein Vergleich der Sequenzlogos mit anderen aus der Literatur bekannten Zinkfingern wie Zif268 zeigt eindeutige Übereinstimmungen. Auffällig war weiterhin das Vorkommen von Kernsequenzen mit einer Länge von 2-3 Basenpaaren in der Mitte der Bindesequenz, die maßgeblich für

---

die Bindung verantwortlich sind. Die Ergebnisse zeigen weiterhin, dass bei den meisten Transkriptionsfaktoren der Austausch dieser Basen hinsichtlich der Proteinbindung als besonders kritisch anzusehen ist. Dies lässt sich durch strukturelle Veränderung in der DNA erklären. Die Bindung des Proteins an die Bindedomäne funktioniert ähnlich dem Schlüssel-Schloss-Prinzip. Für die Spezifität und Stabilität der Bindung sind Wasserstoffbrücken, Ionenbindungen sowie hydrophobe Wechselwirkungen verantwortlich. Wird dieses perfekt aufeinander abgestimmte System verändert und die Struktur des DNA bindenden Motivs passt nicht mehr exakt zu der Form der DNA, ist die Affinität reduziert. Leicht vorstellbar ist dies beispielsweise bei einem Austausch eines Pyrimidins gegen ein Purin. Auch der gegenteilige Fall ist denkbar. Es ist also durchaus erklärbar, dass auch stärker bindende Sequenzen als die Konsensus-Sequenz gefunden wurden. Die häufigste natürlich vorkommende Bindesequenz muss nicht unbedingt die Stärkste sein. Wichtig ist, dass sie schnell und effektiv regulierbar ist.

Mutationen in der Bindesequenz von Transkriptionsfaktoren können somit weitreichende Folgen haben. Bis jetzt sind jedoch nur wenige Mutation in diesen Regionen bekannt, die Krankheiten wie Krebs verursachen können. Mit den aus dieser Arbeit gewonnenen Daten können Vorhersagen getroffen werden, wie sich verschiedene Mutationen in den Promoterregionen hinsichtlich der Proteinbindung und somit auch auf die Expression auswirken können.

Es existieren zur Zeit mehrere Methoden, mit denen die Bindungsspezifitäten von Transkriptionsfaktoren untersucht werden können. Diese basieren zum größten Teil auf Gel-Elektrophorese oder Microarray-Technologie. Aber auch Methoden wie Systematische Evolution von Liganden durch exponentielle Anreicherung (SELEX) können hierfür verwendet werden. Bei der SELEX-Methode werden Molekülbibliotheken mit dem zu untersuchenden Zielmolekül (z.B. ein Transkriptionsfaktor) inkubiert. Nach mehreren Wasch- und Selektionsschritten wird der mit dem Zielmolekül interagierende Ligand (z.B. DNA) mittels Sequenzierung nachgewiesen. Roulet et al. (2002) sequenzierten mehr als 1000 Liganden, um das Bindungsverhalten von CTF/NFI mittels SELEX zu untersuchen. Den Kosten von 10 € pro Sequenzierung stehen zusätzlich viele Wochen Laborarbeit gegenüber.

*DNAse footprinting* und *electrophoretic mobility shift assay* (EMSA) sind Methoden, bei denen die Protein-DNA-Komplexe mittels Gelelektrophorese aufgetrennt und analysiert werden. Hierbei ist die Anzahl der zu untersuchenden Proben sehr begrenzt. Da die Interaktionen nicht direkt gemessen werden können, ist die Sensitivität dieser Methoden sehr beschränkt.

Auch die Microarray-Technologie wird für diese Untersuchungen verwendet. Man kann hier zwischen der Probengewinnung *in vivo* und *in vitro* unterscheiden. Bei *DNA adenine methyltransferase identification*-Technologie (DamID) wird das zu untersuchende Protein als Fusionsprotein mit Dam *in vivo* überexprimiert. Nach Binden des Transkriptionsfaktors an seine Zielsequenz werden in der näheren Umgebung Adenine, in GATC-Sequenzen, durch Dam *in vivo* methyliert. Anschließend können die durch Methyl-spezifischen Restriktionsverdau erhaltenen DNA-Fragmente mit Kontrollen verglichen werden. Dies geschieht nach Amplifizierung und Markierung der DNA ebenfalls durch Microarrayanalyse. Etabliert wurde diese Methode für die Spezies *Drosophila* und *Arabidopsis*. Bei der CHip-Chip-Technologie werden

---

die Protein-DNA Komplexe *in vivo* mit einem Crosslinker wie Formaldehyd inkubiert. Nach dem Lysieren der Zellen können diese Komplexe aufgereinigt und untersucht werden. Der größte Vorteil der ChIP-Chip-Methode ist die Analyse posttranslationalen Modifikationen und kooperativer Effekte anderer Bindungspartner.

Ein Nachteil der oben genannten Methoden ist die Länge der untersuchten DNA. Sie ist größer als die eigentliche Bindesequenz und limitiert das Auflösungsvermögen dieser Techniken.

Mit der Plasmonenresonanzspektroskopie können Protein-DNA-Wechselwirkungen zur Zeit am genauesten untersucht werden. Es ist möglich, sowohl die Bindungskonstanten als auch die Bindungsenergien der Protein-DNA Bindung zu bestimmen. Der größte Nachteil dieser Methode ist die limitierte Anzahl der Proben.

Eine recht neue Methode zur Bestimmung von Protein-DNA-Interaktionen sind die sogenannten Protein bindenden Microarrays (PBMs). Hierbei werden in *E. coli* - Bakterien oder Insektenzellen exprimierten Fusionsproteine auf DNA Microarrays untersucht. Mukherjee et al. (2004) entwickelten eine Methode basierend auf gespotteten DNA-Chips. Im Gegensatz dazu verwenden Bulyk (2006) und Warren et al. (2006) synthetisierte oder kommerziell erhältliche Microarrays, die alle möglichen Permutationen einer 10 bp langen Sequenz beinhalten. Diese Methoden sind sehr effektiv, wenn keine Informationen über die Bindeseiten vorliegen. Wie bei allen PBMs können leider nicht die Einflüsse von Kofaktoren untersucht werden und auch eine Beeinträchtigung der Bindungseigenschaften durch das Fusionsprotein kann nicht ausgeschlossen werden.

Im Gegensatz zu den kommerziell erhältlichen Chips erlaubt das Geniom als Microarray-Plattform durch seine Flexibilität eine freie Synthese von bis zu 125.000 DNA-Sequenzen pro Chip. Da auch die Inkubation, die Waschschrte sowie die Detektion in einem Gerät erfolgen, ist man nicht auf vorgefertigte Microarrays oder externe Scanner angewiesen. Durch die Verwendung eines bestimmten Tags benötigt man auch keine spezifischen Antikörper gegen das zu untersuchende Fusionsprotein. Da der Chip in 8 Arrays aufgeteilt ist, können mehrere Proben parallel inkubiert und detektiert werden. Es können einzelnen Mutationen als auch eine große Anzahl verschiedener Sequenzen analysiert werden. Sofern die DNA-Bindedomäne nicht größer als 9 bp ist und dadurch die zur Verfügung stehende Anzahl an Spots nicht überschritten wird, ist es theoretisch möglich alle Permutationen dieser Sequenz zu untersuchen und zu analysieren. So kann man auch schwache Bindesequenzen erkennen und erhält auch Informationen über Sequenzen, die ein Binden des Proteins nicht erlauben.

## Abstract

This thesis was done as part of the EU-funded project Regulatory Genomics: *Advanced genomics instruments, technology and methods for determination of transcription factor (TF) binding specificities; applications for identification of genes predisposing to colorectal cancer*. The aim of this project is to develop novel genomics tools and methods for determination of transcription factor binding specificity. These tools will be used for the identification of regulatory SNPs that predispose to colorectal cancer, and for characterisation of downstream target genes that are common to multiple oncogenic TFs.

Determination of the sequence of the human genome, and knowledge of the genetic code through which mRNA is translated have allowed rapid progress in the identification of mammalian proteins. However, less is known about the molecular mechanisms that control expression of human genes, and about the variations in gene expression that underlie many pathological states, including cancer. This is caused in part by lack of information about the 'second genetic code', which is the binding specificities of transcription factors (TFs). Deciphering this regulatory code is critical for cancer research, as little is known about the mechanisms by which the known genetic defects induce the transcriptional programs that control cell proliferation, survival and angiogenesis. In addition, changes in binding of transcription factors caused by single nucleotide polymorphisms (SNPs) are likely to be a major factor in many quantitative trait conditions, including familial predisposition to cancer. The information about transcription factor binding site specificity is often incomplete (i.e. only the site with maximal affinity is known), or biased by the prediction methods used (such as alignment of multiple potential binding sites). In addition, the most extensive publicly available collection of relatively high quality transcription factor binding sites (JASPAR2;) contains information about binding specificities of only 108 of more than 2000 predicted TFs in the human genome.

Currently, multiple methods exist for determination of transcription factor binding specificity. The most important methods are SELEX (Systematic Evolution of Ligands by Exponential Enrichment), EMSA (Electrophoretic Mobility Shift Assay), ChIP-chip (chromatin immunoprecipitation-on-chip), DamID (DNA adenine methyltransferase identification) and a luciferase assay developed by one of our cooperation partners. However, also surface plasmon resonance spectroscopy or Quartz Crystal Microbalance with Dissipation Monitoring (QCMD) technology can be used to detect these interactions. All these methods have a common problem: they are limited in the number of samples. For this purpose, a DNA-chip based method would be more cost-effective and by increasing the number of samples analysed in parallel. The future aim of determination of the binding specificities of all human transcription factors could therefore be achieved by using microarray technology.

---

Microarray technology is a high-throughput method with several applications. The major idea is always the same. Specific molecules like DNA, DNA-analogues, peptides, proteins or antibodies are captured on modified surfaces in a defined order. These probes are incubated with target molecules from complex mixtures for the identification of binding partners. The Geniom platform of the company febit is a single benchtop instrument that integrates the complete process of oligonucleotide microarray synthesis, addition of labelled sample, hybridisation, and detection using a CCD camera. The Geniom technology utilises microfluidics and a unique microchannel structure. A digital projector inside the instrument performs light-activated in situ synthesis of high-density oligonucleotide arrays.

The aim of this thesis was the development of a microarray-based method to determine protein-DNA interactions in a high-throughput manner. For this purpose, the Geniom platform was established and refined for transcription factor binding studies in a cooperation with febit biomed GmbH. Initially the development and establishment of these method was performed using the MBP-3Myc-Gli2 fusion protein. In the second part of this thesis, the assay was used to determine the binding specificities of ten transcription factors, which are associated with cancer. These were expressed as fusion proteins in *E. coli* bacteria. For generating the fusion proteins only the binding site of the transcription factor was used, because the full-length protein is too large for the expression in *E. coli*. The structure of the fusion protein is as follows: MBP-3xMyc-TF. The MBP tag leads to a very stable and soluble fusion protein. The MBP-purification is easy to handle and provides good protein yields. The 3xMyc-tag was also used as the common epitope for antibody detection. For increased sensitivity of the binding assays, the antibodies were secondarily labelled with a streptavidin-phycoerythrin conjugate.

In the first part of the thesis, I focused on the development of the method and a proof of principle using the Gli2 fusion protein. These results were also validated using SPR- and QCMD-technology. The choice of using the Gli2 transcription factor had two reasons. Gli2 is one of the 40 known transcription factors, which are associated with cancer. Second, a cooperation partner made supplementary experiments and provided all constructs. Since the transcription factors bind only to double-stranded DNA sequences of approximately 9 bp in length, a special design was used for synthesis. Double-stranded DNA configuration could be attained by direct synthesis of a single-stranded oligonucleotide forming a stem-loop structure. A Geniom chip consists of 54,208 spots divided into 8 arrays with 6,776 spots each. For development, we initially used relatively small numbers of oligonucleotides corresponding to consensus binding sequences of Gli2 and all possible single and double nucleotide substitutions to this sequence.

For the first experiments, the old version of the Geniom platform was used. This Geniom 1X system was limited to 4 arrays with fewer spots. However, for the development of the assay, it had some big advantages. The most important key aspect is the discontinuation of synthesis. By doing this it was possible to fill-up the chemicals and to synthesise oligonucleotides with a length of 48 bp. At that moment, only 30 bp nucleotides could be synthesised with the second generation Geniom platform. Another advantage was the manual control of the washing

---

programs and the fluidic system. This is not possible anymore with the Geniom 2X system, because all the washing profiles are pre-built and tested on the part of febit.

With the experiments at the Geniom 1X system it could be shown that the method can be used to determine protein-DNA interactions. With the discontinuation of the synthesis, it was possible to optimise the synthesis of a 48 bp oligonucleotide. The fusion protein expressed in *E. coli* was purified and bound to the dsDNA sequences on the chip. This could be verified with a biotinylated Anti-cMyc-antibody and Streptavidin-R-Phycoerythrin conjugate. For the validation, all signal intensities were normalised against the consensus sequence in replicates. Also the washing procedure was optimised. In cooperation with febit, a hard- and software-update was realized. It made possible to use the Geniom 2X platform for longer synthesis. With the knowledge of the previous experiments, it was possible to transfer the incubation and washing protocols to this system. Still, the reproducibility of the assays were not guaranteed. This problem could not be solved by optimising the protein expression and purification. To analyse the influence of the DNA density on the chip in respect of the protein binding, a normal glass slide was spotted with the consensus sequence (GACCACCCA) in different concentrations. This slide was incubated with the MBP-3xMyc-Gli2 fusion protein. This kind of assay was used, because the DNA-density on the febit chip can not be changed easily. It could be shown that the DNA concentration on chip has a huge influence on the binding of the transcription factor. Is it too high, no protein-DNA binding can occur because of steric interactions. In cooperation with febit, a chip with only 10% DNA concentration could be constructed and tested. This could be achieved by modifying the amino-silane surface. The problem of the reproducibility could be solved with this chip. The failure of the preceding experiments could now be explained very easily. If the DNA density is too high, an accurate hairpin structure cannot form. Because of the spatial closeness of the molecules, a formation of a network is possible instead. This network decreases the binding specificity of the protein. The successful experiments with the old chip can be explained by a reduced synthesis efficiency.

To study the influence of the large MBP tag with respect to the binding specificity, a His-tagged fusion protein was cloned. The His-Gli2 protein was also expressed in *E. coli* and purified by a Ni-NTA-complex. The microarray experiments were performed under the same conditions with the same nucleotides. After normalisation against the consensus sequence, nearly identical results were achieved. Consequently, an influence of the tag can be excluded.

To validate the microarray experiments, the binding behaviour of the fusion proteins were analysed via Surface Plasmon Resonance Spectroscopy (SPR) and Quartz Crystal Microbalance with Dissipation Monitoring (QCMD) technology. The experiments with QCMD confirmed the results. With the assistance of the SPR, the equilibrium constants of the protein-DNA complex could be determined. Compared to literature ( $10^{-9}$  M) the results ( $10^{-8}$  M) were in the same range. The difference by a factor 10 can be explained by the usage of the fusion protein instead of the whole transcription factor. There are no cooperative effects by using only the binding site of the protein. These results demonstrate that the binding specificity and the functionality of the fusion protein is hereby assigned.

---

In the second part of the thesis, this method was used to determine the binding specificities of ten cancer relevant transcription factors. The plasmids for the expression of the fusion proteins were kindly provided by our cooperation partner Jussi Taipale. Based on the consensus sequence, all possible single and double nucleotide substitutions to this sequence were successfully analysed.

The results obtained in this work reveal that the established method can be used routinely for the measurement of protein-DNA interactions in a high-throughput format. This was shown as a proof of principle for the MBP-3xMyc-Gli2 fusion protein and for ten other transcription factors. Additionally, we found similar core sequences within the binding motives. For the C2H2 zinc finger proteins Gli2 and WT1 basically the base cytosine is responsible for the DNA binding. We have also found core sequences with a length of 2-3 bases in the middle of the binding sequence, which were very important for the binding. Furthermore, the results show that a substitution of one of these sequences is very critical with respect to binding. The binding of the protein to the dsDNA operates similar to the “Lock-and-Key Principle”. Hydrogen-bridges, ion-binding and hydrophobic interactions are responsible for the specificity and stability of the binding. This is a very accurate system, perfectly coordinated. In this system, slightly changed, the affinity can be reduced. But not a decreasing of the affinity is possible, also an increase was observed. The most abundant natural sequence is not necessarily the strongest one. It is important that this sequence is adjustable fast and effective.

Mutations in this binding sequence might have far-reaching consequences. Until now only a few mutations are known which might be cancer-relevant. With the data obtained in this work, a prediction can be made how different mutations in the promoter regions can affect protein binding and gene expression.

Currently, multiple methods exist for determination of transcription factor binding specificity. Largely these methods base on gel electrophoresis or microarray technology. But also methods like SELEX (Systematic Evolution of Ligands by Exponential Enrichment) can be used for this purpose. SELEX also referred to as *in vitro* selection or *in vitro* evolution, is a combinatorial technique in molecular biology for producing oligonucleotides of either single-stranded DNA that specifically bind to a target ligand (e.g. protein). Still, Roulet et al. 2002 performed more than 1000 individual sequencing runs for determination of the binding specificity of CTF/NFI, which at the cost of €10/run, would equal €10 000 per factor in sequencing costs alone, with estimated four to eight weeks of labour in generating the required proteins and clones.

DNase footprinting and electrophoretic shift assay are methods, which use gel electrophoresis to determine protein-DNA interaction. The inherent problem of these methods is that it is not based on actual measurement of the variable of interest but instead infers the affinity of a transcription factor to all possible sequences from statistical treatment of predicted biologically relevant binding sites. In addition to the potential of circular reasoning, the binding sites are affected by investigator bias, and in many cases, the number of binding sites aligned is low, generating high average statistical error (150%).

---

But also the microarray technology can be used for this investigations. There are two possibilities to extract samples: *in vivo* and *in vitro*. DamID (DNA adenine methyltransferase identification) is a molecular biology method used to map the binding sites of DNA-binding proteins in eukaryotes (*drosophila*, *arabidopsis*). DamID identifies binding sites by expressing the proposed DNA-binding protein as a fusion protein with DNA methyltransferase. Binding of the protein of interest to DNA localises the methyltransferase in the region of the binding site. Adenosine methylation does not occur naturally in eukaryotes and therefore adenine methylation in any region can be concluded to have been caused by the fusion protein, implying the region is located near a binding site. Amid is an alternate method to ChIP-on-chip. ChIP is the technique of precipitating a protein antigen out of solution using an antibody that specifically binds to that particular protein. Later this can be analysed on chip. With this method it is possible to determine posttranslational modifications and cooperative effects.

But all this methods have one major disadvantage. The examined binding site is larger than the real binding site. This restricts the resolution of the methods.

With the plasmon resonance spectroscopy it is possible measure the protein-DNA binding most accurately. The equilibrium constants as well as the binding energies can be determined. But this method is limited by the number of samples.

Recently, higher-throughput microarray-based methods for determining transcription factor-binding specificity have been developed. One method is based on the analysis of a labelled transcription factor that binds to long genomic DNA sequences immobilised on a chip (Mukherjee et al. (2004)). Subsequently, sequences that correspond to positive positions on the array are analysed using statistical methods to find conserved motifs. Alignment of the identified sequences is used to deduce an estimate of binding specificity. As the number of identified sequences is large, the method circumvents most of the pitfalls connected to alignment-based methods. Very recently, a related but more direct method was described, in which the DNA-binding specificity is determined by allowing a labelled protein to bind to a DNA microarray that contains every permutation of up to ten positions of a duplex DNA (Bulyk (2006), Warren et al. (2006)). This method is effective for solving the binding specificity when no initial information about the binding site is available and gives complete information about all possible variations of the binding sequence.

In contrast to the commercial chips the Geniom platform allows a free synthesis up to 125,000 sequences. Because the hybridisation as well as the detection takes place in the Geniom, no external scanner is necessary. The microstructured reaction carrier is subdivided into 8 individual channel-like reaction chambers, which are addressed in parallel during synthesis, but may be used as individual arrays for the following hybridisation experiments. It is possible to detect single permutations, but also a large number of sequences can be screened. For a 99 bp long binding sequence, it is possible to screen for all possible permutations. Therefore, one can screen also for weak interactions.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretischer Teil</b>	<b>11</b>
2.1	Aufbau der DNA . . . . .	11
2.2	Transkription und Translation . . . . .	13
2.3	Proteine und ihre Struktur . . . . .	14
2.3.1	Primärstruktur . . . . .	14
2.3.2	Sekundärstruktur . . . . .	14
2.3.3	Tertiärstruktur . . . . .	15
2.3.4	Quartärstruktur . . . . .	15
2.4	Transkriptionsfaktoren und die Strukturelemente DNA bindender Proteine . .	15
2.4.1	Helix-Turn-Helix (HTH) . . . . .	16
2.4.2	Proteine mit Homöodomänen . . . . .	17
2.4.3	Zinkfinger-Motive . . . . .	17
2.4.4	Leucin-Zipper-Motiv . . . . .	17
2.4.5	Helix-Loop-Helix (HLH) . . . . .	18
2.5	Expression von Fusionsproteinen und Aufreinigung über Affinitätschromatographie	18
2.5.1	Proteinexpression in <i>E. coli</i> - Bakterien . . . . .	18
2.5.2	Fusionsproteine . . . . .	19
2.5.3	Proteinaufreinigung . . . . .	19
2.5.3.1	Affinitätschromatographische Aufreinigung über eine Amylose-Harz-Matrix . . . . .	19
2.5.3.2	Affinitätschromatographische Aufreinigung über Ni-NTA Agarose . . . . .	20
2.5.4	Antikörper-Detektion . . . . .	21
2.5.5	Streptavidin-R-Phycoerythrin . . . . .	21
2.6	MALDI-TOF Analyse . . . . .	22
2.7	DNA-Microarrays mit dem Geniom System . . . . .	23
2.7.1	Licht-aktivierte DNA-Synthese . . . . .	24
2.8	Plasmonenresonanzspektroskopie . . . . .	26
2.8.1	Messprinzip . . . . .	27
2.8.2	Kinetik und Auswertung . . . . .	28
2.8.3	Datenauswertung mit der BIAevaluation®-Software . . . . .	29
2.9	Quarzmikrowaage - QCMD . . . . .	29
2.9.1	Massenbestimmung nach Sauerbrey . . . . .	30
2.9.2	Der Dissipations-Faktor . . . . .	30

<b>3</b>	<b>Material und Methoden</b>	<b>31</b>
3.1	Material . . . . .	31
3.1.1	Laborausstattung . . . . .	31
3.1.2	Verbrauchsmaterial . . . . .	32
3.1.3	Enzyme & Proteasen . . . . .	35
3.1.4	Kits . . . . .	35
3.1.5	Puffer und Medien . . . . .	35
3.1.6	Oligonukleotide, Primer und Vektoren . . . . .	37
3.1.7	<i>E. coli</i> - Bakterienstämme . . . . .	38
3.1.8	Antikörper . . . . .	39
3.2	Mikrobiologische Methoden . . . . .	39
3.2.1	Herstellung kompetenter <i>E. coli</i> - Bakterien . . . . .	39
3.2.2	Plasmid-DNA Transformation in <i>E. coli</i> - Bakterien . . . . .	40
3.2.3	Kultivieren von <i>E. coli</i> - Bakterien . . . . .	40
3.2.4	Dauerkulturen von <i>E. coli</i> - Bakterien . . . . .	40
3.3	Molekularbiologische Methoden . . . . .	41
3.3.1	Plasmidisolierung aus <i>E. coli</i> - Bakterien . . . . .	41
3.3.2	Agarose-Gelelektrophorese . . . . .	41
3.3.3	Restriktionsverdau von DNA . . . . .	41
3.3.4	DNA Extraktion aus Agarosegelen . . . . .	42
3.3.5	Polymerasekettenreaktion (PCR) . . . . .	42
3.3.6	Auswahl der Primer . . . . .	42
3.3.7	Klonierung von DNA Fragmenten in Plasmide . . . . .	43
3.3.7.1	Ligation . . . . .	43
3.3.7.2	Blau-Weiß-Selektion . . . . .	43
3.3.8	<i>Colonie-PCR</i> . . . . .	44
3.3.9	DNA Sequenzierung . . . . .	44
3.4	Biochemische Methoden . . . . .	45
3.4.1	Ermittlung der optimalen Expressionsbedingungen . . . . .	45
3.4.1.1	Proteinexpression in <i>E. coli</i> - Bakterien mit dem pMAL-c2X Vektor . . . . .	45
3.4.1.2	Proteinexpression in <i>E. coli</i> - Bakterien mit dem pMAL-p2X Vektor . . . . .	45
3.4.2	Zellaufschluss . . . . .	45
3.4.2.1	French Press . . . . .	45
3.4.2.2	Ultraschall . . . . .	46
3.4.3	Proteinaufreinigung . . . . .	46
3.4.3.1	Affinitätschromatographische Aufreinigung über eine Amylose - Harz - Matrix . . . . .	46
3.4.3.2	Affinitätschromatographische Aufreinigung über Ni - NTA - Agarose . . . . .	47
3.4.3.3	Aufreinigung mittels <i>Fast protein liquid chromatography</i> (FPLC) . . . . .	47

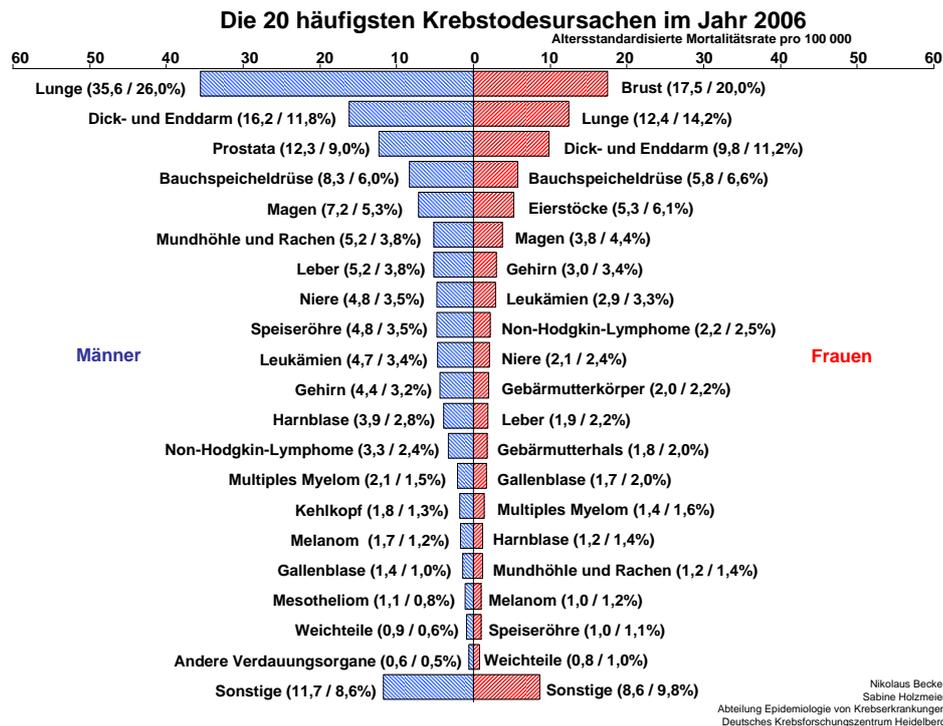
3.4.4	Denaturierende Natriumdodecylsulfat Polyacrylamid Gelelektrophorese	48
3.4.5	Western Blot . . . . .	49
3.4.6	Aufkonzentrieren von Proteinlösungen und Pufferaustausch . . . . .	50
3.4.7	Quantifizierung und Identifizierung von Proteinen . . . . .	51
3.4.7.1	Proteinmengenbestimmung nach Bradford . . . . .	51
3.4.7.2	Proteinmengenbestimmung mittels Bicinchoninsäure (BCA) . . . . .	52
3.4.7.3	Protein-/DNA-Mengenbestimmung mittels Absorption . . . . .	53
3.4.8	Proteinanalyse mittels MALDI-TOF-Massenspektrometrie . . . . .	54
<b>4</b>	<b>Experimente und Ergebnisse</b>	<b>55</b>
4.1	Versuchsaufbau . . . . .	55
4.1.1	<i>E. coli</i> - Bakterien und Plasmide . . . . .	55
4.1.2	Fusionsproteine . . . . .	55
4.1.3	Array-Design . . . . .	56
4.1.4	DNA-Synthese . . . . .	57
4.1.5	Protein-Inkubation . . . . .	57
4.1.6	Detektion . . . . .	58
4.1.7	Auswertung . . . . .	58
4.1.7.1	Gewichtete Positionsmatrix . . . . .	58
4.1.7.2	Sequenzmusteranalyse . . . . .	58
4.2	Etablierung der Methode anhand des Transkriptionsfaktors Gli2 (Gli-Kruppel family member 2) . . . . .	59
4.2.1	Nachweis von Gli2 mittels Sequenzierung . . . . .	59
4.2.2	Klonierung der Gli2- Bindedomäne . . . . .	60
4.2.3	Optimierung der Proteinexpression . . . . .	60
4.2.4	Proteinaufreinigung . . . . .	61
4.2.4.1	Proteinaufreinigung des MBP-Fusionsproteins über eine Amylose-Harz-Matrix . . . . .	61
4.2.4.2	Abspaltung des MBP-Tags . . . . .	62
4.2.4.3	MALDI-TOF Analyse . . . . .	63
4.2.4.4	Proteinaufreinigung des His-Fusionsproteins über Ni-NTA-Agarose . . . . .	65
4.2.4.5	Proteinaufreinigung des His-Fusionsproteins mittels <i>Fast Protein Liquid Chromatography</i> (FPLC) . . . . .	66
4.2.5	Bestimmung der Protein-DNA Wechselwirkungen . . . . .	67
4.2.5.1	Etablierung und Optimierung der Microarrayexperimente . . . . .	69
4.2.5.2	<i>Primer-Extension</i> . . . . .	70
4.2.5.3	Optimierung der Proteinmenge . . . . .	71
4.2.5.4	Optimierung der DNA-Konzentration . . . . .	72
4.2.5.5	Kontrolle der Haarnadelstruktur . . . . .	75
4.2.5.6	Optimierung des Arraydesigns . . . . .	76
4.2.5.7	Analyse der Microarraydaten . . . . .	76

4.2.5.8	Vergleich verschiedener Farbstoffe . . . . .	82
4.2.5.9	Gewichtete Positionsmatrix . . . . .	83
4.2.5.10	Sequenzmuster-Analyse . . . . .	84
4.2.5.11	Reproduzierbarkeit . . . . .	85
4.2.6	Dynamische Detektion . . . . .	86
4.2.7	Universeller 9mer Chip . . . . .	88
4.3	Evaluierung der Ergebnisse mit anderen Methoden . . . . .	91
4.3.1	Plasmonenresonanzspektroskopie (SPR) . . . . .	91
4.3.1.1	Immobilisierung der DNA . . . . .	91
4.3.1.2	Analyse . . . . .	91
4.3.1.3	Auswertung . . . . .	92
4.3.2	Quartz-Kristall-Mikrowaage . . . . .	94
4.3.2.1	Herstellung der Streptavidinoberfläche . . . . .	95
4.3.2.2	Immobilisierung der DNA . . . . .	95
4.3.2.3	Analyse . . . . .	95
4.3.2.4	Auswertung . . . . .	95
4.4	Expression der Fusionsproteine in unterschiedlichen <i>E. coli</i> - Stämmen. . . . .	99
4.5	Bestimmung der Bindungsspezifitäten krebsrelevanter Transkriptionsfaktoren	102
4.5.1	PBX1 (Pre-B-Cell Leukemia Transcription Factor 1) . . . . .	102
4.5.2	RBPJK (Recombination Signal Binding Protein for Immunoglobulin Kappa J region) . . . . .	104
4.5.3	ATF1 (Cyclic AMP-dependent transcription factor 1) . . . . .	106
4.5.4	TCF4 (Transcription Factor-4; Immunoglobulin Transcription Factor-2 (ITF2)) . . . . .	108
4.5.5	WT1 (Wilms tumor protein 1) . . . . .	110
4.5.6	CBFB (Core-Binding Factor, beta subunit) . . . . .	112
4.5.7	SMAD4 (Mothers Against Decapentaplegic Homolog 4) . . . . .	114
4.5.8	TFE3 (Transcription Factor E3) . . . . .	116
4.5.9	p53 (Cellular tumor antigen p53) . . . . .	118
<b>5</b>	<b>Diskussion und Ausblick</b>	<b>121</b>
<b>6</b>	<b>Literaturverzeichnis</b>	<b>130</b>
<b>7</b>	<b>Abkürzungsverzeichnis</b>	<b>136</b>
<b>8</b>	<b>Abbildungsverzeichnis</b>	<b>139</b>
<b>9</b>	<b>Anhang</b>	<b>144</b>

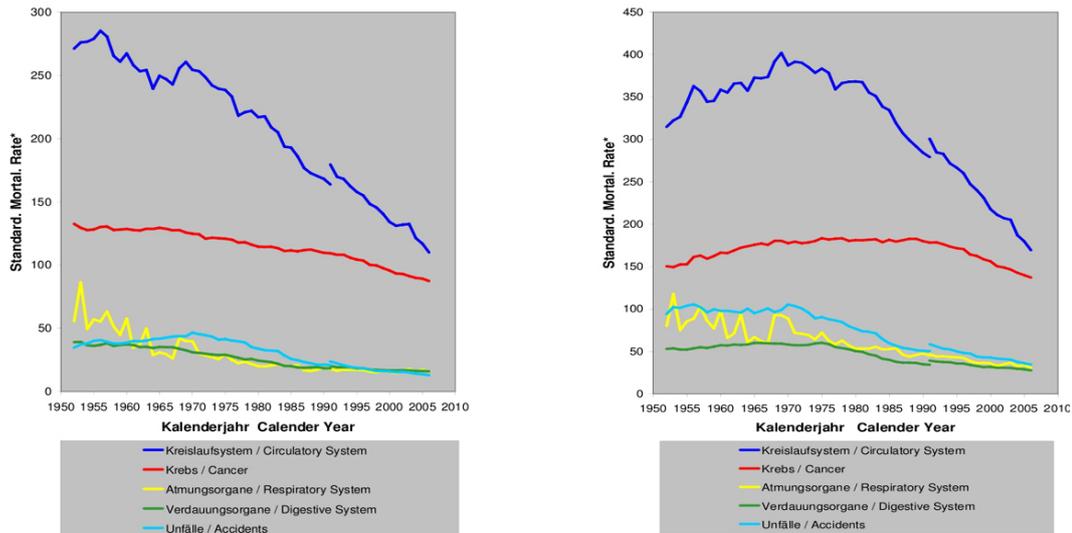
# 1 Einleitung

Krebs gilt heutzutage als zweithäufigste Todesursache nach Herz-Kreislaufkrankungen bei Frauen und Männern in Deutschland. In naher Zukunft wird die Zahl der Neuerkrankungen weiter steigen, da bösartige Neubildungen sehr vielfältig vorkommen und sich schnell wandeln. Im Jahr 2005 war bei Frauen der Brustkrebs die häufigste Todesursache, bei Männer hingegen war der Lungenkrebs der häufigster Grund für den Tod (Abbildung 1.1).

Im Gegensatz dazu geht die altersbereinigte Sterblichkeit an bösartigen Neubildungen in Deutschland erstmals seit dem Jahr 1990 für Männer zurück; bei Frauen ist dieser rückläufige Trend bereits seit Beginn der 50er Jahre zu beobachten. Dies kann durch die stetige Weiterentwicklung der medizinischen Versorgung sowie durch die bessere Früherkennung erklärt werden. Zur Zeit stirbt ungefähr jeder Dritte an einer Krankheit des Kreislaufsystems und jeder Vierte an einer Krebserkrankung (siehe Abbildung 1.2).



**Abbildung 1.1:** Die 20 häufigsten Krebstodesursachen in Deutschland im Jahr 2006. (Quelle: [www.dkfz.de/de/krebsatlas](http://www.dkfz.de/de/krebsatlas))



**Abbildung 1.2:** Trendgraphiken zu den fünf häufigsten Todesursachengruppen in Deutschland für Frauen (links) und Männer (rechts). (Quelle: [www.dkfz.de/krebsatlas](http://www.dkfz.de/krebsatlas))

Obwohl die verschiedenen Krebskrankheiten hinsichtlich ihrer Entstehung und Entwicklung, der Möglichkeit rechtzeitiger Diagnostik und Therapie sowie der Heilungsaussichten im einzelnen sehr unterschiedlich sind, konnten auf der epidemiologischen Ebene Risikofaktoren bzw. Gruppen von Risikofaktoren identifiziert werden, die jeweils für eine ganze Reihe von Krebsarten von Bedeutung sind, und deren Beitrag zum Gesamtkrebsgeschehen zumindest in ihrer Größenordnung abgeschätzt werden kann. Eine Auswahl dieser Risikofaktoren ist in folgend zusammengefasst:

- Rauchen
- Ernährungsgewohnheiten
- Alkohol
- Berufliche Exposition mit kanzerogenen Chemikalien
- Genetische Faktoren
- Schadstoffbelastung der Umwelt
- Infektiöse Erreger
- Weitere Faktoren

Die vorliegende Doktorarbeit war Teil des “*Regulatory Genomics*” EU Projektes:

*“Advanced genomics instruments, technology and methods for determination of transcription factor binding specificities; applications for identification of genes predisposing to colorectal cancer”.*

Ziel des Projektes ist die Entwicklung neuartiger Methoden zur Bestimmung der Bindungsspezifitäten krebsrelevanter Transkriptionsfaktoren. Diese Methoden werden später zur Identifizierung regulatorischer Mutationen benutzt, die in Bezug zu Darmkrebs stehen. Desweiteren werden sie zur Charakterisierung von Genen dienen, die gemeinsame Transkriptionsfaktoren besitzen.

Die Entwicklung einer auf Microarrays basierenden Methode zur Bestimmung von Protein-DNA-Wechselwirkungen im “*high-throughput* Format” war Teil dieses Projektes und Hauptbestandteil der vorliegenden Arbeit. Diese Methode wurde dann zur Identifizierung von *single nucleotide polymorphisms* (SNPs) in der Bindedomäne von Transkriptionsfaktoren benutzt, die im Verdacht stehen, Krebs zu verursachen. Die Entwicklung dieser Methode wurde in enger Zusammenarbeit mit der febit biomed GmbH (Heidelberg) durchgeführt, die ebenfalls Partner in diesem Projekt ist. Neben febit waren noch weitere Gruppen beteiligt (siehe Abbildung 1.3).

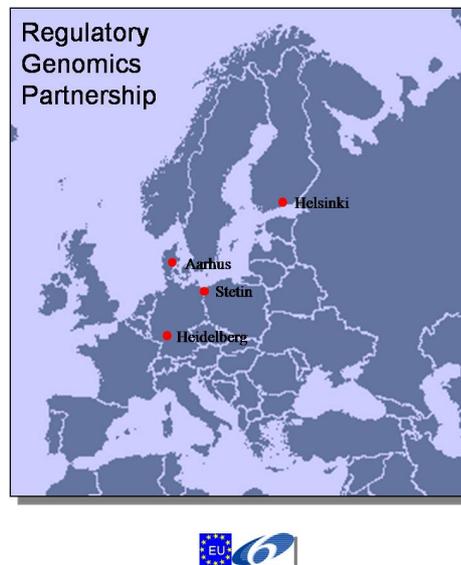


Abbildung 1.3: .

Partner-Institute des “*Regulatory Genomics*” EU Projektes: Jussi Taipale (Finnland); Lauri Aaltonen und Esko Ukkonen (Finnland), Jan Lubinski (Polen)

Jussi Taipale aus Helsinki entwickelt eine Methode zur Bestimmung der Bindungsspezifitäten von Transkriptionsfaktoren im Microtiterplatten-Format. Lauri Aaltonen und Esko Ukkonen (Universität Helsinki, Finnland) sind für die Entwicklung eines *SNP-genotyping*-Chips und für die dafür notwendigen bioinformatischen Algorithmen zuständig. Die Patientenproben für den SNP- Chip werden von Jan Lubinski (Pommersche Medizinische Universität, Polen) bereitgestellt und charakterisiert.

Die hier verwendete Microarray-Technologie hat sich aus dem in den 70er Jahren entwickelten Southern-Blot heraus zu einer Hochdurchsatz-Methode mit vielseitigen Einsatzgebieten entwickelt (Southern (1975)). Bei beiden Methoden ist die Grundidee gleich. Bestimmte Mole-

küle (DNA, DNA-Analoga, Proteine, Peptide, Antikörper) werden auf speziell modifizierten Oberflächen in einer definierten Anordnung fixiert und anschließend mit Proben inkubiert. Durch die Inkubation mit Proben ist es möglich, aus komplexen Gemischen das komplementäre Zielmolekül nachzuweisen. In den 90er Jahren wurden dann schrittweise andere Oberflächen aus Kunststoff und Glas eingeführt. Der Vorteil dieser Oberflächen liegt in der Verwendung geringerer Probenkonzentrationen. Auf einem Spot konnten somit Proben im Pico-/ Nanoliterbereich aufgebracht werden, während bei der Verwendung von porösen Membranen die 10-fache Menge benötigt wurde. Durch unterschiedliche Modifikationen der Oberflächen und die Verwendung verschiedener Lösungsmittel kann die Form und Größe der Tropfen und somit die der Spots verändert werden. Durch diesen Wechsel der Oberflächen konnte der Umstieg von Macroarrays (Spotdurchmesser größer als 300  $\mu\text{m}$ ) auf Microarrays (Spotdurchmesser kleiner als 200  $\mu\text{m}$ ) entschieden vorangetrieben werden. Diese Derivatisierung der Oberflächen und die zusätzliche Verwendung von Fluoreszenzfarbstoffen zur Detektion ermöglicht es nun tausende von Zielmolekülen auf einer Fläche von wenigen Quadratzentimetern zu untersuchen. Es können so inzwischen mehr als 200 000 Spots auf 1  $\text{cm}^2$  generiert werden (Skena et al. (1995), Pease et al. (1994)). Je nach gewünschter Anwendung werden Oberflächen und die verwendeten Proben aufeinander abgestimmt (Hoheisel (2006)). Aufgrund dieser Erfindungen und technischen Verbesserungen in den verschiedensten Bereichen führte dieser Ansatz etwa 20 Jahre nach der Entdeckung des Southern Blots zur Entwicklung der Microarray - Technologie (Jacobs and Fodor (1994), Skena et al. (1995)). Neben dem Auftragen der DNA durch Pin- oder kontaktfreie Piezospotter auf beispielsweise mit Epoxy- oder Aminosilangruppen beschichteten, chemisch modifizierten Objektträgern, gibt es bei relativ kurzen Oligonukleotiden (bis etwa 60 bp) desweiteren die Möglichkeit, diese mittels Licht-aktivierter Synthese *in-situ* direkt auf eine Trägeroberfläche zu synthetisieren (Beier and Hoheisel (2005)). Die Länge dieser *in-situ* synthetisierten Oligonukleotide ist jedoch durch die Ausbeute der einzelnen Syntheseschritte begrenzt, da die Kopplungsausbeuten nicht 100% sind. Dadurch entstehen bei jedem Synthesesyklus auf der Oberfläche Abbruchsequenzen, die mit zunehmender Länge des Zielmoleküls zunehmen und somit einen immer größeren Anteil an der Summe aller Moleküle auf dem Spot ausmachen. Das Haupteinsatzgebiet der Microarray-Technologie waren zuerst Expressionsanalysen (DNA-Microarray), mit denen die Auswirkung unterschiedliche Faktoren wie Umwelteinflüsse, Medikamente oder genetische Faktoren sowie Krankheiten untersucht werden können. Im Laufe der Zeit haben sich noch andere Anwendungen im Bereich der Microarray-Technologie entwickelt. Antikörper-, Protein-, PNA- oder Peptid-Microarrays werden heutzutage standardmäßig zum "*high-throughput-screening*" verwendet (Kusnezow and Hoheisel (2002), Jacob et al. (2004), Brandt et al. (2003)).

In dieser Arbeit sollte das Geniom System von febit für die Untersuchung von Wechselwirkungen zwischen Transkriptionsfaktoren und deren DNA-Bindeseite weiterentwickelt werden. Transkriptionsfaktoren sind Proteine, die durch DNA-Bindung die Transkription im Genom regulieren. Sie sind für die Initiation der RNA-Polymerase sowie bei der Regulation der Elongation und Termination von Bedeutung. Transkriptionsfaktoren können an die DNA binden und den Promotor aktivieren oder reprimieren. Es gibt auch Transkriptionsfaktoren, die nicht

direkt an die DNA, sondern zum Beispiel an andere DNA-bindende Proteine binden. Die Regulierungen durch Transkriptionsfaktoren können sehr vielseitig sein. So sind Phosphorylierungen an verschiedenen Stellen genauso möglich wie Dephosphorylierungen, Acetylierungen, Glykosylierungen oder proteolytische Spaltungen (Polevoda and Sherman (2007), Riobo et al. (2006)). Dies muss nicht nur den Transkriptionsfaktor selbst betreffen, sondern auch mit ihm assoziierte Moleküle wie Liganden, Kofaktoren, Inhibitoren oder andere Partnerproteine. Es sind zur Zeit etwa 40 Transkriptionsfaktoren bekannt, bei denen eine Veränderung der Genregulation mit Krebs in Verbindung gebracht werden. Es entsteht durch Mutationen in Genen, deren Proteine an der Kontrolle des Zellwachstums beteiligt sind. Durch diese Mutationen verlieren die Proteine ihre normale Funktion bei der Kontrolle des Zellwachstums. Diese Gene nennt man Onkogene, die entsprechenden Proteine Onkoproteine. Die Zelle beginnt sich unkontrolliert zu teilen (Proliferation) und wird dadurch unsterblich (Immortalisation). Zugleich verliert die Zelle Form und Funktion ihrer Vorläuferzelle. Der Mechanismus mit dem diese onkogenen Transkriptionsfaktoren das Wachstum regulieren ist bis heute noch nicht ganz verstanden. Nur wenige Ausnahmen, wie z. B. der Östrogen Rezeptor oder der Androgen Rezeptor wurden bis jetzt charakterisiert. Onkoproteine können durch Basenaustauschmutationen und Rastermutationen entstehen. Bei Rastermutationen werden Basen eingeschoben oder weggelassen. Hierbei wird das Leseraster verschoben; es entsteht eine völlig neue Aminosäurekette und das Protein ist funktionsunfähig. Bei Basenaustauschmutationen handelt es sich um den Austausch einer Base in der DNA. Bei neutralen Mutationen bleiben die Aminosäuren gleich und das Protein funktionsfähig. Bei Fehlsinmmutationen entsteht ein fehlerhafter Sinn, da durch den Austausch der Basen eine andere Aminosäure in das Protein eingebaut wird. Das Protein kann dadurch funktionsunfähig werden. Durch den Austausch einer Base kann auch ein Stopp-Codon entstehen, welches die Proteinbiosynthese vorzeitig beendet und somit das Protein stark verkürzt. Damit ist das Protein unvollständig und somit funktionsunfähig. Eine Veränderung der Genexpression und somit eine Entscheidung über das Proliferationsverhalten kann nach der Mutation nur über mehrere Molekülstationen umgesetzt werden. Dies kann anhand der Signal-Transduktionskette für einen Wachstumsfaktor gut veranschaulicht werden. Physiologischerweise werden die Moleküle dieser Kette an- oder abgeschaltet. In Abwesenheit des Wachstumsfaktors bleiben sie inaktiv und werden erst durch ihn aktiviert. Bei der Onkogenaktivierung verändert eine Mutation eines dieser Moleküle so, dass es immer im angeschalteten Zustand vorliegt, also nicht mehr abgeschaltet werden kann. Für die Zelle wird durch die Mutation das Vorhandensein eines Wachstumssignals vorgetäuscht und reagiert darauf mit Proliferation. Die mit Krebs in Verbindung gebrachten Transkriptionsfaktoren sind in folgenden Tabellen aufgeführt.

FAKTOR	FAMILIE, LIGAND	UNGEWÖHNLICH REGULIERT IN
TCF/Lef	Wnt	Kolonkarzinom
GLI	Hedgehog	Basalzellkarzinom oder Medulloblastom
CSL	Notch	akute T-Zell lymphoblastische Leukämie
Smad Proteine	TGF-B Superfamily	Darm- und Pankreaskrebs
Androgen Rezeptor	Testosteron	Prostata Krebs
Östrogen Rezeptor	Östrogen	Brustkrebs
Retinalsäure Rezeptor Alpha	Retinalsäuren	akute promyelozyten Leukämie

**Tabelle 1.1:** Transkriptionsfaktoren, deren Mutationen eine Veränderung der Signal-Weitergabe verhindern.

Bei weitem nicht alle Mutationen in einem Protoonkogen haben diesen Effekt: die meisten Mutationen führen dazu, dass das kodierte Protein nicht mehr funktioniert (*loss of funktion*) oder immer aktiviert (*gain of funktion*) ist. Folgende Tabellen zeigen Beispiele für diese Effekte.

FAKTOR	AKTIVITÄT BEI KREBS VERÄNDERT	TUMOR
Myc family	erhöht	Viele, aber nicht alle
E2F family/DP1	erhöht	Viele, aber nicht alle
p53	verloren oder erniedrigt	Viele Tumortypen, insbesondere epitheliale

**Tabelle 1.2:** Transkriptionsfaktoren, die den Zellzyklus und die DNA Reparatur regulieren.

FAKTOR	EFFEKT DER MUTATION	TUMOR
Smad4	Funktionsverlust	Darm- und Pankreaskrebs
WT1	Funktionsverlust	Wilms Tumor
MLL	Translokation und Funktionszunahme	Verschiedene Abstammungen (Rückenmark, Lymphknoten); Leukämie
sarkom	Translokation und Funktionsverlust	AML
PBX1	Translokation	B-Zellen-Lymphom
BCL6	Translokation und Funktionszunahme	B-Zellen-Lymphom

SCL/TAL1	Translokation und Funktionszunahme	T-Zellen ALL
Myc family	Translokation und Funktionszunahme	T-Zellen ALL, Burkitt's lymphoma, Neuroblastoma
Hox 11 (homeobox)	Translokation und Funktionszunahme	T-Zellen ALL
FLI1 (ETS family)	Translokation und Funktionszunahme	Ewings-Sarkom, AML
Erg (ETS family)	Translokation und Funktionszunahme	Ewings-Sarkom, AML
ETV1 (ETS family)	Translokation und Funktionszunahme	Ewings-Sarkom
ETS1 (ETS family)	Translokation und Funktionszunahme	AML, ALL
FEV (ETS family)	Translokation und Funktionszunahme	Maxillary Tumor
Pax3 or Pax7 (homeobox)	Translokation und Funktionszunahme	Alveolar Rhabdomyosarkom
ATF1	Translokation und erhöhte Aktivität	Klarzellsarkom
CHOP	Translokation und Funktionszunahme	Myxoid Liposarkom
NR4A3	Translokation und Funktionszunahme	Extraskeletal myxoides Chondrosarkom
TFE3	Translokation und Funktionszunahme	Renales Adenokarzinom

**Tabelle 1.3:** Transkriptionsfaktoren, deren Aktivität direkt durch Mutationen verändert wird.

**Quelle:** Hanahan and Weinberg (2000), Taipale and Beachy (2001), Howe et al. (1998).

Der Einfluss von Mutationen in der DNA Bindesequenz, welche das Binden von Transkriptionsfaktoren beeinflussen, ist noch nicht vollständig geklärt. Bis jetzt ist jedoch nur wenig über Mutationen in diesen Regionen bekannt, die Krankheiten wie z.B. Krebs verursachen können. Nach erfolgter Aktivierung eines Transkriptionsfaktors bindet dieser über spezielle DNA-Bindedomänen an seine Zielsequenz im Genom. Bei dieser Bindung werden nicht direkt Abfolgen von Nukleotiden erkannt, sondern die daraus resultierenden räumlichen Strukturen der DNA (Laity et al. (2001)). Die Spezifität und die Stärke der Bindung werden durch Bildung von Wasserstoffbrücken, Ionenbindungen, sowie hydrophobe Wechselwirkungen beeinflusst. Dementsprechend ist es auch möglich, dass ein und derselbe Faktor an mehrere, leicht variierende Sequenzen binden kann. Wie stark dadurch das Bindungsverhalten geändert wird, hängt von Art und Position der ausgetauschten Basen ab. Die Erkennungssequenzen sind von

Faktor zu Faktor unterschiedlich und können sich zusätzlich in ihrer Länge unterscheiden. Die biologischen Eigenschaften eines Proteins hängen also von seinen physikalischen Wechselwirkungen mit anderen Molekülen ab. In manchen Fällen ist diese Bindung sehr fest, in manchen anderen Fällen ist sie sehr locker und kurzlebig. Jedoch ist diese Bindung in allen Fällen spezifisch. Die Spezifität und die Stärke der Bindung werden wie bereits oben erwähnt durch eine Vielzahl von Wechselwirkungen bestimmt. Da diese einzelnen Wechselwirkungen alleine nur sehr schwach sind, müssen um eine effektive Wechselwirkung zu erzielen, viele Bindungen gleichzeitig gebildet werden. Dies ist jedoch nur möglich, wenn die verschiedenen räumlichen Strukturen und somit die dadurch induzierten Wechselwirkungen sehr genau übereinstimmen. Somit wird ein sehr vielseitiger und spezifischer Mechanismus gebildet. Werden diese räumlichen Strukturen durch Mutationen gestört, kann die Transkription nicht richtig reguliert werden. Dies kann sich wie bereits oben erwähnt auch auf die Signalkaskaden auswirken und zur Proliferation führen.

Die Information über diese spezifischen Transkriptionsfaktorbindestellen sind meist unvollständig oder es ist nur die Sequenz mit der höchsten Affinität bekannt. Viele dieser so genannten Konsensus Sequenzen geben nur eine einzelne Sequenz an. Dies entspricht nicht der Realität, da die wenigsten Transkriptionsfaktoren ausschließlich an eine bestimmte Sequenz binden. Theoretische Modelle können diese Sequenzen nur unvollständig beschreiben oder vorhersagen. Von den im menschlichen Genom geschätzten 2000 Transkriptionsfaktoren sind bislang nur ein Bruchteil der Bindestellen bekannt, gut charakterisiert und in öffentlichen Datenbanken wie JASPAR (Stormo (2000), Wasserman and Sandelin (2004)) oder TRANSFAC (Knueppel et al. (1994), Matys et al. (2006)) zugänglich.

Es existieren zur Zeit mehrere Methoden, mit denen die Bindungsspezifitäten von Transkriptionsfaktoren untersucht werden können. Diese basieren zum größten Teil auf Gel-Elektrophorese oder Microarray-Technologie. Aber auch Methoden wie Systematische Evolution von Liganden durch exponentielle Anreicherung (SELEX) können hierfür verwendet werden (Ellington and Szostak (1990)). Bei der SELEX-Methode werden Molekülbibliotheken mit dem zu untersuchenden Zielmolekül (z.B. ein Transkriptionsfaktor) inkubiert. Nach mehreren Wasch- und Selektionsschritten wird der mit dem Zielmolekül interagierende Ligand (z.B. DNA) mittels Sequenzierung nachgewiesen. Roulet et al. (2002) sequenzierten mehr als 1000 Liganden, um das Bindungsverhalten von CTF/NFI mittels SELEX zu untersuchen. Den Kosten von 10 € pro Sequenzierung stehen zusätzlich viele Wochen Laborarbeit gegenüber.

*DNAse footprinting* (Senear et al. (1986)) und *electrophoretic mobility shift assay* (EMSA; Garner and Revzin (1981)) sind Methoden, bei denen die Protein-DNA-Komplexe mittels Gelelektrophorese aufgetrennt und analysiert werden. Hierbei ist die Anzahl der zu untersuchenden Proben sehr begrenzt. Da die Interaktionen nicht direkt gemessen werden können, ist die Sensitivität dieser Methoden sehr beschränkt.

Auch die Microarray-Technologie wird für diese Untersuchungen verwendet. Man kann hier zwischen der Probengewinnung *in-vivo* und *in-vitro* unterscheiden. Bei der *DNA adenine methyltransferase identification*-Technologie (DamID) wird das zu untersuchende Protein als

Fusionsprotein mit Dam *in vivo* überexprimiert. Nach Binden des Transkriptionsfaktors an seine Zielsequenz werden in der näheren Umgebung Adenine, in GATC-Sequenzen, durch Dam *in-vivo* methyliert. Anschließend können die durch Methyl-spezifischen Restriktionsverdau erhaltenen DNA-Fragmente mit Kontrollen verglichen werden. Dies geschieht nach Amplifizierung und Markierung der DNA ebenfalls durch Microarrayanalyse. Etabliert wurde diese Methode für die Spezies *Drosophila* (van Steensel et al. (2001)) und *Arabidopsis* (Tompa et al. (2002)). Bei der ChIP-Chip-Technologie werden die Protein-DNA Komplexe *in-vivo* mit einem Crosslinker wie Formaldehyd inkubiert. Nach dem Lysieren der Zellen können diese Komplexe aufgereinigt und untersucht werden (Lieb et al. (2001), Ren et al. (2000)). Der größte Vorteil der ChIP-Chip-Methode ist die Analyse posttranslationaler Modifikationen und kooperativer Effekte anderer Bindungspartner. Ein Nachteil der oben genannten Methoden ist die Länge der untersuchten DNA. Sie ist größer als die eigentliche Bindesequenz und limitiert das Auflösungsvermögen dieser Techniken.

Mit der Plasmonenresonanzspektroskopie können Protein-DNA-Wechselwirkungen zur Zeit am genauesten untersucht werden. Es ist möglich, sowohl die Bindungskonstanten als auch die Bindungsenergien der Protein-DNA Bindung zu bestimmen.

Der größte Nachteil all dieser Methode ist die limitierte Anzahl der zu untersuchenden Proben.

Eine recht neue Methode zur Bestimmung von Protein-DNA-Interaktionen sind die sogenannten Protein bindenden Microarrays (PBMs). Hierbei werden in *E. coli* - Bakterien oder Insektenzellen exprimierte Fusionsproteine auf DNA Microarrays untersucht. Mukherjee et al. (2004) entwickelten eine Methode basierend auf gespotteten DNA-Chips. Im Gegensatz dazu verwenden Bulyk (2006) und Warren et al. (2006) synthetisierte oder kommerziell erhältliche Microarrays, die alle möglichen Permutationen einer 10 bp langen Sequenz beinhalten. Diese Methoden sind sehr effektiv, wenn keine Informationen über die Bindeseiten vorliegen. Wie bei allen PBMs können leider nicht die Einflüsse von Kofaktoren untersucht werden und auch eine Beeinträchtigung der Bindungseigenschaften durch das Fusionsprotein kann nicht ausgeschlossen werden.

## Zielsetzung der Arbeit

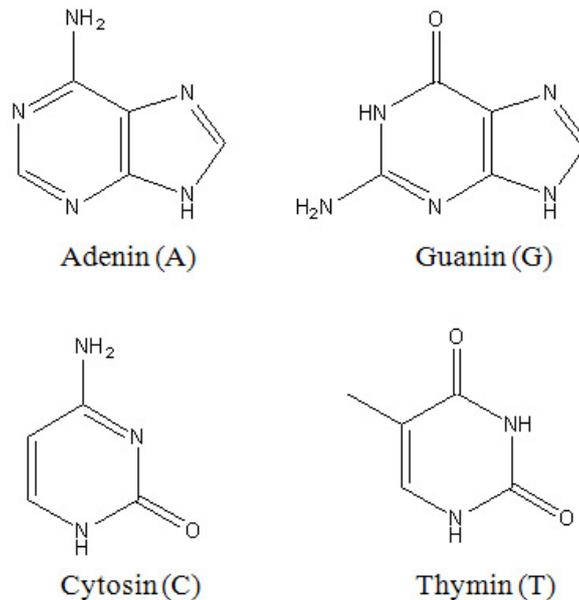
Ziel der hier vorliegenden Arbeit war Entwicklung und Etablierung einer auf Microarray basierenden Methode zur Bestimmung von Protein-DNA-Wechselwirkungen im *high throughput* Format. Hierzu sollte die Geniom-Plattform in Zusammenarbeit mit der febit biomed GmbH weiterentwickelt und für diese Experimente etabliert werden. Zunächst wurde anhand des Fusionsproteins MBP-Gli2 die Entwicklung und Etablierung der Methode durchgeführt. In zweiten Teil wurde diese Methode genutzt, um die Bindungseigenschaften von 10 Transkriptionsfaktoren zu untersuchen. Diese wurden allesamt in *E. coli* - Bakterien als Fusionsproteine exprimiert. Da der vollständige Transkriptionsfaktor aufgrund seiner Größe nicht als Fusionsprotein in *E. coli* exprimiert werden kann, wurde für die Generierung nur die für die DNA Bindung notwendige Bindedomäne verwendet. Die Fusionsproteine sind alle nach folgendem Schema aufgebaut: MBP-3xMyc-TF. Das Maltose bindende Protein (MBP) wurde für die Aufreinigung der Fusionsproteine benötigt. Dadurch wurde zusätzlich die Löslichkeit und Stabilität des Fusionsproteins erhöht, da in vielen Fällen die Bindedomäne des Transkriptionsfaktors alleine schlecht löslich und instabil ist. Mit Hilfe der 3-fach Myc-Markierung (3xMyc) kann unter der Verwendung eines Anti-Myc-Antikörpers und des entsprechend fluoreszenz-markierten 2. Antikörpers das Fusionsprotein detektiert werden.

## 2 Theoretischer Teil

Im folgenden Kapitel soll auf die theoretischen Grundlagen eingegangen werden.

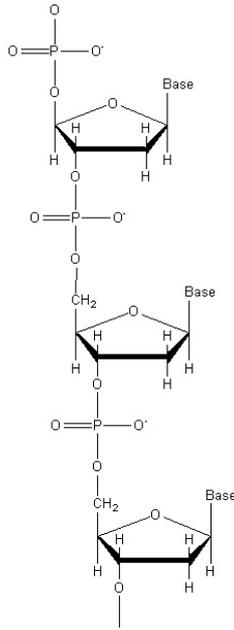
### 2.1 Aufbau der DNA

Die Nukleinsäure ist die Quelle der biologischen Information. Die Formen und Aktivitäten individueller Zellen sind durch genetische Konstruktionen bestimmt, die in der DNA enthalten sind. Bei allen Zellen ist Erbinformationen in Form von doppelsträngigen DNA (Desoxyribonukleinsäure) Molekülen gespeichert, die aus den gleichen Basen besteht:



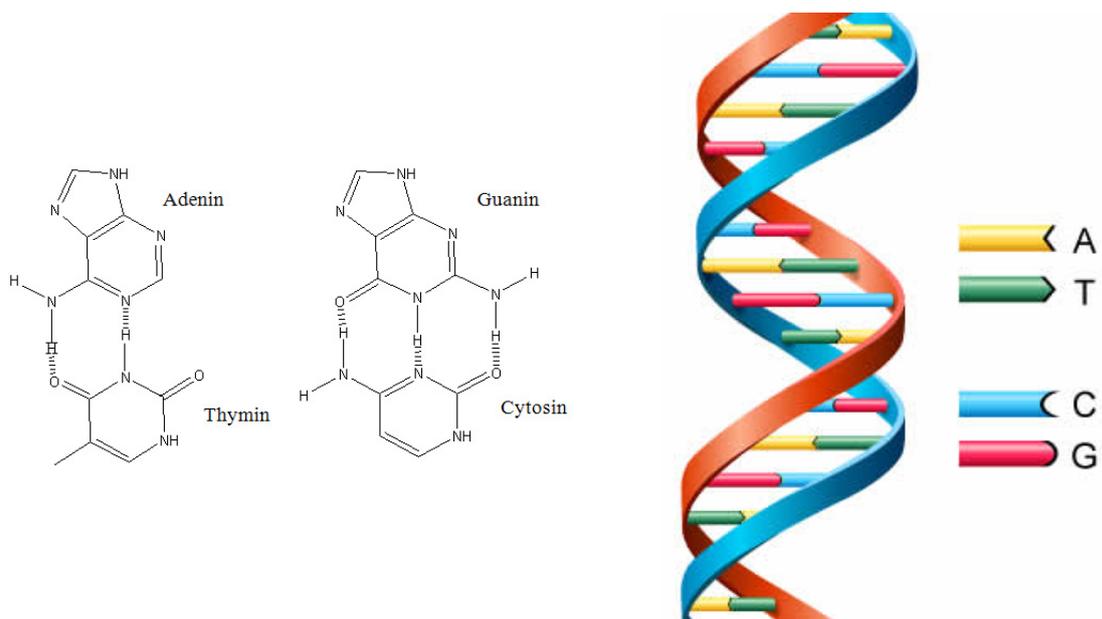
**Abbildung 2.1:** Die vier Basen der DNA.

Die DNA ist ein Polymer bestehend aus Deoxyribose, die durch Phosphodiesterbindungen miteinander verknüpft sind. Zwischen dem Zucker (Ribose) und der Phosphatgruppe entsteht eine Esterbindung (siehe Abbildung 2.2).



**Abbildung 2.2:** Zuckerphosphatrückgrat der DNA mit beliebigen Nucleobasen (B).

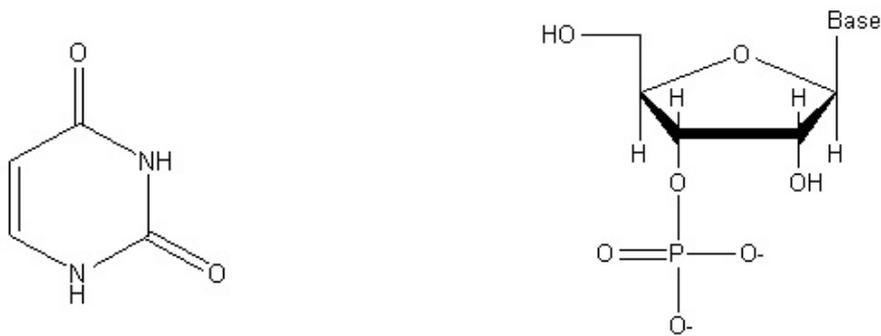
Die DNA ist in ihrer Struktur dynamisch und kann verschiedene Konformationen annehmen. In Zellen kommt die DNA nicht als Einzelstrang, sondern als Doppelstrang vor. Die hierfür notwendige Basenpaarung folgt strengen Regeln. So werden zwischen Thymin und Adenin zwei und zwischen Cytosin und Guanin drei Wasserstoffbrückenbindungen ausgebildet. Die beiden Stränge drillen sich umeinander und bilden eine Doppelhelix (siehe Abbildung 2.3). Watson und Crick stellten dies zum ersten Mal 1953 fest (Watson and Crick (1953)).



**Abbildung 2.3:** Basenpaarung der Basen Adenin und Thymin bzw. Cytosin und Guanin (links). DNA-Helix bestehend aus den vier Nucleotiden (Adenin A, Thymin (T), Cytosin (C) und Guanin (G) (rechts).

## 2.2 Transkription und Translation

Als Transkription (v. lat. trans = jenseits, hinüber + scribere = schreiben) wird in der Genetik die Synthese von RNA anhand einer DNA als Vorlage bezeichnet. Dies wird durch das Binden von Transkriptionsfaktoren an die DNA reguliert. Dabei wird ein Gen abgelesen und als RNA-Molekül vervielfältigt, das heißt ein spezifischer DNA-Abschnitt dient als Vorlage zur Synthese eines neuen RNA-Strangs. Bei diesem Vorgang werden die Nukleinbasen der DNA (A,T,G,C) in die Nukleinbasen der RNA (A,U,G,C) umgeschrieben. Anstelle des Thymins kommt Uracil und anstelle der Desoxyribose kommt Ribose in der RNA vor (siehe Abbildung 2.4).



**Abbildung 2.4:** Uracil (links) und Ribose (rechts).

Man kann zwischen mehreren Arten von RNA unterscheiden. Die m-RNA (engl. messenger RNA) kopiert die in einem Gen auf der DNA liegende Information und überträgt sie zum Ribosom, wo mit Hilfe dieser Information die Proteinbiosynthese stattfinden kann. Die Transfer-RNA (t-RNA) kodiert keine genetische Information, sondern dient als Hilfsmolekül bei der Proteinbiosynthese. Sie nimmt eine einzelne Aminosäure aus dem Cytoplasma auf und transportiert sie zum Ribosom. Die ribosomale RNA (r-RNA), trägt, ähnlich wie die t-RNA, keine genetische Information, sondern ist am Aufbau des Ribosoms beteiligt und ist bei der Knüpfung der Peptidbindung auch katalytisch aktiv.

Während der Transkription werden also aus RNA-Monomeren anhand der DNA-Matrize m-RNA-Polymermoleküle synthetisiert, deren Nukleotidsequenz einen komplementären Teil der Erbinformation der Zelle wiederholt.

Als Translation bezeichnet man die Übersetzung der genetischen Information der RNA in Proteine. Die Information in der RNA Sequenz liegt in Gruppen von drei Nukleotiden (Tripletts oder Codon) vor. Jedes dieser Codons codiert für eine Aminosäure in einem zugehörigen Protein. Da es 64 ( $= 4 \times 4 \times 4$ ) mögliche Codons gibt, aber nur 20 Aminosäuren, muss es notwendigerweise Fälle geben, in denen mehrere Codons für ein und dieselbe Aminosäure stehen. Die t-RNA besitzt eine spezifische Sequenz von drei Nukleotiden (Anticodon), die es ihr ermöglicht, durch Basenpaarung ein bestimmtes Codon auf der m-RNA zu erkennen. Zur Synthese eines Proteins müssen also die mit ihrer entsprechenden Aminosäure beladenen t-RNAs mit einer m-RNA im Ribosom zusammengebracht werden. Die t-RNAs binden mit ihren

Anticodons über komplementäre Basenpaarung an die jeweils passenden Codons der m-RNA. Dann werden die Aminosäuren miteinander verknüpft, um die wachsende Polypeptidkette zu verlängern.

### 2.3 Proteine und ihre Struktur

Proteine sind die strukturell komplexesten und funktionell am feinsten abgestimmten Moleküle, die wir kennen. Sie bestehen aus 20 verschiedenen Aminosäuren, von denen jede ihre eigenen chemischen Eigenschaften besitzt. Ein Proteinmolekül besteht aus einer langen Aminosäurekette, in der jeden Aminosäure mit ihren Nachbarn über eine Peptidbindung kovalent verknüpft ist. Die sich wiederholende Abfolge von Atomen entlang des Rückgrades entlang einer Polypeptidkette wird Polypeptid-Grundgerüst genannt. Peptide mit einer Kettenlänge von 2 bis ca. 20 werden als Oligopeptide bezeichnet. Längere Peptide bezeichnet man allgemein als Polypeptide, wobei der Übergang Oligo- Poly- fließend ist. Eine Kette mit über 100 Aminosäuren nennt man Protein.

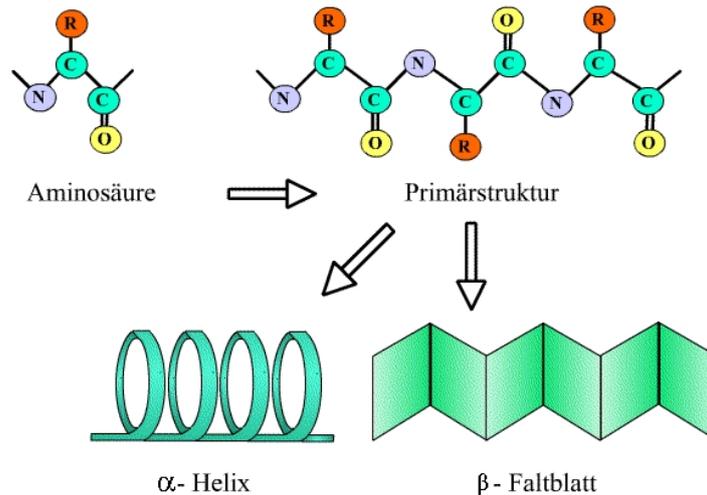
#### 2.3.1 Primärstruktur

Die Primärstruktur eines Proteins ist durch seine Aminosäuresequenz gekennzeichnet. Beginnend mit der N-terminalen Aminosäure werden alle Aminosäuren bis zur letzten, C-terminalen Aminosäure an einer Kette aufgereiht. Aus dieser Struktur leiten sich dann alle weiteren Strukturen des Proteins ab. Leider lässt sich allein aus dieser Sequenz zur Zeit nicht die endgültige Struktur des Proteins vorhersagen.

#### 2.3.2 Sekundärstruktur

Die Sekundärstruktur beschreibt die räumliche Anordnung von Aminosäuren einer Polypeptidkette, die nahe beieinander liegen. Diese Struktur wird durch Wasserstoffbrückenbindungen und andere Wechselwirkungen bestimmt. Man kann hierbei zwischen der  $\alpha$ -Helix und der  $\beta$ -Faltblattstruktur unterscheiden.

Eine schematische Darstellung dieser Strukturen ist in Abbildung 2.5 gezeigt..



**Abbildung 2.5:** Schematische Darstellung der räumlichen Proteinstrukturen.

### 2.3.3 Tertiärstruktur

Unter der Tertiärstruktur eines Proteins versteht man die räumliche Anordnung der Aminosäureketten. In der Tertiärstruktur sind viele  $\alpha$ - Helix-,  $\beta$ - Faltblatt- und ungeordnete Bereiche räumlich dargestellt. Die Tertiärstruktur verleiht dem Protein erst seine Funktion, ob es ein Strukturprotein oder ein Enzym ist, ob es sich um ein globuläres Protein oder ein fadenförmiges Molekül handelt.

### 2.3.4 Quartärstruktur

Wenn ein Proteinmolekül aus mehr als einer Polypeptidkette zusammengesetzt ist, bezeichnet man diesen Komplex als Quartärstruktur. Er besteht somit aus verschiedenen Untereinheiten.

## 2.4 Transkriptionsfaktoren und die Strukturelemente DNA bindender Proteine

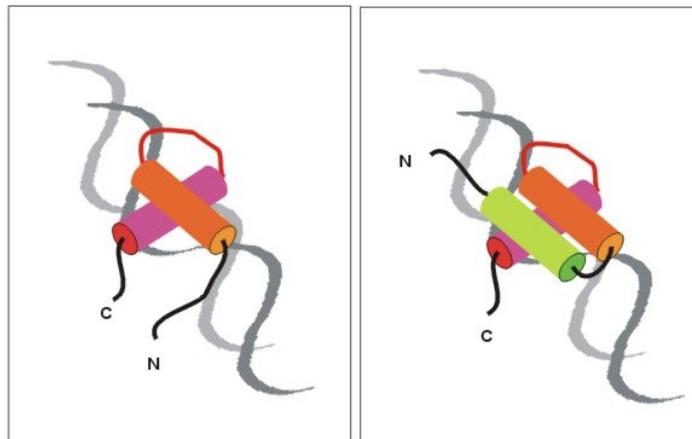
Transkriptionsfaktoren sind Proteine, die durch DNA-Bindung die Transkription im Genom regulieren. Der vereinfachte schematische Aufbau eines Transkriptionsfaktors kann in drei Untereinheiten aufgeteilt werden:

- Aktivierungsdomäne
- Signalweitergabe-Domäne
- DNA-Bindedomäne

Die Aktivierungsdomäne enthält Bindestellen für andere Proteine, um mit dem Transkriptionsapparat interagieren zu können, während die Signalweitergabe-Domäne für die Weiterverarbeitung der Signale verantwortlich ist. Nach erfolgter Aktivierung eines Transkriptionsfaktors bindet dieser über eine spezielle DNA-Bindedomäne an seine Zielsequenz im Genom. Diese Bindedomäne besteht aus Aminosäuren, die spezifisch die räumliche Struktur der DNA-Basen in der Nähe des Transkriptionsstartes erkennen. Diese Erkennungssequenzen können sich von Faktor zu Faktor in den verschiedenen Nukleotidabfolgen aber auch in ihrer Länge unterscheiden. Obwohl jedes Beispiel der Protein/DNA-Wechselwirkung für sich im Detail einmalig ist, deuten die mittels Röntgenstrukturanalyse und Kernspinresonanzspektroskopie durchgeführten Untersuchungen an, dass viele dieser Proteine bestimmte Gruppen DNA bindender Struktur motive enthalten können. Jedes diese Motive benutzt entweder  $\alpha$ -Helices oder  $\beta$ -Faltblätter, um an die große Furche der DNA zu binden. Typische DNA-Bindedomänen sind das Helix-Turn-Helix-Motiv, das basische Helix-Loop-Helix-Motiv, der Leucin-Zipper oder der C2H2-Zinkfinger.

#### 2.4.1 Helix-Turn-Helix (HTH)

Das Helix-Turn-Helix Motiv ist das erste DNA bindende Proteinmotiv, das entdeckt wurde. Es wird durch zwei  $\alpha$ -Helices gebildet, die durch eine kurze, gestreckte Aminosäurekette verbunden sind, die die Umkehrschleife oder den 'Turn' bildet (siehe Abbildung: 2.6). Dieses Motiv bindet somit als symmetrisches Dimer an die DNA. Dadurch wird die Bindungsaffinität deutlich erhöht, da die Anzahl der Kontaktstellen verdoppelt wird. Dies bewirkt eine Vervierfachung der Affinitätskonstante.



**Abbildung 2.6:** Bindung von Helix-Turn-Helix-Proteinen (links) und homöotischen Proteine (rechts) an eine DNA-Bindungsstelle in der großen Furche der DNA. (Quelle: <http://biochemie.web.med.uni-muenchen.de>)

Außerhalb der HTH-Region kann die Struktur der verschiedenen DNA bindenden Proteine, die dieses Motiv enthalten, variieren. Auf diese Weise wird die Vielseitigkeit des HTH-Motivs verstärkt.

### 2.4.2 Proteine mit Homöodomänen

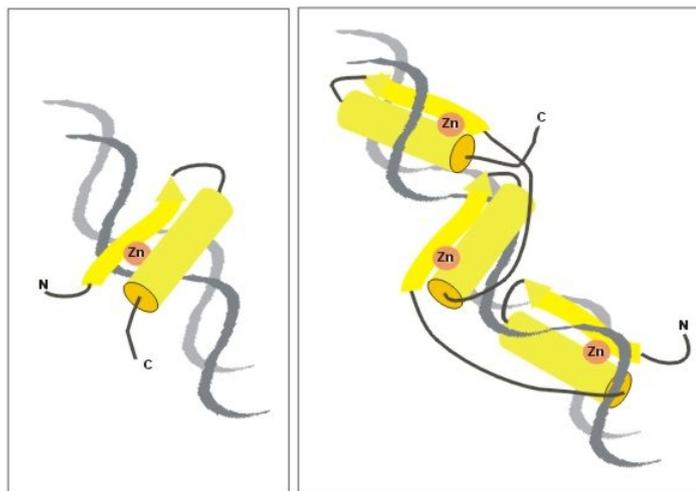
Die Homöodomäne ist eng mit der HTH-Domäne verwandt. Sie ist im Gegensatz dazu zu drei  $\alpha$ -Helices gefaltet, die durch hydrophobe Wechselwirkungen zusammengehalten werden. Das HTH-Motiv ist in dieser Struktur eingebettet und konserviert. Die schematische Struktur ist in Abbildung 2.6 gezeigt.

### 2.4.3 Zinkfinger-Motive

Das Zink-Finger-Motiv benutzt zur Bindung an die DNA ein oder mehrere Zinkionen als Strukturkomponente. Es können bei dieser Struktur zwei verschiedenen Typen unterschieden werden:

- Eine  $\alpha$ -Helix und ein  $\beta$ -Faltblatt werden durch ein Zink Ion zusammengehalten.
- Zwei  $\alpha$ -Helices sind um ein Zink Atom gefaltet.

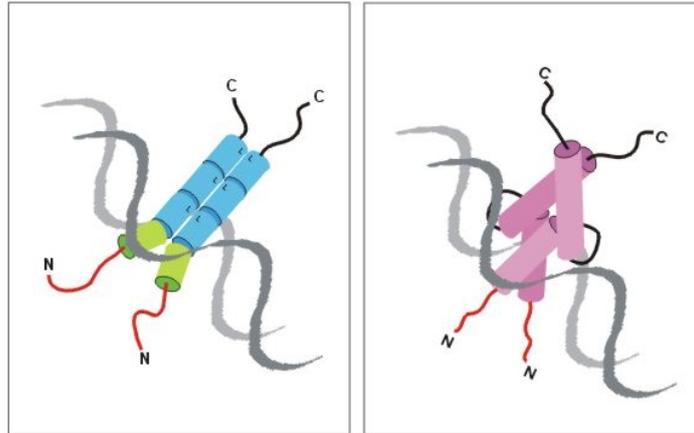
Eine schematische Darstellung ist in Abbildung 2.7 zu sehen.



**Abbildung 2.7:** Bindung eines Zinkfingers (links) und eines Zinkfingerproteins mit mehreren Zinkfingern (rechts) an ihre Bindungssequenz(en) in der DNA. (Quelle: <http://biochemie.web.med.uni-muenchen.de>)

### 2.4.4 Leucin-Zipper-Motiv

Das Leucin-Zipper-Motiv bindet als Dimer an die DNA. Im Gegensatz zu den anderen Motiven ist hier der gleiche Teil des Proteins für die Dimerbildung und die DNA Bindung verantwortlich. Normalerweise sind zwei verschiedene Teile des Proteins für diese Wechselwirkungen zuständig. Die beiden  $\alpha$ -Helices eines jeden Monomeres werden so miteinander verbunden, dass sie eine kurze Doppelwendel bilden (siehe Abbildung 2.8). Die Helices werden durch hydrophobe Wechselwirkungen der Seitenketten zusammengehalten. Oftmals ist hierfür das Leucin zuständig.



**Abbildung 2.8:** Bindung eines Leucinzipper-Proteins (links) und eines Helix-Loop-Helix-Proteins (rechts) an ihre Bindungssequenzen in der DNA. Leucinzipper sind blau gezeichnet, basische  $\alpha$ -Helices grün. (Quelle: <http://biochemie.web.med.uni-muenchen.de>)

#### 2.4.5 Helix-Loop-Helix (HLH)

Das Helix-Loop-Helix-Motiv ist mit dem Leucin-Zipper-Motiv verwandt. Im Unterschied dazu sind bei dieser Bindung vier  $\alpha$ -Helices an der DNA Bindung beteiligt, wobei je zwei Helices eine Schleife bilden. Diese können als Homodimer oder als Heterodimer spezifische Kontakte mit der DNA ausbilden. Dies wird in Abbildung 2.8 verdeutlicht.

### 2.5 Expression von Fusionsproteinen und Aufreinigung über Affinitätschromatographie

#### 2.5.1 Proteinexpression in *E. coli* - Bakterien

Für die Proteinexpression werden im Allgemeinen *E. coli* - Bakterien verwendet. Den vielen Vorteilen wie Ausbeute, einfache und billige Kultivierung stehen auch einige schwerwiegende Nachteile wie z.B. keine posttranslationalen Modifikationen gegenüber. Nachfolgend seien einige dieser Vor- und Nachteile erwähnt:

##### Vorteile

- hohe Ausbeute
- einfache, billige Kultivierung
- gute Raum/Zeit-Ausbeute
- Organismus umfassend bekannt, einfach zu handhaben und zu manipulieren
- zahlreiche Kontrollelemente auf Transkriptions- und Translationsebene bekannt

## Nachteile

- keine posttranslationalen Modifikationen wie in eukaryotischer Zellen, daher häufig keine bioaktiven Proteine
- Expressionsprodukte eventuell schwerlöslich in Form von sog. „inclusion bodies“

In der Regel werden Fusionsproteine für die Expression in *E. coli* - Bakterien verwendet.

### 2.5.2 Fusionsproteine

Ein Fusionsprotein, auch Hybridprotein genannt, entsteht durch die gemeinsame Expression zweier Gene oder Genteile, die hintereinander im Genom liegen. Durch Entfernen des Stopp-Codons hinter dem ersten Gen oder durch eine Verschmelzung durch eine Chromosomenveränderung (bspw. Translokation) werden beide Gene so abgelesen, als ob es sich um ein einziges Gen handeln würde. Dadurch ist es möglich verschiedene Proteinuntereinheiten zu verknüpfen.

Klassische Fusionstags, d.h. Proteinteile, mit denen Proteine fusioniert werden, sind durch verschiedene Eigenschaften gekennzeichnet (Consortium et al. (2008)). Sie sollten selbst gut exprimierbar sein und die Löslichkeit von Proteinen nicht negativ beeinflussen. Weiterhin sollten sie eine hohe Affinität zu einer bestimmten Matrix besitzen, die eine Aufreinigung ermöglicht. Dazu sollte sich die Bindung an diese Matrix mit einem Elutionsmittel wieder gut lösen lassen. Die folgende Auflistung zeigt eine Auswahl klassischer Fusionstags und der jeweiligen Möglichkeit zur Elution:

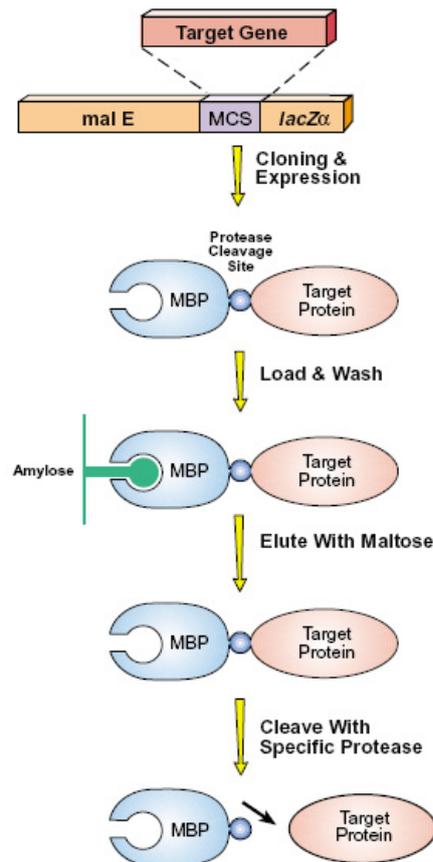
- MBP: *Maltose Binding Protein*. Aufreinigung über Amylose-Harz-Matrix. Elution mit Maltose.
- GST: Glutathion-S-Transferase. Aufreinigung über Glutathion-Sepharose. Elution mit reduziertem Glutathion.
- His: Histin. Aufreinigung über Nickel ( $\text{Ni}^{2+}$ ). Elution mit Imidazol.
- CBP: *Calmodulin Binding Peptide*. Aufreinigung über Calmodulin. Elution mit EGTA.

Die Affinitäts-Tags können somit zur Aufreinigung eines Fusionsproteins als auch für die Detektion mittels Antikörper verwendet werden.

### 2.5.3 Proteinaufreinigung

#### 2.5.3.1 Affinitätschromatographische Aufreinigung über eine Amylose-Harz-Matrix

Das Maltose-bindenden-Protein kann zur Aufreinigung von gekoppelten Transkriptionsfaktoren über eine Amylose-Harz-Matrix verwendet werden. Hierfür nutzt man die natürliche Affinität des Maltose-bindenden-Proteins an eine Amylose-Harz-Matrix aus. Eine schematische Darstellung dieser Aufreinigung ist in Abbildung 2.9 gezeigt.

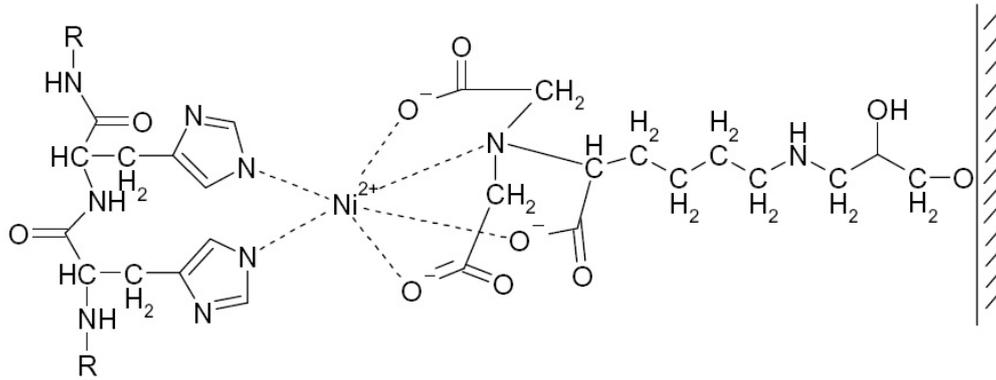


**Abbildung 2.9:** Aufreinigung der Fusionsproteine mit Hilfe eines Maltose-bindenden-Proteins (MBP). Das MBP kann an die Amylose-Harz-Matrix binden. Aufgrund der stärkeren Bindung an Maltose im Vergleich zur Amylose kann das Fusionsprotein nach mehreren Waschschritten mittels Maltose wieder eluiert werden. Das MBP kann später mit einer spezifischen Protease abgeschnitten werden

Diese Aufreinigungsmethode bietet im Vergleich zu anderen Affinitätschromatographischen Aufreinigungen viele Vorteile. Zum einen erhöht der MBP-Ligand dank seiner Hydrophilie die Wasserlöslichkeit des Fusionsproteins. Aber auch die Stabilität wird durch den Liganden erhöht. Die starke MBP-Amylose-Bindung ermöglicht desweiteren eine saubere Aufreinigung der Expressionsprodukte.

### 2.5.3.2 Affinitätschromatographische Aufreinigung über Ni-NTA Agarose

Bei der Aufreinigung über Ni-NTA Agarose macht man sich die Bildung des Komplexes zwischen dem Histidin-Tag eines Proteins und der Ni-NTA-Agarose zu Nutze:



**Abbildung 2.10:** Struktur des Nickel-Chelat-Komplex. R = Protein.

Dieser His-Tag bindet spezifisch an ein Säulenmaterial mit zweiwertigen Nickel-Ionen. Das  $\text{Ni}^{2+}$ -Ion ist in einem oktaedrischen Komplex mit dem Trägermaterial und Wasser-Molekülen gebunden. Diese können durch Histidin verdrängt werden, wodurch ein sehr stabiler Chelat-Komplex entsteht.  $\text{Ni}^{2+}$  ist auf diesen Säulen durch Nitrilotriessigsäure-Reste (Ni-NTA) gebunden und kann mit zwei Histidin-Resten des Proteins im Austausch gegen Wasser interagieren. Diese Spezifität gewährleistet, dass nur das Fusionsprotein an das Säulenmaterial bindet. Die Elution des Fusionsproteins erfolgt mit Imidazol, das das Histidin des Proteins kompetitiv aus dem Chelat-Komplex verdrängt. Alternativ kann auch ein Puffer mit niedrigem pH-Wert verwendet werden, der die Protonierung des Imidazol-Rings im Histidin bewirkt. Der enthaltene Stickstoff kann dann nicht mehr als Koordinationspartner fungieren.

Da Imidazol mit den Histidinen des Proteins um den Ni-Chelat-Komplex konkurriert, kommt es durch das Ansteigen der Imidazolkonzentration in den verschiedenen Wasch- und Elutions-Puffern zuerst zu einer Aufreinigung, und letztlich zur Elution des gewünschten Proteins.

#### 2.5.4 Antikörper-Detektion

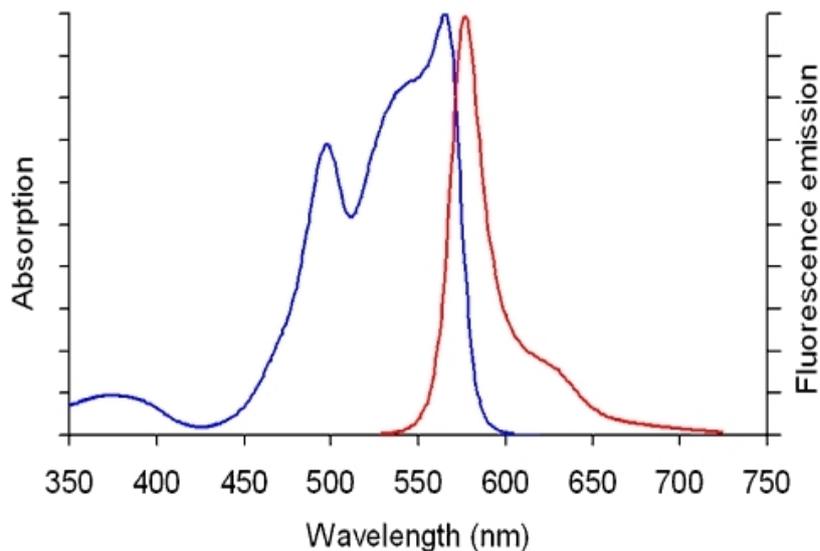
Antikörper, die gegen einen Fusions-Tag gerichtet sind, können zur Detektion von Fusionsproteinen verwendet werden. Genau wie die Bindung eines Substrates an ein Enzym ist auch die Bindung eines Antigens an einen Antikörper reversibel, aber trotzdem sehr stark ( $K_a \sim 10^{10} - 10^{12} 1/M$ ). Sie ist die Summe vieler relativ schwacher, nicht kovalenter Kräfte wie hydrophobe Wechselwirkungen, Wasserstoffbrücken, van-der-Waalskräfte und Ionenbindungen. Sie wirken nur wie die Schlüssel-Schloss-Prinzip, d.h. Antikörper und Antigen müssen sich sehr nahe kommen und "einrasten".

#### 2.5.5 Streptavidin-R-Phycoerythrin

Das R-Streptavidin-Phycoerythrin-Konjugat (SAPE) wird in der Microarray-Technologie zur Generierung des Fluoreszenz-Signals verwendet (Wilson et al. (2002)). Es besteht aus

zwei Untereinheiten, welche für die Bindung an das Zielmolekül und für die Fluoreszenz verantwortlich sind:

- Streptavidin (SA) ist ein aus Streptomyeten (*Streptomyces avidinii*) isoliertes Protein mit einer Größe von 60 KDa. Es besteht aus vier identischen Untereinheiten. Jede dieser Untereinheiten kann mit sehr hoher Affinität ( $K_a \sim 10^{14} - 10^{15} 1/M$ ) ein Biotin-Molekül (244 Da) binden, wobei pro SA-Molekül 4 Biotine parallel gebunden werden können. Diese Bindung ist eine der stärksten nicht kovalenten biologischen Bindungen.
- R-Phycoerythrin ist ein sehr hell leuchtendes Protein mit einem Absorptionsmaximum bei 496, 546 und 565 nm. Es emittiert das Licht bei einer Wellenlänge von  $\sim 578$  nm. Durch seinen sehr hohen Extinktionskoeffizienten  $\epsilon = 2.400.000 L * mol^{-1} * cm^{-1}$  und der hohen Quantenausbeute von 0,98 ist es für Microarrayexperimente bestens geeignet. Das Spektrum ist in Abbildung 2.11 gezeigt. Ein weiterer Vorteil der Phycobiliproteine ist ein großer Stokes-Shift und eine gute Wasserlöslichkeit. Eine kovalente Bindung des Fluorophors an das Protein verhindert zusätzlich das Auftreten unerwünschter Fluoreszenzlöschungen.

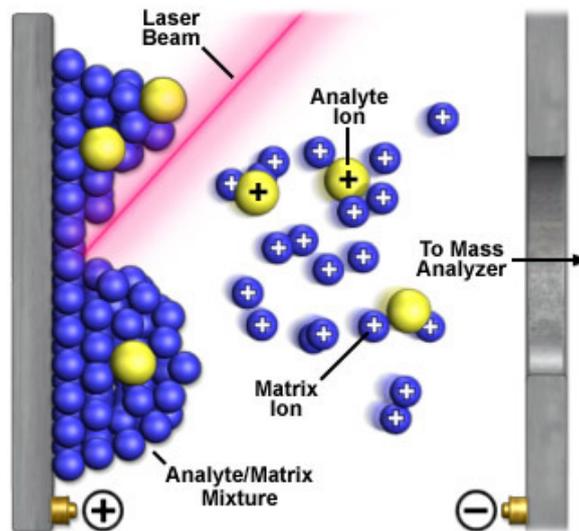


**Abbildung 2.11:** Fluoreszenzspektrum von R-Phycoerythrin. Blaue Linie: Absorption. Rote Linie: Emission. (Quelle: <http://probes.invitrogen.com/media/spectra/801ph75.gif>)

## 2.6 MALDI-TOF Analyse

*Matrix assisted laser desorption and ionisation time of flight* Massenspektrometrie (MALDI-TOF) wurde im Jahre 1988 zeitgleich von Michael Karas, Franz Hillenkamp und Koichi Tanaka entwickelt (Tanaka et al. (1988), Hillenkamp et al. (1991), Karas and Hillenkamp (1988)). Es ist eine Methode zur massenspektrometrische Bestimmung von Biomolekülen (DNA, Peptide, Proteine) oder großen organischen Molekülen. Die Proben werden durch Kokristallisation in eine UV-absorbierende Matrix eingebettet und durch Laserbeschuss in die

Gasphase desorbiert. Als Matrix verwendet man für Proteinanalysen 2,5-Dihydroxybenzoesäure oder Sinapinsäure. Für DNA werden 3-Hydroxy-Picolinsäure oder deren Derivate verwendet. Durch Protonenübertragung von Matrixmolekülen auf die Proteine/Peptide entstehen Protein-/Peptidionen, die in einem elektrischen Feld beschleunigt werden. Eine schematische der Ionisation und Desorption zeigt Abbildung 2.12.



**Abbildung 2.12:** Schema der der Ionisation und Desorption bei der MALDI-TOF Analyse

Durch Messung der Flugzeit (tof) der Ionen auf einer Driftstrecke lässt sich auf die Masse der erzeugten Protein-/Peptidionen zurückschließen. Die Geschwindigkeit der Ionen ist dabei proportional zu eins durch die Wurzel aus Masse/Ladung.

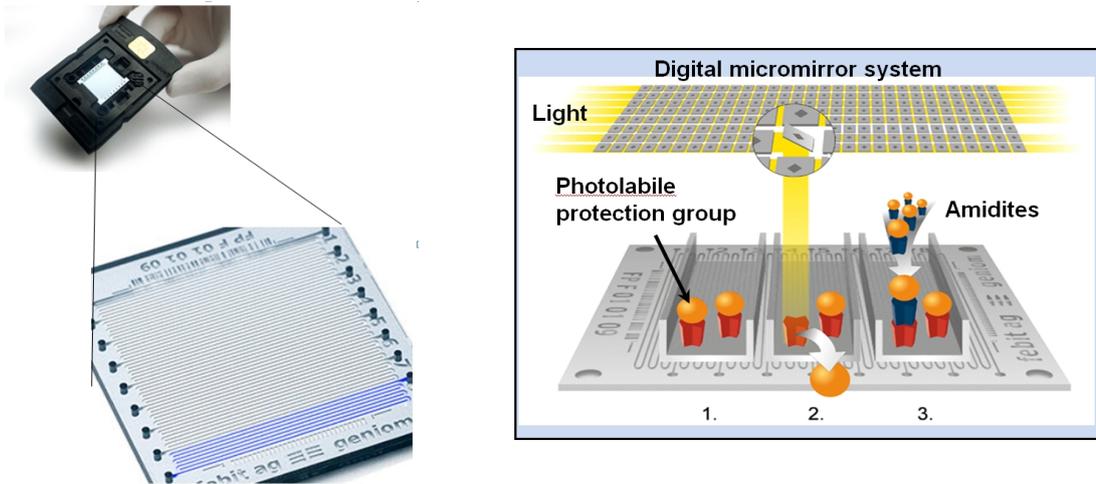
$$\text{tof} \sim \sqrt{m/z}$$

Weiterhin kann durch Fragmentierung einzelner Peptidionen die Aminosäuresequenz des jeweiligen Peptids abgeleitet werden.

## 2.7 DNA-Microarrays mit dem Geniom System

Wie bereits in der Einleitung erwähnt, wurde in dieser Arbeit die Microarray-Plattform der febit biomed GmbH verwendet. Der Chip ist mit einer speziellen Silan-Oberfläche beschichtet und in 8 Kanäle unterteilt, die jeweils mit verschiedenen Proben inkubiert werden können (siehe Abbildung 2.13). In jedem dieser Kanäle lassen sich durch Licht-aktivierte Synthese auf einem Standardarray 6776 beliebige Sequenzen bis zu einer Länge von etwa 60 bp herstellen (Beier and Hoheisel (2000)). Auf einem Array können somit bis zu 54.208 verschiedene DNA-Sequenzen synthetisiert werden. Durch Optimierung des optischen Systems war es gegen Ende der Doktorarbeit möglich, insgesamt 125.000 unterschiedliche Oligonukleotide pro Array zu

erzeugen. Erreicht wurde dies durch eine veränderte Spiegelkonfiguration. Der Spotdurchmesser wurde von 34  $\mu\text{m}$  auf 17  $\mu\text{m}$  reduziert.



**Abbildung 2.13:** Typischer DNA-Microarray (links). Lichtaktivierte DNA-Synthese mit Hilfe eines Mikrospiegelsystems (rechts).

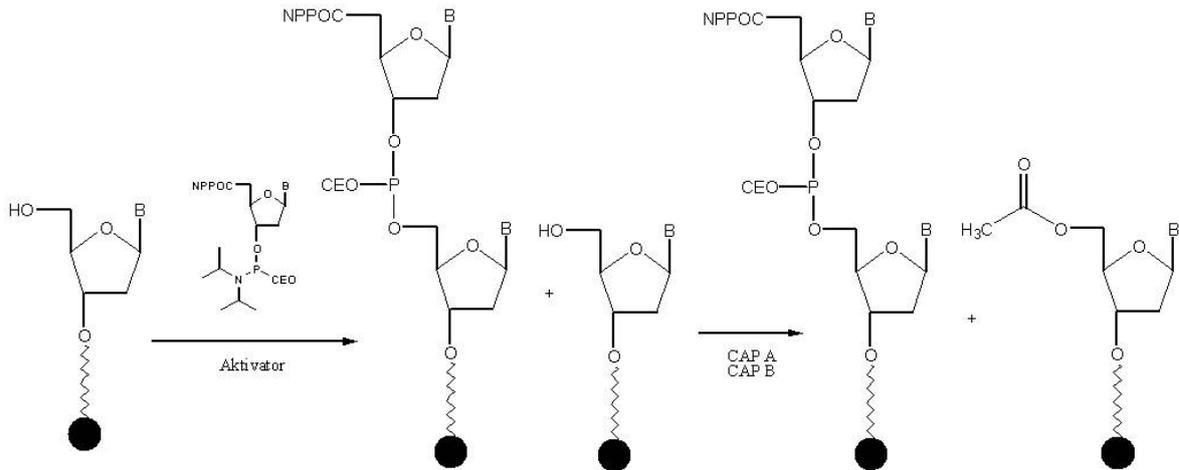
### 2.7.1 Licht-aktivierte DNA-Synthese

Die *in-situ* DNA-Synthese erfolgte über standard Phosphoramiditchemie an der festen Phase, wobei anstelle der DMT Schutzgruppe am 5'-OH Monomere mit einer photolabilen NPPOC Schutzgruppe eingesetzt werden. Die Entschützung erfolgt durch die Verwendung eines Mikrospiegelsystems (siehe Abbildung 2.13). Die Syntheserichtung der DNA erfolgte von 3' → 5'.

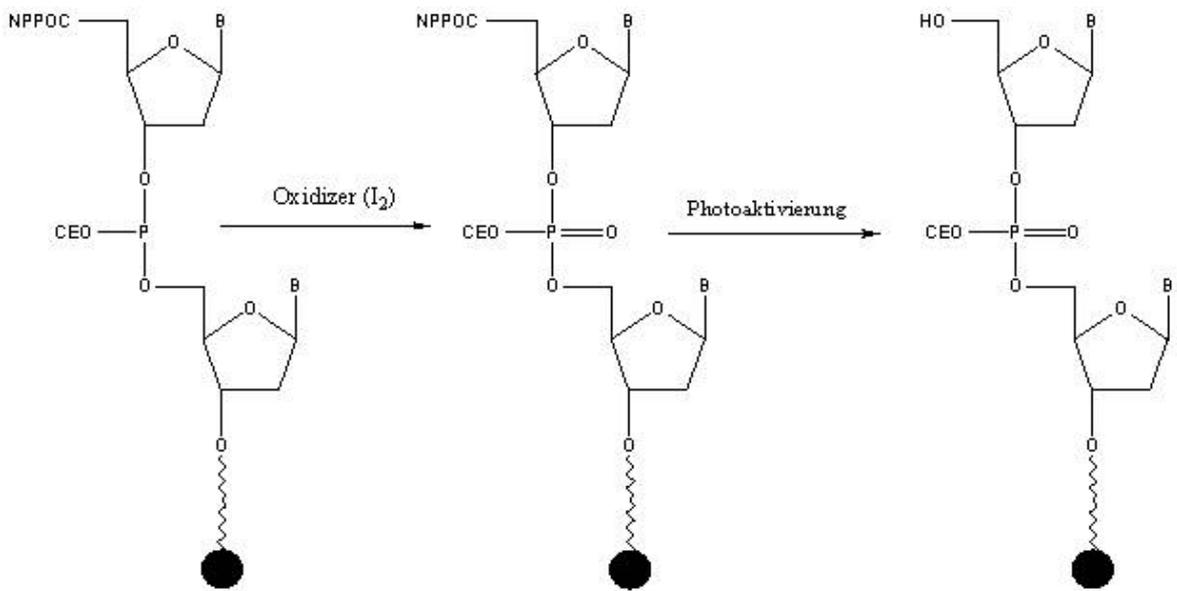
Nachfolgend sind die einzelnen Schritte der DNA-Synthese aufgeführt:

- Entschützung: Entfernen der NPPOC-Schutzgruppe
- Aktivierung der freien 5'-OH-Gruppe
- Kopplung der Phosphoramidite
- Veresterung der freien nicht abreagierten OH-Gruppen
- Oxidation des Phosphatrückgrates von P(III) zu P(V)

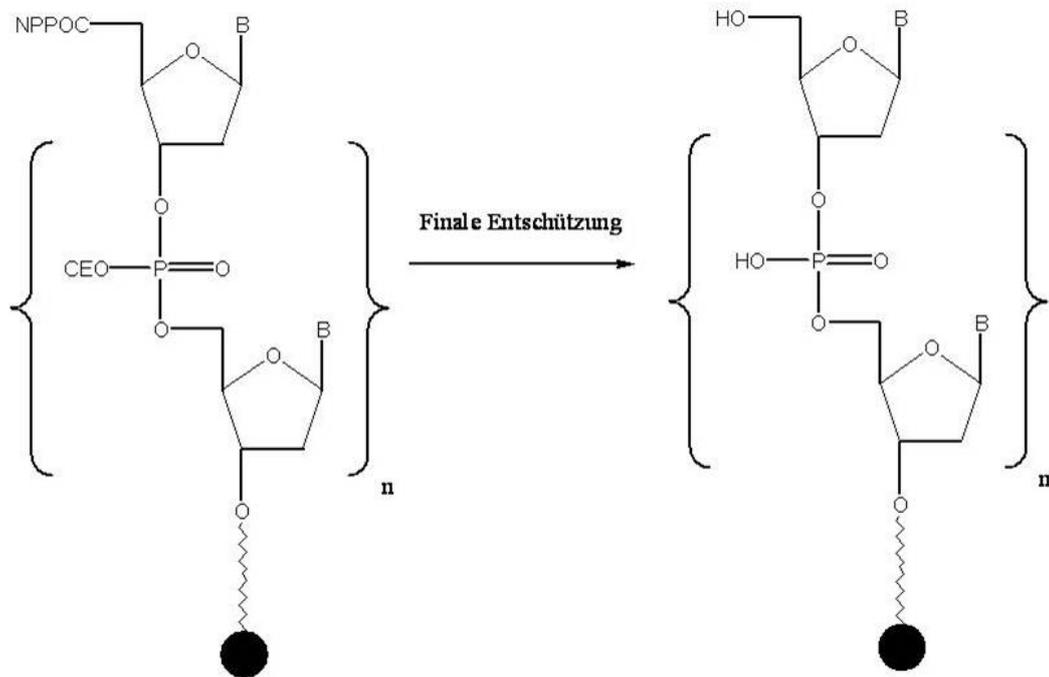
Durch Wiederholen der oben genannten Schritte können somit die verschiedenen Sequenzen auf dem Array synthetisiert werden. Eine finale Entschützung erfolgt am Ende der Synthese mittels 2-Aminoethanol.



**Abbildung 2.14:** Die am 3'-Ende an der Oberfläche gekoppelten Nucleotide werden am freien 5'-Ende mit dem Amidit umgesetzt. Die Ausbeute bei diesem Syntheseschritte beträgt  $> 99\%$ . Die nicht abreagierten Monomere werden verestert, um ungewünschte Syntheseprodukte zu vermeiden. NPPOC: 3'-Nitrophenylpropyloxycarbonyl-Schutzgruppe. CEO:  $\beta$ -Cyano-Ethyl-Schutzgruppe.



**Abbildung 2.15:** Das instabile P(III) wird mit Iod zum stabilen P(V) oxidiert. Darauf folgend wird die photolabile NPPOC Schutzgruppe am 5'-OH abgespalten. Ein neuer Synthesesyklus beginnt.



**Abbildung 2.16:** Am Ende der DNA Synthese erfolgt eine finale Entschützung der Basenschutzgruppen.

Für die Synthese verschiedenen von 60 bp langen Oligonukleotiden sind abhängig von der Sequenz bis zu 180 Synthesyklen notwendig.

## 2.8 Plasmonenresonanzspektroskopie

Die Nutzung von optischen Biosensoren erlaubt das Untersuchen von biologischen Interaktionen auf Oberflächen. Protein-Protein oder Protein-DNA-Wechselwirkungen können hiermit sehr einfach nachgewiesen werden. Es ist möglich diese Interaktionen nicht nur qualitativ, sondern auch quantitativ und ohne vorheriges Markieren nachzuweisen (Joensson et al. (1991)). Die Firma BIAcore® hat ein Biosensor-Instrument entwickelt, das auf der Basis der Plasmonenresonanz solche Untersuchungen möglich macht (Kretschmann (1971)). Die verschiedenen Stadien der Interaktionen (Assoziation, Gleichgewicht, Dissoziation) können dabei in Echtzeit abgebildet werden.

Das Biacore T100-System detektiert Interaktionen von Molekülen ab einem Molekulargewicht von etwa 5000 Dalton. Es arbeitet nach dem Prinzip der Oberflächen- Plasmonresonanz (SPR, engl. surface plasmon resonance). Auf einem Chip wird ein Ligand mit definierter Menge immobilisiert. Diese verhält sich hierbei proportional zu der detektierten Beladungsdichte. Der Chip wird mit polarisiertem Licht bestrahlt, das in einem bestimmten Winkel reflektiert wird und zu einer Abnahme der Intensität des polarisierten Lichtstrahls führt. Ein potentieller Bindungspartner (Analyt, in diesem Falle eine Protein) der immobilisierten DNA wird anschließend in laminarem Pufferfluss über den Liganden geführt und vergrößert in bestimmten Maße

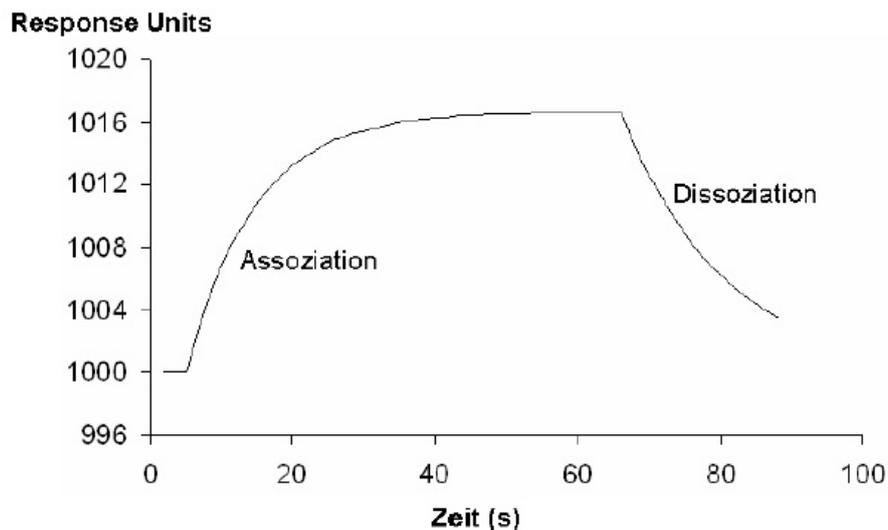
den Austrittswinkel des reflektierten Lichtstrahls. Entsprechend der Menge des gebundenen Analyten ändert sich der Ausfallwinkel des polarisierten Lichts. Aus der Differenz ergibt sich eine Signaländerung und im zeitlichen Verlauf ein Sensorgramm.

Aus den Sensorgrammen der verschiedenen Konzentrationen lassen sich unter Verwendung der Software BIAevaluation 3.0 die Assoziations- bzw. Dissoziationskonstanten einer jeweiligen Protein-DNA-Wechselwirkung berechnen.

### 2.8.1 Messprinzip

Die Plasmonenresonanz beruht auf dem quantenmechanischen Phänomen der optischen Schwingungsanregung von Elektronen in dünnen Metallfilmen. Durch die Wechselwirkung der Elektronen mit Photonen entstehen Schwingungen, sog. Plasmonen, die sich in der Metallschicht wellenförmig ausbreiten können (Johnsson et al. (1991)). Der Aufbau des Biacore Systems besteht aus einem miniaturisierten Flussinjektionssystem und einem integrierten optischen Sensor, der die Oberflächenplasmonenresonanz zur Detektion nutzt.

Mit Hilfe dieser Plasmonenresonanz lässt sich der Brechungsindex einer dünnen Schicht über dem Chip bestimmen. Der Sensorchip an sich besteht aus einem Glasträger mit einem ca. 50 nm dünnen Goldfilm, an dessen Oberfläche eine Matrix aus carboxymethyliertem Dextran gekoppelt ist. Auf dieser Matrix ist eine Streptavidinschicht immobilisiert, die eine Ankopplung der biotinylierten Liganden erlaubt. Während der Messung wird der im Laufpuffer gelöste Analyt über den an den Sensorchip gekoppelten Liganden gegeben. Der Brechungsindex hängt dann von der Protein/DNA-Masse in der jeweiligen Schicht ab. Die Zunahme der verschiedenen Massen (Ligand und Analyt) kann somit optisch anhand der aufgenommenen Sensorgramme verfolgt werden, in dem Signalhöhe und Zeit angezeigt werden (siehe Abbildung 4.40).



**Abbildung 2.17:** Sensorgramm einer Biacore-Messung. Auftragung: Messsignal als “*Response Units*” (Bindungseinheiten – RU) gegenüber der Zeit. Assoziationsphase, Sättigungsbereich, sowie die Dissoziationsphase sind zu sehen.

### 2.8.2 Kinetik und Auswertung

Die Adsorption des Proteines [P] an der auf der Oberfläche fixierten DNA [D] wird unter konstantem Fluss als eine Reaktion pseudo-erster Ordnung angenommen. Da die Konzentration der DNA als konstant anzusehen ist, kann die Interaktion des Proteines mit der DNA durch folgende Gleichungen (1) beschrieben werden.



Daraus ergibt sich folgende Gleichung (2).

$$(2) \quad K_D = \frac{[P] * [D]}{[PD]}$$

Unter Berücksichtigung der Assoziationskonstanten  $K_a$  und der Dissoziationskonstanten  $K_b$  folgen daraus Gleichung (3) und (4).



$$(4) \quad \frac{d[PD]}{dt} = K_a * [P] * [D] - K_d * [PD]$$

Die Assoziations- und Dissoziationsraten  $k_a$  bzw  $k_d$  können durch Experimente mit verschiedenen Protein-Konzentrationen (C) ermittelt werden. Durch Messung der Bindungsrate  $dR/dt$  als Funktion des SPR-Signals R kann Gleichung (4) in Gleichung (5) umgeschrieben werden, wobei  $R_{max}$  das maximale Signal der Protein-Bindung ist.

$$(5) \quad \frac{d[PD]}{dt} = K_a * C * R_{max} - (K_a * C + K_d) * R$$

Für eine Reihe von Proteinkonzentrationen liefert die Auftragung von  $dR/dt$  auf R den Term  $k_a C + k_d$  als Steigung, der als  $k_s$ -Wert bezeichnet wird. Auftragung des  $k_s$ -Wertes über der Konzentration C liefert eine Gerade, deren Steigung dem  $k_a$ -Wert entspricht. Der  $K_D$ -Wert kann durch Gleichung (6) und (7) ermittelt werden:

$$(6) \quad \ln \frac{R_0}{R_t} = k_d * (t - t_0)$$

$R_t$  ist das Signal zur Zeit  $t$ ,  $R_0$  das Startsignal zur Zeit  $t_0$ . Durch Auftragung des Terms  $\ln(R_0/R_t)$  auf  $(t - t_0)$  wird eine Gerade mit der Steigung  $k_d$  erhalten. Die Gleichgewichtskonstanten der Assoziation und Dissoziation ergeben sich als Quotienten der Reaktionsraten.

$$(7) \quad K_A = k_a/k_d \text{ bzw. } K_D = k_d/k_a = 1/K_A$$

Die Ermittlung der kinetischen Parameter erfolgte durch die vom Hersteller mitgelieferten Software BIAevaluation.

### 2.8.3 Datenauswertung mit der BIAevaluation®-Software

Die mit BIAcontrol® aufgenommenen Kurven können in das Programm BIAevaluation® importiert werden. Injektionsstart und -stopp als auch die Bereiche der Assoziation und Dissoziation können manuell festgelegt werden, um die richtige Berechnung der Parameter zu gewährleisten. Den einzelnen Kurven können deren Konzentration und das entsprechende Bindungsmodell (z.B. 1:1 Langmuir, 1:1 Langmuir mit Drift der Basislinie, Reaktion mit mehreren Analyten oder mehreren Liganden) zugeordnet und die Berechnung initiiert werden. Sie kann sowohl global (mehrere Kurven – „Global Fit“), als auch lokal (jede Kurve separat – „Local Fit“) für verschiedene Parameter erfolgen.

## 2.9 Quarzmikrowaage - QCMD

Die QCMD-Technik (*Quartz Crystal Microbalance and Dissipation*) basiert hauptsächlich auf der Arbeit von Rodahl et al (Rodahl (1995)). Dabei wird der von Pierre and Jacques Curie im Jahre 1880 entdeckte Piezoelektrische Effekt ausgenutzt. Wenn man an eine Elektrode eine elektrische Spannung anlegt, kann ein Quarzkristall in Schwingung angeregt werden. Somit kann man über den Energieabfall einer gedämpften Schwingung die Masse und die Dissipation bestimmter Materialien auf dem Kristall bestimmen. Schon im Jahre 1959 entdeckte Sauerbrey, dass Frequenzänderung eines Quarzkristalls proportional zu der Masse dieses Kristalls ist (Sauerbrey (1959)). Zu diesem Zeitpunkt wurden jedoch noch alle Versuche im Vakuum oder in der Gasphase durchgeführt.

Erst Nomura et al. erweiterten im Jahre 1980 die Technik auch auf Flüssigkeiten (Nomura and Hattori (1980)). Rodahl erforschte später während seiner Doktorarbeit nicht nur die Resonanzfrequenzen der Kristalle sondern auch die Dissipation. Diese Kombination ermöglichte viele neue Anwendungen. Man konnte nicht nur die Masse der Materialien auf dem Kristall bestimmen, es war auch möglich die Elastizität dieser Oberflächen mit Hilfe der Dissipation zu bestimmen. Damit konnten auch biologische System wie z.B. Zelladhäsion untersucht werden (Hook (1997)). Eines der Hauptanwendungsgebiete ist jedoch die Polymerforschung (Garbassi et al. (1998)).

### 2.9.1 Massenbestimmung nach Sauerbrey

Sauerbrey schlug im Jahre 1959 ein einfaches physikalisches Modell für die beobachtete Proportionalität zwischen der Massenzunahme und der induzierten Frequenzänderung (Sauerbrey (1959)) vor. Er nahm an, dass kleine Massenänderungen des Film als Massenänderung des gesamten Quarzkristalls angenommen werden können.

Die induzierten Schwingungen führen zum Aufbau eines elektrischen Feldes im Kristall. Nach Abschalten der Wechselfeldspannung tritt eine Dämpfung dieses elektrischen Feldes auf, welche gemessen werden kann. Die Masse des adsorbierten Films kann nach der Sauerbrey Gleichung (Gleichung 8) berechnet werden:

$$(8) \quad \Delta m = -\frac{v_q * \rho_q}{2f_0^2} * \frac{\Delta f_n}{n} = -C \frac{\Delta f_n}{n}$$

$n$  = n-ter Oberton (ungerade: 1, 3, 5, ...)

$v_q$  = Schallgeschwindigkeit im Kristall (3340 m/s)

$\rho_q$  = Dichte des Quarzes (2,65 g/cm<sup>3</sup>)

$f_0$  = Eigenfrequenz des Quarzes ( $f_0 = v_q / 2 t_q$ , 5 MHz)

$\Delta f_n$  = Frequenzänderung des Obertons  $n$

$t_q$  = Dicke des Kristalls

$C$  = Massensensitivitätskonstante (17,7 ng/cm<sup>2</sup> \* Hz)

Die Gleichung ist für starre Filme gültig. Zusätzlich muss die Masse des Filmes klein gegenüber der Masse des Gesamtkristalls sein.

### 2.9.2 Der Dissipations-Faktor

Der Dissipations-Faktor ist die Inverse des bekannteren Q-Faktors (Gleichung (9)).

$$(9) \quad D = \frac{1}{Q} = \frac{E_{dissipated}}{2\pi E_{stored}}$$

Er beschreibt die viskoelastischen Eigenschaften eines adsorbierten Materials. Handelt es sich um eine flache und rigide Masse, so schwingt der Kristall nur langsam ab. Dagegen zeigen weiche und viskose Massen einen hohen Dissipations-Wert, sodass die Kristallschwingung deutlich schneller abklingt. Ein Beispiel hierfür wäre die Wassereinlagerung in einem viskosen Film.

## 3 Material und Methoden

### 3.1 Material

#### 3.1.1 Laborausstattung

NAME	HERSTELLER
Bioanalyzer 2100	Agilent, Böblingen, Deutschland
Elektrophoresekammer /Agarosegel Model D3	OQI Separation Systems, Porthmouth Holland
Erlenmeyer-Kolben	Schott, Mainz, Deutschland
Glasflaschen	Schott, Mainz, Deutschland
Hoefer TE 70 (Semi dry Wester-Blot)	Amersham Bioscience, Piscataway, USA
Microarray-Plattform, Geniom2x	febit, Heidelberg Deutschland
Mini-Protean®3 Elektrophorese Kammer und Zubehör für Gele	BioRad, Waltham, USA
Netzteil	Feather Volt 500, Stratagene, Deutschland
Ofen	Haereus, Hanau, Deutschland
PCR-Maschine PTC200	M J Research BioRad, Waltham, USA
pH-Meter	MP 230, Mettler Toledo, Deutschland
Photometer (UV/Vis)	Ultrospec 2000, Pharmacia Biotech, Freiburg Deutschland
ND 1000, NanoDrop, Spectrophotometer	NanoDrop, Wilmington, DE USA
Pipetten	Pipetboy acu, Integra Bioscience
Pipetten (0,5 µl-10 µl)	Eppendorf Research, Hamburg Deutschland
Pipetten (20/200/1000 µl)	Pipetman, Gilson, Middleton, USA
Rotoren, SL-50T, Sorvall	Eppendorf, Hamburg Deutschland
Rotoren, A-4-62, Sorvall	Eppendorf, Hamburg Deutschland
Scanner Scan Array 3000	Packard, Billerica, MA, USA
Scanner Scan Array 3000	Packard, Billerica, MA, USA
Schüttler	GFL 3020, GFL
Slidebooster	Slidebooster, Advalytix, München Deutschland
Spotter Nano Plotter	Gesim, Großerkmannsdorf Deutschland
Thermomixer 5436	Eppendorf, Hamburg Deutschland
Vortexer Reax 2000	Heidolph, Kehlheim, Deutschland

Waagen, 440-45	Kern; Balingen Deutschland
Waagen, 440-53	Kern; Balingen Deutschland
Waagen, 434, Kern	Balingen Deutschland
Zentrifuge	5810R, Eppendorf, Hamburg Deutschland
Zentrifuge	5415D, Eppendorf, Hamburg Deutschland

**Tabelle 3.1:** Laborausstattung**3.1.2 Verbrauchsmaterial**

NAME	HERSTELLER
384er Mikrotiterplatte	Genetix, New Milton Hampshire, UK
96er Filterplatte MADVN 6510	Millipore, Bedford, USA
96er Mikrotiterplatte	Nunc, Roskilde, Dänemark
Einmalküvetten	UVette, Eppendorf, Hamburg Deutschland
Eppendorf Reaktionsgefäße	Safe Lock Tubes 1,5ml und 2ml, Eppendorf, Hamburg Deutschland
Falcon Röhrchen	15ml und 50ml Falcons, BD
Geniom - Biochip	Sigma, Seelze, Deutschland
Handschuhe	Latex, Blossom Mexpo, Hayward, USA
Handschuhe	Nitril, Microflex, Wien, Österreich
His Trap FF Säulen (1 ml)	GE Healthcare, Uppsala Schweden
Kryoröhrchen 2 ml	Nunc, Wiesbaden, Deutschland
Lazy-L-Spreaders	Sigma-Aldrich, St.Louis, USA
MILLEX®GV-Filter, 0,22 µm	Millipore, Bedford, USA
Nitrozellulose Membran	Sigma, Seelze, Deutschland
Objektträger Starfrost	Knittel Gläser, Braunschweig, Deutschland
Parafilm	PM 996, Pechiney Plastic packaging
Pasteurpipetten 150mm	Lager
Polypropylen-Säulen, 5 ml	Qiagen, Hilden, Deutschland
Präzisionstücher	Kimberly-Clark, Reigate, UK
Protein LoBind Tubes 0,5 ml	neoLab, Heidelberg, Deutschland
Sterilfilter 500 ml	Nalgene, Rochester, USA
Transferpapier	Sartorius Neolab, Heidelberg, Deutschland
Zellkultur-Petrischalen 96x20mm	Lager
Zentrifugenröhren Amicon Ultra 10 KDa (UFC801024)	Millipore, Bedford, USA
2,5-Dimethylformamid (DMF)	SDS, Peypin, Frankreich
2-Propanol	Fluka Sigma-Aldrich, Seelze, Deutschland
Acrylamid/Bisacrylamid	Roth, Karlsruhe, Deutschland
Agar	Roth, Karlsruhe, Deutschland
Agarose	Invitrogen, Carlsbad, USA

Ammoniumpersulfat (APS)	Roth, Karlsruhe, Deutschland
Ampicillin (Natrium Salz)	Genaxxon, München, Deutschland
Amylose Resin	NEB, Frankfurt, Deutschland
Braun Wasser	Braun, Melsungen, Deutschland
Brij58	Sigma-Aldrich, St.Louis, USA
BSA (Rinderalbumin)	Sigma-Aldrich, Schnelldorf Deutschland
Chloramphenicol	Genaxxon, München, Deutschland
Colorburst Elektrophorese Marker (Protein)	Sigma-Aldrich, Schnelldorf Deutschland
<i>complete mini EDTA free protease Inhibitor</i>	Roche, Mannheim, Deutschland
Dikaliumhydrogenphosphat ( $K_2HPO_4$ )	Merck, Darmstadt, Deutschland
Dithiothreitol (DTT)	Invitrogen, Carlsbad, USA
DNA-Ladepuffer Loading Dye Solution	Fermentas, St. Leon Roth, Deutschland
DNA-Ladepuffer Gel-Pilot-Loading Dye	Qiagen, Hilden, Deutschland
DNA-Marker: Gen Ladder 100 bp + 1,5 kb + 2 kb	Genaxxon, München, Deutschland
DNA-Marker: Gene Ruler™ 100bp Ladder6	Fermentas, St. Leon Roth, Deutschland
DNA-Marker: Low Molecular Weight Marker	NEB, Frankfurt, Deutschland
DNA-Marker: Mass Ruler™ DNA-Ladder- Mix	Fermentas, St. Leon Roth, Deutschland
DNA-Marker: $\lambda$ -DNA/PstI-Marker	Fermentas, St. Leon Roth, Deutschland
Essigsäure	Fluka Sigma-Aldrich, Seelze, Deutschland
Ethanol, absolut	Riedel de Haën Sigma, Seelze, Deutschland
Ethidiumbromid	AppliChem, Darmstadt, Deutschland
Ethylendiamintetraessigsäure (EDTA)	Sigma-Aldrich, St.Louis, USA
Formamid	Merck, Darmstadt, Deutschland
Glucose D (+)	AppliChem, Darmstadt, Deutschland
Glycerol	J.T.Baker, Deventer, Holland
Glyzin	Gerbu, Gaiberg, Deutschland
Hefeextrakt	Gerbu, Gaiberg, Deutschland
Hepes	Roth, Karlsruhe, Deutschland
Imidazol	Roth, Karlsruhe Deutschland
Incidin Extra	Ecolab, Wien, Österreich
Isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG)	Roth, Karlsruhe, Deutschland
Kaliumchlorid 1 M	Roth, Karlsruhe, Deutschland
Kanamycin	Genaxxon, München, Deutschland
Ligase Puffer	Roche, Mannheim, Deutschland
Magnesiumdichlorid	Fluka Sigma-Aldrich, Seelze, Deutschland
Magnesiumdichlorid ( $MgCl_2$ ) 25 mM	Genaxxon, München, Deutschland
Magnesiumsulfat ( $MgSO_4$ )	Merck, Darmstadt, Deutschland
Maltose	Serva, Heidelberg, Deutschland
Methanol	VWR, Darmstadt, Deutschland
Milchpulver	Fluka Chemie, Steinheim, Deutschland

Natriumchlorid	Riedel de Haën Sigma, Seelze, Deutschland
Natriumdihydrogenphosphat ( $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$ )	Roth, Karlsruhe, Deutschland
Natriumdodecylsulfat (SDS)	Riedel de Haën Sigma, Seelze, Deutschland
Natronlauge (NaOH)1 M	KMF, Lohmar, Deutschland
Natronlauge (NaOH)10 M	Riedel de Haën Sigma, Seelze, Deutschland
Ni-NTA-Agarose	Qiagen, Hilden, Deutschland
PBS 1x Gibco	Invitrogen, Grand Island N.Y., USA
PCR-Puffer E + $\text{MgCl}_2$ 10x	Genaxxon, München, Deutschland
Pfu-Polymerase-Puffer + 20 mM $\text{MgSO}_4$ 10x	Promega, Madison, USA
Protein-Marker: Broad Range	NEB, Frankfurt Deutschland
Roti-Blue (5x Konzentrat)	Roth, Karlsruhe Deutschland
Saccharose	Merck, Darmstadt, Deutschland
Salzsäure (HCl) 1N	VWR, Darmstadt, Deutschland
Salzsäure 37%	VWR, Darmstadt, Deutschland
Tetracyclin Hydrochlorid	Böhringer Mannheim, Deutschland
TEMED	Roth, Karlsruhe, Deutschland
Tetracyclin	Genaxxon, München, Deutschland
Triton X 100	Gerbu, Gaiberg, Deutschland
Trizma®-base	Sigma-Aldrich, St.Louis, USA
Trizma®-hydrochloride (Tris-HCl)	Sigma-Aldrich, St.Louis, USA
Trypton/Pepton aus Casein pankreatisch ver- daut	Roth, Karlsruhe, Deutschland
Tween 20	Sigma-Aldrich, St.Louis, USA
X-Gal	Roth, Karlsruhe Deutschland
Zinksulfat ( $\text{ZnSO}_4$ )	Serva, Heidelberg, Deutschland

Tabelle 3.2: Verbrauchsmaterial

### 3.1.3 Enzyme & Proteasen

BEZEICHNUNG	HERSTELLER
Antarctic Phosphatase (5000 U/ml)	NEB, Frankfurt Deutschland
Benzonase® Nuclease (2,5 kU)	Novagen, Darmstadt, Deutschland
BspHI (500 U)	NEB, Frankfurt Deutschland
HindIII (10000U)	NEB, Frankfurt Deutschland
Lysozym	Roth, Karlsruhe, Deutschland
NcoI (1000 U)	NEB, Frankfurt Deutschland
Pfu DNA Polymerase (3U/µl)	Promega, Madison, USA
T4 Ligase	Roche, Mannheim, Deutschland
Taq DNA Polymerase (etwa 5U/µl)	Im Labor hergestellt
Taq Polymerase (5 U/µl) lot: 11873183	Qiagen, Hilden, Deutschland
Thermosequenase	Amersham Bioscience, Piscataway, USA

**Tabelle 3.3:** Enzyme und Proteasen

### 3.1.4 Kits

BEZEICHNUNG	HERSTELLER
BCA™ Protein Assay Kit	Pierce, Rockford, USA
Bioanalyzer	Agilent, Böblingen, Deutschland
Electrophoretic Mobility Shift Assay (EMSA)	Invitrogen, Karlsruhe Deutschland
peqGOLD GelPure™ Kit	peqLab, Erlangen, Deutschland
Qiagen QIAprep Spin Miniprep Kit	Qiagen, Hilden, Deutschland
Qiagen QIAquick Gel Extraction Kit	Qiagen, Hilden, Deutschland
Qiagen QIAquick PCR Purification Kit	Qiagen, Hilden, Deutschland
Synthese-Kit für Geniom 2x	Sigma, Seelze, Deutschland

**Tabelle 3.4:** Kommerziell erhältliche Kits

### 3.1.5 Puffer und Medien

NAME	HERSTELLER
2-YT Medium	16 g Trypton/Pepton, 10 g Hefeextrakt, 5 g NaCl, ad 1 l mit H <sub>2</sub> O, pH 7,2 mit 1 M NaOH
Binde Puffer 5x	0,7 M KCl, 25 mM NaCl, 5 mM K <sub>2</sub> HPO <sub>4</sub> , 10 mM MgSO <sub>4</sub> , 100 mM Hepes, 350 µM EDTA, 5 µM ZnSO <sub>4</sub> , ad ddH <sub>2</sub> O
Block-Puffer (Western blot)	3% BSA, 0,1% Tween 20, ad 50 ml 1xTBS

EB Puffer	10 mM Tris-Cl, pH 8,5
Elutions-Puffer	Siehe Lyse-Puffer, 250 mM Imidazol
EMCS-Aktivatorlösung	20 mM EMCS in 10% diisopropylethylamin in DMF
Ethidiumbromidbad	0,5 µg/ml Endkonzentration
Ethylendiamintetraessigsäure (EDTA) 0,5 M pH 8,0	93 g EDTA, ad 500 ml ddH <sub>2</sub> O, mit 10 M NaOH auf pH 8,0
Laemmli Puffer	30,1 g Tris Base, 144,2 g Glycin, 50 ml SDS (20%), ad 1 l dH <sub>2</sub> O
LB-Agar	LB-Medium + 1,5% (w/v) Agar
LB-Medium	10 g Trypton/Pepton, 5 g Hefeextrakt, 10 g NaCl, ad 1 l ddH <sub>2</sub> O mit NaOH auf pH 7,2
Lösung 1	5µl Lysozym (10mg/ml), 12,5 µl Brij 58 (1%), 7,5µl Lyse Puffer A
Lösung 2	0,3 µl 1 M <i>MgCl</i> <sub>2</sub> , 0,1 µl Benzonase (25 U/µl), 24,6 µl 50 mM Tris-HCl pH 8,0
Lyse Puffer A	50 mM Tris-HCl pH 8,0, 0,3 M NaCl, 0,1mM EDTA
Lyse-Puffer	50 mM <i>NaH</i> <sub>2</sub> <i>PO</i> <sub>4</sub> x H <sub>2</sub> O, 300 mM NaCl, 10 mM Imidazol, ad 800 ml ddH <sub>2</sub> O, auf pH 8,0 mit NaOH (10 M/1 M), ad 1 l ddH <sub>2</sub> O
Magnesiumsulfat ( <i>MgSO</i> <sub>4</sub> ) 1M	24,67 g <i>MgSO</i> <sub>4</sub> * 7H <sub>2</sub> O, ad 100 ml mit ddH <sub>2</sub> O
PBS 10x	80 g NaCl, 2 g KCl, 26,8 g <i>Na</i> <sub>2</sub> <i>HPO</i> <sub>4</sub> , 2,4 g <i>KH</i> <sub>2</sub> <i>PO</i> <sub>4</sub> , auf 1l mit ddH <sub>2</sub> O und mit HCl auf pH 7,4
Säulen Puffer	20 ml Tris-HCl 1M (pH 7,4), 11,7 g NaCl, 2 ml EDTA 0,5 M, ad 1 l ddH <sub>2</sub> O, vor Gebrauch Zugabe von 1 mM DTT (Endkonz.)
TB Lösung	10 mM Pipes, 55 mM <i>MnCl</i> <sub>2</sub> x 4H <sub>2</sub> O, 75 mM <i>CaCl</i> <sub>2</sub> x 2 H <sub>2</sub> O, 250 mM KCl, pH 6,7 mit 1 M KOH
TBE 10x	108 g Tris, 55 g Borsäure, 40 ml 0,5 M <i>Na</i> <sub>2</sub> <i>EDTA</i> (pH8,0) mit ddH <sub>2</sub> O auf 1l
TBS (für Western)	10 mM Tris HCl, 150 mM NaCl
TBS 10x	50 mM Tris, 150 mM NaCl mit HCl auf pH 7,5, ad 1 l ddH <sub>2</sub> O
TBS-Tween 20 mM (TBS/T)	Tris-HCl, 500 mM NaCl, 0,05% Tween20, ddH <sub>2</sub> O
TBS-Tween-Triton (TBS/TT)	20 mM Tris-HCl, 500 mM NaCl, 0,05% Tween20, 0,2% TritonX100,ddH <sub>2</sub> O

Transferpuffer	150 mM Glycin, 25 mM Tris Base, 10% Ethanol, ad 1 l ddH <sub>2</sub> O
Tris/HCl 1 M	pH 6,8 / 7,4 / 8,0 121,14 g Tris-OH ad 1 l ddH <sub>2</sub> O, mit HCl 10 M auf pH
Wasch-Puffer	Siehe Lyse-Puffer, 20 mM Imidazol
Waschpuffer A	50 mM Tris-HCl pH 8,0, 0,3 mM NaCl, 20 mM Imidazol
X-Gal 4%	500 mg X-Gal + 12,5 ml DMF

**Tabelle 3.5:** Puffer und Medien

### 3.1.6 Oligonukleotide, Primer und Vektoren

GLI2; 5'-THIOL	SEQUENZ
consensus Sequenz	5'-SH-tttttttt attaat GACCACCCA aa-tactttttgtatt tgggtggc attaat-3'
1MM +	5'-SH-tttttttt attaat CACCACCCA aa-tactttttgtatt tgggtggc attaat-3'
1MM -	5'-SH-tttttttt attaat GATCACCCA aa-tactttttgtatt tgggtggc attaat-3'

**Tabelle 3.6:** Oligonukleotide, die am 5'-Ende mit einem Thiol modifiziert sind.

GLI2; 5'-BIOTIN	SEQUENZ
Gli2 consensus Sequenz	5'-Biotin-tttttttt attaat GACCACCCA aa-tactttttgtatt tgggtggc attaat-3'
Gli2 1MM +	5'-Biotin-tttttttt attaat CACCACCCA aa-tactttttgtatt tgggtggc attaat-3'
Gli2 1MM -	5'-Biotin-tttttttt attaat GATCACCCA aa-tactttttgtatt tgggtggc attaat-3'

**Tabelle 3.7:** Oligonukleotide, die am 5'-Ende mit einem Biotin modifiziert sind.

NAME	SEQUENZ
pMal <i>forward</i>	5'- GGT CGT CAG ACT GGA TGA AGC C- 3'
pMal <i>reverse</i>	5'- TTG TAA AAC GAC GGC CAG TGC C- 3'
pQE Trisystem Strep1 HisC Gli2 <i>forward</i>	5'-TTA GGA GCA TCA TGA AGC AGG AGG CTG AGG TGG TCA TCT ACG AG- 3'
pQE Trisystem Strep1 HisC Gli2 <i>reverse</i>	5'-TGC TCC TAA AAG CTT CAG CCC GAG CTC TCG GTC TTG A-3'
pQE Trisystem <i>forward</i>	5'-ATA CTG TAA ATT ACA TTT TAT TTA CAA TCA AAG GAG ATA T-3'
pQE Trisystem <i>reverse</i>	5'-GCA CGT GTA TAC AGC TGT GCG GC-3'

Tabelle 3.8: Primer

3.1.7 *E.coli*- Bakterienstämme

BAKTERIENSTAMM	EIGENSCHAFTEN
XL1-Blue (Stratagen)	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Te- tr)]
Rosetta2(DE3)pLysS (Novagen)	F <sup>-</sup> ompT hsdS <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) gal dcm λ (DE3 [la- cI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysS- RARE (CamR)
Rosetta-gami B(DE3)pLysS	F <sup>-</sup> ompT hsdS <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) gal dcm lacY1 ahpC (DE3) gor522:Tn10 trxB pLysSRARE (Cam <sup>R</sup> , Kan <sup>R</sup> , Tet <sup>R</sup> )
M15 (pREP4) (Qiagen)	NaIS, StrS, RifS, Thi <sup>-</sup> ,Lac <sup>-</sup> , Ara <sup>+</sup> , Gal <sup>+</sup> , Mtl <sup>-</sup> , F <sup>-</sup> , RecA <sup>+</sup> , Uvr <sup>+</sup> , Lon <sup>+</sup>
SG13009 (pREP4) (Qiagen)	NaIS, StrS, RifS, Thi <sup>-</sup> ,Lac <sup>-</sup> , Ara <sup>+</sup> , Gal <sup>+</sup> , Mtl <sup>-</sup> , F <sup>-</sup> , RecA <sup>+</sup> , Uvr <sup>+</sup> , Lon <sup>+</sup>

Tabelle 3.9: *E.coli*- Bakterienstämme

### 3.1.8 Antikörper

ANTIKÖRPER	FIRMA
Penta-His-Antikörper, Maus, Material-Nr 34660	Qiagen, Hilden, Deutschland
Anti Maus IgG Alkalische Phosphatase Konjugat, Material-Nr. A3562	Sigma-Aldrich, Schnelldorf Deutschland
Sigma Monoklonaler Anti-c-Myc-Biotin Antikörper, Maus, Klon 9E10	Sigma Aldrich, Schnelldorf Deutschland
Anti Biotin Alkalische Phosphatase Konjugat	Sigma-Aldrich, Schnelldorf Deutschland
Sigma Monoklonaler Anti-c-Myc-Cy3 Antikörper, Maus, Klon 9E10	Sigma Aldrich, Schnelldorf Deutschland
Sigma Monoklonaler Anti-c-Myc-Alexa555 Antikörper, Maus, Klon 9E10	Sigma Aldrich, Schnelldorf Deutschland

**Tabelle 3.10:** Verwendete Antikörper

## 3.2 Mikrobiologische Methoden

### 3.2.1 Herstellung kompetenter *E. coli* - Bakterien

Durch die Behandlung von Bakterienzellen mit  $\text{CaCl}_2$  steht eine recht einfache Methode zur Verfügung, mit der kompetente Zellen selbst herzustellen sind.

#### Protokoll nach Inoue et al. (1990):

- 39 ml SOB-Medium
- 0,4 ml 1 M  $\text{MgCl}_2$
- 0,4 ml 2 M Glucose
- 0,1 ml Kanamycin (10 mg/ml)
- 10  $\mu\text{l}$  Dauerkultur von M15 / SG13009- Zellen

Bei einer OD 600 zwischen 0,2 und 0,4 wurden die Zellen bei 3000 g und  $4^\circ\text{C}$  für 3 min pelletiert. Im Anschluss an die Abnahme des Überstandes wurden das halbe Volumen der Ausgangslösung an kalter TB-Lösung zugegeben. Nach einem erneuten Zentrifugationsschritt unter den selben Bedingungen wurden die Zellen nun in 1/20 des Ausgangsvolumens TB-Lösung mit 20% Glycerol resuspendiert und anschließend im Ethanol-Stickstoffbad in 100  $\mu\text{l}$  Aliquots eingefroren.

### 3.2.2 Plasmid-DNA Transformation in *E. coli* - Bakterien

Eine Möglichkeit Plasmide in chemisch kompetente Zellen einzuführen, ist ein abrupter Temperaturwechsel, die sogenannte Hitzeschock-Methode. Die auf Eis aufgetauten Zellen wurden mit der aufzunehmenden DNA versetzt und sehr vorsichtig durchmischt. Einer 30 minütigen Inkubation auf Eis folgten 90 Sekunden bei 42°C. Das Plasmid sollte auf diese Weise in das Bakterium aufgenommen werden. Da hierbei auch meist eine zusätzliche Antibiotikaresistenz vermittelt wird, konnte nach Zugabe von 400 µl SOC-Medium und einer 1-3 stündigen Erholungsphase bei 37°C, anschließend auf Agarplatten mit dem betreffenden Antibiotikum selektiert werden. Bei einer kürzeren Inkubationsdauer ist nicht gewährleistet, dass dem Bakterium genügend Zeit zur Verfügung steht, die für die Resistenz benötigten Proteine zu exprimieren.

### 3.2.3 Kultivieren von *E. coli* - Bakterien

Auf Grund ihrer verschiedenen Eigenschaften bieten manche Bakterienstämme Vorteile im Hinblick auf die Expression von Proteinen, während andere eine höhere Ausbeute bei Plasmidpräparationen versprechen. Die Kultivierung unterscheidet sich jedoch nicht grundlegend. Aus einer angetauten Glycerol Dauerkultur wurde mittels eines sterilen Zahnstochers eine Übernachtskultur angeimpft. Hierzu wurden zwischen 10 und 50 ml LB oder 2-YT Medium mit den entsprechenden Antibiotika zur Selektion in ein 50 ml Falcon-Gefäß bzw. bei mehr als 10 ml Medium in einen 250 ml Glaskolben vorgelegt. Die Inkubation erfolgte bei 37°C auf dem Schüttler bei etwa 200 rpm über Nacht. Am nächsten Tag erfolgte das Animpfen einer Großkultur mit einer 1:20 Verdünnung der Übernachtskultur (ÜNK).

Verwendet wurden die folgenden Antibiotika in entsprechender Endkonzentration:

ANTIBIOTIKUM	KONZENTRATION
Ampicillin	100 µg/ml
Kanamycin	25 µg/ml
Tetracyclin	10 µg/ml
Chloramphenicol	15 µg/ml

**Tabelle 3.11:** In der Arbeit verwendete Antibiotika und ihre Konzentration

### 3.2.4 Dauerkulturen von *E. coli* - Bakterien

Um eine lange Lagerung von Bakterien zu ermöglichen, wurde 1 ml einer Übernachtskultur mit 250 µl Glycerol (80%) gut durchmischt und anschließend bei -80°C gelagert.

### 3.3 Molekularbiologische Methoden

#### 3.3.1 Plasmidisolierung aus *E. coli* - Bakterien

Die Durchführung erfolgte mit dem QIAprep Miniprep Kit von Qiagen, wobei die Vorgehensweise der im QIAprep Miniprep Handbuch (12/2006) entspricht. Es wurden 2 ml einer Übernachtskultur bei 6800 g für 3 min pelletiert und das aufgereinigte Plasmid in 25 µl EB-Puffer eluiert. Diese Art der Plasmidaufreinigung basiert auf alkalischer Lyse zum Zellaufschluss, dem Binden der DNA unter hohen Salzkonzentrationen an eine spezielle Silika-Membran, dem mehrmaligen Waschen und einer Elution unter salzärmeren Bedingungen.

#### 3.3.2 Agarose-Gelelektrophorese

Diese Art der Elektrophorese dient der Trennung von DNA Fragmenten verschiedener Größe und kann damit beispielsweise zur Überprüfung von PCR Reaktionen oder Restriktionsverdauen genutzt werden. Nach Anlegen eines elektrischen Feldes an die Elektrophoresekammer, wandern kleinere DNA-Fragmente wegen ihrer negativen Eigenladung und ihrer geringeren Größe schneller durch das auspolymerisierte Agarose-Gel als größere. Die Gele wurden je nach Agarosekonzentration für etwa 30-120 min in 1x TAE-Puffer laufen gelassen. Anschließend erfolgte die Färbung in einem Ethidiumbromidbad für etwa 15 min. Ethidiumbromid interkaliert in DNA-Doppelstränge und ermöglicht so eine Detektion unter UV-Licht. Durch Verwendung von kommerziellen Größenmarkern konnten Größe und Konzentration des DNA-Fragments bestimmt werden. Von den Proben wurden je nach Taschengröße meist 5 µl zusammen mit 1 µl Ladepuffer (6x) aufgetragen.

FRAGMENTLÄNGE	AGAROSEKONZENTRATION
1 bis 30 kb	0,5%
0,5 bis 7 kb	1,0%
0,1 bis 2 kb	2,0%

**Tabelle 3.12:** Abhängigkeit der Agarosekonzentration von der Länge des DNA-Fragments.

#### 3.3.3 Restriktionsverdau von DNA

Restriktionsenzyme haben die Eigenschaft DNA an spezifischen Sequenzen zu schneiden. So lassen sich beispielsweise Plasmide verifizieren oder spezielle Klonierungen vornehmen. Hierzu wurde die jeweilige Enzymmenge in Abhängigkeit von der zu verdauenden DNA Menge so gewählt, dass ein theoretischer 3- bis 5-facher Überverdau vorlag. Der Puffer wurde je nach Herstellerangaben und Enzym gewählt. Inkubiert wurde bei 37°C für 1-3 Stunden, wobei das Reaktionsvolumen zwischen 10 und 30 µl betrug. Anschließend wurde der Verdau mittels Agarose-Gelelektrophorese überprüft.

### 3.3.4 DNA Extraktion aus Agarosegelen

Nach einem Restriktionsverdau und anschließender Agarose-Gelelektrophorese wurden die DNA-Fragmente zur Weiterverarbeitung aufgereinigt. Die gewünschte Bande wurde unter UV-Licht schnellstmöglich mit einem Skalpell ausgeschnitten und in ein 2 ml Eppendorf-Gefäß überführt. Für die Extraktion wurde das QIAquick Gel Extraction Kit von Qiagen entsprechend dem Handbuch verwendet. Es zeigte sich jedoch zum einen, dass es Vorteile hat, einen 1 minütigen Inkubationsschritt bei 50°C zur vollständigen Entfernung des Ethanols einzufügen, als auch in vorgewärmten EB-Puffer (25 µl) zu eluieren. Ebenfalls verwendet wurde das peqGold GelPure™ Kit von peqLab, das beim Spektrum der aufgereinigten DNA im Bereich um 230 nm einen deutlichen Unterschied zum Qiagen kit zeigte. In der Effizienz unterschieden sich die beiden Ansätze jedoch nur unwesentlich.

### 3.3.5 Polymerasekettenreaktion (PCR)

Die Polymerasekettenreaktion ist eine Technik zur Amplifikation bestimmter DNA Sequenzen. Die zu amplifizierende DNA wird zusammen mit kleinen, für den Start und die Spezifität der Reaktion benötigten DNA-Sequenzen (Primer), einzelnen dNTPs sowie einem für die Verlängerung des DNA-Doppelstranges benötigten Enzym (Polymerase) in ein Reaktionsgefäß gegeben. Unter geeigneten Pufferbedingungen und mit, an die einzelnen Reaktionsschritte angepassten Temperaturen, kommt es dann zu einer exponentiellen Vermehrung des gewünschten DNA-Abschnittes. Die Taq-Polymerase reißt in einer Minute etwa 2000 Nukleotide aneinander. Die Pfu-Polymerase ist auf Grund ihrer 3'-5'Exonukleaseaktivität deutlich langsamer, hat dafür jedoch eine geringe Fehlerrate, die für die sich anschließende Protein-Expression von Bedeutung ist.

MIX	ZYKLUS	TEMP.	ZEIT.
1 µl dNTP (10 mM)	1.	95°C	2 min
5 µl PCR-Puffer (10x)	2.	95°C	40 sek
1 µl Vektor (etwa 200 ng/µl)	3.	55°C	40 sek
1,6 µl Vorwärts-Primer (10 µM)	4.	72°C	1,5 min für Taq-Polymerase
1,6 µl Rückwärts-Primer (10 µM)			3 min für Pfu-Polymerase
29,3 µl ddH <sub>2</sub> O Wasser	5.		30 Wiederholungen ab 2.
0,5 µl Taq/Pfu Polymerase (5 U/µl)	6.	72°C	10 min

**Tabelle 3.13:** Protokoll des Standard PCR Ansatzes und das verwendete PCR Programm.

### 3.3.6 Auswahl der Primer

Die Primer wurden mit Hilfe des Programms Vector NTI der Firma Invitrogen entworfen. So war es möglich, schon im Vorfeld ihre Spezifität zu überprüfen und die Schmelztemperatur für

die dann folgende PCR zu kalkulieren. Auf diesem Weg wurden auch zusätzliche Schnittstellen, Stopcodons sowie ein 8xHis-Tag in die Konstrukte eingefügt.

### 3.3.7 Klonierung von DNA Fragmenten in Plasmide

#### 3.3.7.1 Ligation

Ligasen haben die Eigenschaft freie 3'-OH-Gruppen eines Nukleotids mit einer freien 5'-Phosphatgruppe eines anderen Nukleotids verknüpfen zu können. Dies ermöglicht das Zusammenfügen vorher linearisierter DNA-Fragmente. Im vorliegenden Fall wurde der pQE-TriSystem His-Strep 1 Vektor mit NcoI und HindIII verdaut. Im einzufügenden Fragment wurden komplementäre Überhänge durch einen Verdau mit BspHI und HindIII erzeugt. Entscheidend für das Gelingen einer Ligation ist auch das molare Verhältnis von Insert zu Vektor, welches mit folgender Formel berechnet werden kann. Getestet wurden Verhältnisse zwischen 1:1 und 1:10 (Vektor/Insert), da je nach Größe des Inserts das optimale Verhältnis stark variieren kann.

$$\frac{\text{Menge Vektor [ng]} * \text{Insertgröße [kb]}}{\text{Vektorgröße [kb]}} * \frac{\text{Insert}}{\text{Vektor}} = \text{Menge Insert [ng]}$$

PLASMIDE, INSERTS	GRÖSSE [KB]
TriSystem His-Strep 1	5,7
pGEM T easy	3,0
Gli2 His-tag	0,7

**Tabelle 3.14:** Die Größen der verwendeten Plasmide und des Inserts.

Ligiert wurde bei Raumtemperatur (22°C) für 1-3 Stunden oder bei 16°C über Nacht. Die Aktivität der Ligase wurde durch 10 minütiges Erhitzen auf 65°C gestoppt.

#### 3.3.7.2 Blau-Weiß-Selektion

Unter der Verwendung von bestimmten Plasmiden, ermöglicht diese Technik auf recht einfache Art und Weise die Effizienz einer Klonierung zu überprüfen. Von Bedeutung ist vor allem die den aminoterminalen Teil der  $\beta$ -Galactosidase kodierende DNA-Sequenz (das  $\alpha$  Fragment des lacZ-Gens), die meist in einer Multiple Cloning Site (MCS) liegt. Als Konsequenz einer erfolgreichen Insertion des gewünschten DNA-Fragments in die MCS wird keine  $\beta$ -Galactosidase mehr gebildet, was durch Zugabe von X-Gal überprüft werden kann. Sofern das Enzym vorhanden ist, wird X-Gal abgebaut und in einen Indigo-Farbstoff umgewandelt, was letztlich zu einer Blaufärbung der Kolonien führt. Im gegenteiligen Fall verändert die Kolonie nicht ihre Farbe und bleibt weiß. IPTG ist wie Lactose in der Lage, den lac-Repressor zu binden und somit für die Induktion des lac-Operons zu sorgen. Der Unterschied zwischen beiden Substanzen

ist jedoch, dass IPTG kein Substrat der  $\beta$ -Galactosidase ist und seine Konzentration somit konstant bleibt.

Verwendet wurden LB-Agar Platten, die mit 25  $\mu$ l 1 M IPTG und 25  $\mu$ l einer 4%-igen X Gal-Lösung überzogen und vor der weiteren Verwendung bei Raumtemperatur getrocknet wurden.

#### Protokoll der Ligation:

- 50 ng Vektor
- 3  $\mu$ l 10x Ligase Puffer
- 1  $\mu$ l T4 Ligase
- X  $\mu$ l Insert (je nach Verhältnis)
- ad 30  $\mu$ l ddH<sub>2</sub>O

#### 3.3.8 *Colonie-PCR*

Diese Methode ermöglicht es eine relativ große Zahl von Bakterien auf das Vorhandensein eines gewünschten Inserts hin zu untersuchen. Nach erfolgter Blau-Weiß-Selektion werden Kolonien hierzu mit sterilen Zahnstochern in PCR-Reaktionsgefäße transferiert. Der Ansatz selbst ist identisch zu einer gewöhnlichen PCR. Die verwendeten Primer flankieren das zu kontrollierende DNA-Fragment. Der einzige Unterschied liegt in der verlängerten Inkubationszeit bei 95°C zu Beginn der PCR. Dieser dient dem Aufschluss der Zellen.

MIX	ZYKLUS	TEMP.	ZEIT.
1 $\mu$ l dNTP (10mM)	1.	95°C	10 min
5 $\mu$ l PCR-Puffer mit MgCl <sub>2</sub> (10x)	2.	95°C	1 min
Zellen einer Kolonie	3.	55°C	1 min
2 $\mu$ l Forward Primer (10 $\mu$ M)	4.	72°C	2 min Taq-Polymerase
2 $\mu$ l Reverse Primer (10 $\mu$ M)	5.		30 Wiederholungen 2.
39 $\mu$ l ddH <sub>2</sub> O Wasser	6.	72°C	10 min
1 $\mu$ l Taq Polymerase			

**Tabelle 3.15:** Protokoll des Standard *colony*-PCR Ansatzes und des verwendeten PCR Programms.

#### 3.3.9 DNA Sequenzierung

Die DNA-Sequenzierung wurde in einer DKFZ-internen Abteilung durchgeführt.

Genomics & Proteomics Core Facilities / Sequencing / Andreas Hunziker

## 3.4 Biochemische Methoden

### 3.4.1 Ermittlung der optimalen Expressionsbedingungen

#### 3.4.1.1 Proteinexpression in *E. coli* - Bakterien mit dem pMAL-c2X Vektor

Zur Proteinexpression wurde eine 50 ml ÜNK angesetzt, die am nächsten Tag mit 950 ml Medium und den entsprechenden Antibiotika in den geeigneten Konzentrationen gemischt und auf vier 1 l Erlenmeyerkolben verteilt wurde. Die Großkulturen wurden bei einer OD 600 von etwa 0,5 mit 0,3 mM IPTG induziert und für 3 Stunden bei 37°C auf dem Schüttler inkubiert. Anschließend wurde die Bakterienlösung bei 4000 g und 4°C für 15 min zentrifugiert, das Pellet in 50 ml Säulen-Puffer aufgenommen und bei -20°C gelagert. Nach dem Zellaufschluss durch Ultraschall oder mittels French press folgte die Proteinaufreinigung.

#### 3.4.1.2 Proteinexpression in *E. coli* - Bakterien mit dem pMAL-p2X Vektor

Die Großkulturen wurden annähernd wie der pMAL c2X Vektor behandelt. Das Pellet wurde statt Säulen Puffer jedoch in 50 ml einer Tris HCl – Saccharose Lösung aufgenommen und EDTA mit einer Endkonzentration von 1 mM zugesetzt. Nach einer Inkubation von etwa 10 Minuten bei Raumtemperatur folgte ein Zentrifugationsschritt bei 8000 g, wobei das Pellet diesmal in einer kalten 5 mM MgSO<sup>4</sup> Lösung resuspendiert wurde. Nach einer letzten Zentrifugation bei 8000 g für erneut 15 Minuten wurde dem, die Proteine des Periplasmas enthaltene Überstand, noch 1 ml Tris-HCl pH 7,4 zugesetzt. Dann erfolgte die Aufreinigung über eine Amylose Harz Matrix.

### 3.4.2 Zellaufschluss

#### 3.4.2.1 French Press

Bei dieser Methode werden Zellen vom Probenzylinder, der unter hohem Druck steht, durch ein sehr dünnes Verbindungsstück in einen anderen Kolben gepresst, der lediglich unter Atmosphärendruck steht. Durch diesen abrupten Druckwechsel wirken enorme Scherkräfte, die zum Zerreißen der Zellwand führen. Die aufgetaute Zellsuspension wird in die über Nacht vorgekühlte French press gefüllt und der Druckregler auf etwa 40 psi eingestellt. Dieser Wert kann je nach Probe variieren, letztlich sollte jedoch ein Druck von etwa 1000 – 1500 bar erreicht werden. Das Einstellen dieses Wertes ist nicht immer sofort möglich. Sobald er jedoch konstant ist, genügt ein 3 maliges Wiederholen der Prozedur, um einen vollständigen Aufschluss der Zellen zu gewährleisten. Man erkennt eine erfolgreiche Lyse auch an der Veränderung von Farbe und Viskosität der Probe. Zwischen der Lyse von zwei unterschiedlichen Proben wurde das Gerät unter ständiger Variation des Drucks gereinigt. Hierzu wurden folgende gekühlte Flüssigkeiten verwendet:

- 50-100 ml Wasser
- 2x je 50-100 ml 0,5 M NaOH
- 50-100 ml 10mM HCl
- 50-100 ml 1x PBS als Äquilibrierungspuffer

Nach der letzten Lyse wird statt dem Äquilibrierungspuffer noch einmal mit 20% Ethanol gewaschen und letztlich erneut Ethanol nachgefüllt, der dann bis zur nächsten Benutzung im Kolben verbleibt. Das Lysat wurde für 1 Stunde bei 10.000 g und 4°C zentrifugiert und vor der Proteinaufreinigung über einen Filter (0,22 µm, MILLEX®GV) aufgereinigt.

### 3.4.2.2 Ultraschall

Die abzentrifugierten Bakterien wurden in anwendungsspezifischen Puffern resuspendiert und durch Ultraschallbehandlung aufgeschlossen. Puffer, sowie Dauer und Anzahl der Ultraschall-Impulse sind in folgender Tabelle aufgeführt:

PUFFER	ULTRASCHALLBEHANDLUNG
20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM Azid, 10 mM β-ME or 1 mM DTT	30%, 8x 15sec

**Tabelle 3.16:** Protokoll der Ultraschallbehandlung

### 3.4.3 Proteinaufreinigung

#### 3.4.3.1 Affinitätschromatographische Aufreinigung über eine Amylose - Harz - Matrix

Bis zur Aufreinigung unterschied sich die Behandlung der beiden pMAL-Vektoren bisweilen grundlegend, da das Protein einmal im Cytoplasma und einmal im Periplasma lokalisiert war. Die Aufreinigung selbst unterscheidet sich jedoch nicht und wurde ebenfalls wie im Handbuch des pMAL Protein Fusions- und Aufreinigungssystem von der Firma New England Biolabs beschrieben, durchgeführt. Das MBP-Fusionsprotein bindet hierbei an den Amyloseteil der Matrix, wird durch mehrfaches Waschen mit Säulen-Puffer gereinigt und letztlich mit Säulen-Puffer und 10 mM Maltose eluiert. In eine 2,5 x 10 cm Säule wurden 1,5 ml Amylose Harz gegeben und mit dem 4-fachen des Säulenvolumens mit Säulen-Puffer gewaschen. Die zuvor aufbereiteten Überstände wurden aufgebracht, und mit dem 8-fachen Volumen Puffer gewaschen und das aufgereinigte Protein anschließend mit 15 ml 10 mM Maltose in Säulen Puffer eluiert, wobei die Durchflussgeschwindigkeit etwa 1 ml pro Minute betrug. Die auf diese Weise erhaltenen je 1,5 ml großen Fraktionen wurden mit SDS Gelen analysiert.

### 3.4.3.2 Affinitätschromatographische Aufreinigung über Ni - NTA - Agarose

Bei der Aufreinigung über Ni-NTA Agarose macht man sich die Bildung des Komplexes zwischen dem His-tag eines Proteins und der Ni-NTA-Agarose zu Nutze. Da Imidazol mit den Histidinen des Proteins um den Ni-Chelat-Komplex konkurriert, kommt es durch das Ansteigen der Imidazolkonzentration in den verschiedenen Wasch- und Elutions-Puffern zuerst zu einer Aufreinigung, und letztlich zur Elution des gewünschten Proteins. Es wurden je 5 ml des Bakterien-Lysates mit 1,5 ml Ni-NTA-Agarose in einem 15 ml Falcon-Gefäß gemischt (batch Verfahren) und für 1 bis 3 Stunden bei 4°C auf dem Rollschüttler inkubiert. Anschließend wurde die Suspension auf eine 5 ml Polypropylen-Säule übertragen und eluiert, wobei darauf geachtet wurde, dass immer etwas Flüssigkeit über der Agarose zurück bleibt. In der Folge wurde zwei Mal mit je 4 ml Wasch-Puffer, mit verschiedenen Imidazolkonzentrationen, gewaschen und letztlich vier Mal mit je 500 µl Elutions-Puffer eluiert. Alle Fraktionen wurden gesammelt und mittels SDS-PAGE analysiert.

### 3.4.3.3 Aufreinigung mittels *Fast protein liquid chromatography* (FPLC)

Ein großer Vorteil dieser Methode, im Vergleich zu einer manuellen Aufreinigung mit dem batch Verfahren, ist die Möglichkeit, einen kontinuierlichen Imidazol-Gradienten aufzubauen. Dies sorgt, durch eine punktgenaue Elution des gewünschten Proteins für eine bessere Aufreinigung. Als Plattform diente eine Äkta Prime Plus unter der Verwendung von HisTrap FF 1 ml Säulen, beides von GE Healthcare.

#### Reinigen des Gerätes

Da das Gerät zu Beginn mit 20% Ethanol gefüllt war, begann jede Aufreinigung mit dem Durchlauf des „System-Wasch-Programms“. Hierbei wurden auch alle verwendeten Schläuche unter hohem Druck und mit schneller Durchflussgeschwindigkeit zuerst mit 20% Ethanol und anschließend mit dH<sub>2</sub>O gespült.

#### Äquilibrieren der Säule

Weil auch die Säule zu Beginn mit 20% Ethanol gefüllt war, wurde zuerst mit 3- bis 5 fachem Säulenvolumen Wasser und anschließend mit mindestens dem 4- fachen Volumen des späteren Bindepuffers gewaschen. Die Durchflussgeschwindigkeit beträgt hierbei und auch für die weiteren Schritte stets 1 ml/min bei 0,3 MPa.

#### Beladen der Säule

Das Beladen der Säule mit der Probe erfolgte über eine externe Schleife mit Hilfe einer Spritze. Das Fassungsvermögen betrug hier jeweils 5 ml, so dass das gesamte Probenvolumen in 3 Schritten zugegeben werden konnte. Zwischen den einzelnen Zugaben wurde so lange gewartet, bis der unmittelbar hinter der Säule angebrachte UV-Detektor kein Protein mehr anzeigte.

### **Aufreinigung und Elution**

Als Startpunkt diente der Lyse-Puffer mit 10 mM Imidazol, von dem aus über ein Volumen von 20 ml ein linearer Gradient bis hin zum Elutionspuffer mit 250 mM Imidazol aufgebaut wurde. Aufgefangen wurden die Fraktionen in 500  $\mu$ l pro Eppendorf-Gefäß. Auch hier erfolgte eine Analyse der Proben durch SDS-PAGE.

### **Reinigen der Säulen und des Gerätes**

Zuerst wurde die Säule mit etwa 20 ml dH<sub>2</sub>O und einem entsprechenden Volumen an 20% Ethanol gewaschen und dann luftdicht aufbewahrt. Dann wurde das Gerät durch Ausführen des „System-Wasch-Programms“ mit Wasser und anschließend noch ein weiteres Mal mit 20% Ethanol gereinigt.

#### **3.4.4 Denaturierende Natriumdodecylsulfat Polyacrylamid Gelelektrophorese**

SDS-Gele erlauben die Auftrennung von Proteinen nach ihrem Molekulargewicht. Als Matrix dient hierbei ein Gemisch aus Acrylamid und Bis-Acrylamid, durch deren Konzentration die Porengröße und letztlich das Trennverhalten des Gels bestimmt wird. SDS als anionisches Detergenz hat die Eigenschaft sich in regelmäßigen Abständen, statistisch alle 1,6 AS, an ein Protein anzulagern und sorgt somit für denaturierenden Bedingungen, ein proportionales Verhältnis von Ladung zu Molekulargewicht und letztlich für eine korrekte Auftrennung. Durch den Zusatz von  $\beta$ -Mercaptoethanol in den Ladepuffer, werden desweiteren Disulfidbrücken gespalten, was zu einem Zerfall in die einzelnen Untereinheiten führt. Wie bei anderen Elektrophorese-Arten üblich, wandern auch hier kleine Moleküle schneller durch die Poren der Matrix als Größere. Durch Hinzufügen eines niederprozentigen Sammelgels kann man die Auflösung des Gels verbessern.

Die zu untersuchenden Proben wurden für 10 min bei 95°C in Ladepuffer denaturiert und bis zum Auftragen auf Eis gelagert. Die Gele wandern bei etwa 100 V, bis das Bromphenolblau aus dem Ladepuffer das untere des Gels erreicht. Durch den Einsatz von kommerziellen Markern konnte das Molekulargewicht eines Proteins bestimmt werden. Die möglichen Färbetechniken unterscheiden sich in ihrer Nachweisgrenze und sind individuell wählbar.

### **Coomassie-Färbung**

Zur Färbung der Proteinbanden wurde Roti®-Blue, eine kolloidale Coomassie-Brillant-Blau G250-Färbung verwendet, deren Nachweisgrenze laut Hersteller bei etwa 30 ng Protein liegt.

#### **Coomassie-Färbelösung:**

- 20 ml Roti-Blue (5xKonzentrat)
- 20 ml Methanol
- 60 ml dH<sub>2</sub>O

Da nur kleine Gele in der Größe 12 cm x 9 cm verwendet wurden, mussten keine weiteren Fixierungsschritte oder ähnliches vorgenommen werden. Die Färbezeit variierte zwischen 2 und 16 Stunden, wobei der Hintergrund durch angepasste Entfärbezeiten in Wasser wieder minimiert wurde.

#### **Instant-Blue-Färbung**

Eine andere Möglichkeit zur Proteinfärbung bietet die Firma Expedeon an. Der Farbstoff Instant Blue ist ein Derivat des Comassie Farbstoffes und färbt Proteine in SDS Gelen innerhalb weniger Minuten an. Der Hintergrund ist hierbei so gering, dass ein weiterer Entfärbeschritt nicht notwendig ist.

#### **3.4.5 Western Blot**

Proteine, die zuvor mittels SDS-PAGE nach ihrem Molekulargewicht aufgetrennt wurden, konnten mit der semi-dry-Western-Blot-Technik auf eine Nitrocellulose-Membran übertragen und dann mit speziellen Antikörpern inkubiert und nachgewiesen werden. Als 1. AK wurden ein anti-Maus IgG Antikörper bzw. ein Anti-Biotin Antikörper verwendet. Mit einem 2. Antikörper, der mit alkalischer Phosphatase als Konjugat gekoppelt wurde, konnte man später die Banden detektieren. Hierzu wurde BCIP/NBT als Substrat für die alkalische Phosphatase verwendet.

Zu Beginn wurden einige Stücke Whatman-Papier und ein Stück Nitrocellulose-Membran, die alle etwa die Größe des Gels haben, wenige Minuten in Transferpuffer eingelegt. Auf Grund der negativen Ladung der Proteine ist der Aufbau wie folgt:

#### **Kathode:**

- 3 Stücke Whatman-Papier
- SDS-Gel
- Nitrocellulose-Membran
- 3 Stücke Whatman-Papier

#### **Anode:**

Kleine Luftblasen wurden durch vorsichtiges Ausstreichen entfernt. Um einen guten Stromfluß zu gewährleisten, wurde zusätzlich noch etwas Transferpuffer nachgelegt. Durch Anlegen eines elektrischen Feldes bei 35 mA für 1 Stunde wurden die Proteine auf die Nitrocellulose-Membran übertragen.

#### **Protokoll zur Durchführung eines Western Blots:**

- 1 Stunde mit 2% BSA in TBS/T blocken
- 3 x für je 10 min mit TBS/T/T waschen
- Inkubation Über-Nacht mit 1. AK (1:1000) bei 4°C (0,3 g BSA, 10 µl Tween20, 10 µl AK, ad 10 ml TBS; Eingeschweißt in Plastik / auf dem Überkopftaumler)
- 2 x für je 10 min mit TBS/T/T waschen
- 10 min mit TBS waschen
- Inkubation für 1 Stunde mit 2. AK (1:20000) bei RT; 0,5 g BSA, 0,5 µl AK, ad 10 ml TBS
- 2 x für je 10 min mit TBS/T/T waschen
- 10 min mit TBS waschen

Die Detektion erfolgte mit BCIP/NBT, einem Substrat, bei dessen Reaktion mit der alkalischen Phosphatase ein Indigo-Farbstoff gebildet wird. Durch kurzes Waschen in Wasser konnte die Reaktion gestoppt werden.

#### **3.4.6 Aufkonzentrieren von Proteinlösungen und Pufferaustausch**

Da der für die Aufreinigung eines Proteins die verwendete Puffer für weitere Experimente teilweise störten, war es notwendig, ihn gegen einen Anderen auszutauschen. Im vorliegenden Fall der His-Tag Aufreinigung galt es vor allem Imidazol zu entfernen, das für die Elution des Proteins verwendet wurde. Imidazol würde die folgende Proteinmengenbestimmung mittels BCA oder die Absorptions-Messung stören. Hierfür wurden Amicon® Ultra 4 Filter-Zentrifugenröhrchen (10 KDa) verwendet. Moleküle mit einem Molekulargewicht von weniger als 10 KDa können die Filtermembran passieren, größere jedoch nicht. Dies führt zu einer Abnahme der Salzkonzentration bei nahezu unveränderter Proteinkonzentration. Die Proteinlösung wurde bei 4000 g und 4°C etwa 15 min zentrifugiert, bis ein Restvolumen von ungefähr 200 µl übrig blieb. Je nach Proteinkonzentration konnte diese Zeit jedoch etwas variieren. Im Anschluss wurde mit 1x Binde-Puffer auf 4 ml aufgefüllt und erneut zentrifugiert. Nach einer dritten Wiederholung wurde die Proteinlösung aus dem Filter abpipettiert und in ein Eppendorfgefäß überführt. Die weitere Lagerung der Proteine erfolgte auf 4°C oder als Aliquotes auf -70°C.

Die gleichen Amicon® Ultra 4 Filter-Zentrifugenröhrchen (10 KDa) konnten auch zum Aufkonzentrieren der Proteine benutzt werden.

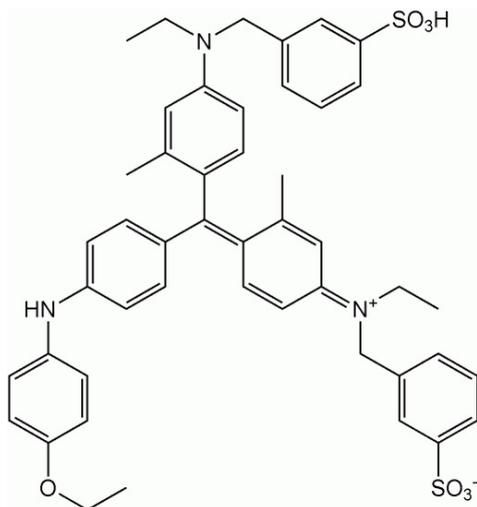


Abbildung 3.1: Coomassie Brilliant Blue G-250

### 3.4.7 Quantifizierung und Identifizierung von Proteinen

#### 3.4.7.1 Proteinmengenbestimmung nach Bradford

Die Methode zur Proteinkonzentrationsbestimmung nach Bradford (Bradford (1976)), auch Bradford Protein Assay genannt, basiert auf der Komplexbildung des Farbstoffes Coomassie Blue und kationischen sowie unpolaren Seitenketten von Proteinen in saurerer Lösung (Tal et al. (1985)). Dabei verschiebt sich das Absorptionsmaximum des ungebundenen Farbstoffes mit rötlicher Färbung, das zunächst bei 465 nm liegt, nach 595 nm in der gebundenen Form mit blauer Farbe. Der Anstieg der Absorption bei 595 nm ist proportional zu der Menge an gebundenem Farbstoff und somit der Menge oder Konzentration an Protein in der Probe. Diese Untersuchung ist sehr gut reproduzierbar und schnell durchzuführen, da die Farbstoffbindung in den meisten Fällen nach zwei Minuten abgeschlossen ist und für eine Stunde stabil bleibt.

Weiterhin wird Coomassie Blue wenig bis überhaupt nicht durch Kationen, wie Kalium oder Natrium und Kohlenwasserstoffverbindungen wie Saccharose beeinträchtigt. Lediglich Detergenzien wie SDS oder Triton X-100 beeinflussen die Farbreaktion deutlich (Boccaccio and Quesada-Allue (1989)).

#### Regular Bradford Protein Assay

Der normale Bradford-Assay wird für einen Konzentrationsbereich zwischen ungefähr 100-8000  $\mu\text{g}/\text{mL}$  verwendet, da ein Volumenverhältnis von Bradford-Reagenz zu Probe von 50:1 eingesetzt wird. In dieser Arbeit wurden zunächst BSA-Lösungen als Standard mit mindestens acht verschiedenen Konzentrationen hergestellt und dann deren Absorption bei 595 nm gemessen. Um eine hohe Linearität der Standardkurve zu garantieren und einen großen Konzentrationsbereich abdecken zu können. Durch Zuordnung der Absorptionswerte zu den jeweiligen Konzentrationen wurde eine Gerade generiert und diese für die Berechnung der unbekannt Konzentrationen herangezogen.

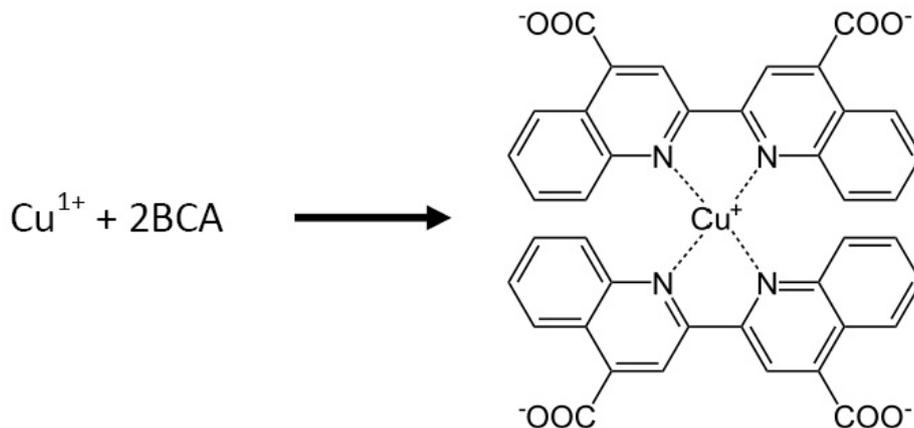
### Mini Bradford Protein Assay

Der Mini-Bradford-Assay ist für Konzentrationen zwischen 15-100 µg/mL geeignet, da ein Volumenverhältnis von 1:1 von Reagenz zu Probe eine höhere Sensitivität in diesem Bereich bedeutet. In diesem Fall wurden ebenso zuerst mindestens acht BSA-Lösungen verschiedener Konzentration hergestellt, die jeweilige Absorption bei 595 nm gemessen und daraus anschließend eine Standardkurve berechnet. Aus dieser konnten die Proteinkonzentrationen unbekannter Lösungen bestimmt werden.

Da manche Detergenzien diesen Test stören können (Boccaccio and Quesada-Allue (1989)), kann als Alternative eine BCA-Test gemacht oder die Absorption bei 260 nm gemessen werden.

#### 3.4.7.2 Proteinmengenbestimmung mittels Bicinchoninsäure (BCA)

Die BCA-Reaktion (Smith et al. (1985), Wiechelman et al. (1988)) ist eine alternative Methode zur Bestimmung von Proteinkonzentrationen in Lösungen. Sie gründet auf der Reduktion von  $\text{Cu}^{2+}$  - zu  $\text{Cu}^{1+}$ -Ionen durch Peptidbindungen in Proteinen, wobei die Menge an reduzierten Kupfer-Ionen proportional der vorhandenen Proteinmenge in der alkalischen Lösung ist. Weiterhin bilden zwei Moleküle der Bicinchoninsäure (BCA) mit einem  $\text{Cu}^{1+}$ -Ion einen Komplex (siehe Abbildung 3.2), welcher von violetter Farbe ist und Licht bei einer Wellenlänge von 562 nm sehr stark absorbiert. Über den Vergleich der Absorptionsspektren von Proben bekannter Konzentration, lässt sich die Menge an Protein in unbekanntenen Proben bestimmen.



**Abbildung 3.2:** Struktur eines Kupfer-Bicinchoninsäure- Komplexes

Ansetzen der Arbeitslösung aus dem BCA™ Protein Assay Kit:

- 98 µl Lösung A ( Carbonat-Puffer mit BCA-Reagenz)
- 2 µl Lösung B (Kupfersulfat-Lösung)
- 5 µl Proteinlösung

### 3.4.7.3 Protein-/DNA-Mengenbestimmung mittels Absorption

Die Konzentrationsbestimmung von unbekanntem DNA- und Proteinlösungen erfolgte mit dem NanoDrop®ND-1000 Spektrophotometer der Firma NanoDrop Technologie. Zur Messung von DNA wird ein Probenvolumen von 1  $\mu\text{l}$  benötigt, bei Proteinen hingegen sind es 2  $\mu\text{l}$ . Das Gerät misst ein Absorptionsspektrum von 220 nm bis 350 nm, was ein Überprüfen der Lösung auf eventuelle Verunreinigungen ermöglicht. Lösungsmittel und manche niedermolekulare Substanzen absorbieren beispielsweise bei etwa 230 nm, wohingegen das Absorptionsmaximum von DNA und Proteinen bei 260 nm bzw. 280 nm liegt. Als „rein“ angesehen werden DNA-Lösungen mit einem Verhältnis der Absorptionen 260/280 von etwa 1,8. Das Verhältnis von 260/230 sollte zwischen 1,8 und 2,2 liegen.

Zur Proteinmengenbestimmung musste der Extinktionskoeffizient und das Molekulargewicht der Fusionsproteine bestimmt werden. Da durch die Sequenzierung der Plasmide die DNA Sequenz bekannt war, konnte diese in die Proteinsequenz übersetzt werden. Daraus konnte dann das Molekulargewicht und der Extinktionskoeffizient der Proteine berechnet werden. Berechnungen von verschiedenen Programmen aus dem Internet ergaben leicht abweichende Werte von  $\epsilon$ . Dies lässt sich durch die Abhängigkeit von den verschiedenen Pufferbedingungen erklären. Der Extinktionskoeffizient der Proteine berechnet sich aus der Summe der Extinktionskoeffizienten der einzelnen Aminosäuren Tryptophan, Thyrosin und Cystein (siehe Abbildung 3.3).

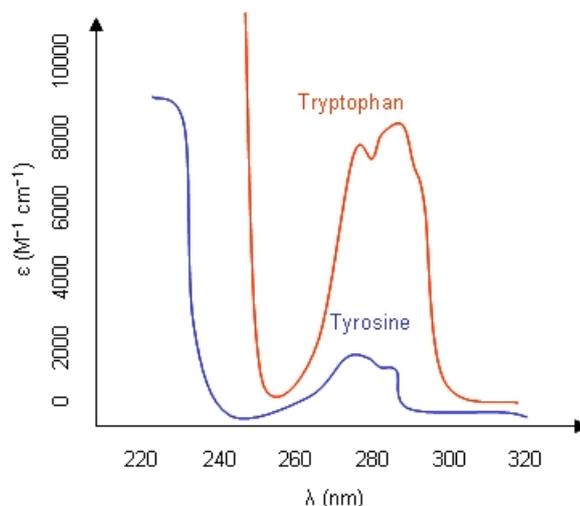


Abbildung 3.3: UV-Absorption von Tryptophan und Thyrosin.

#### Durchführung:

Hierbei wurden 2  $\mu\text{l}$  Probe auf den Probensteller pipettiert und die Absorption bei 260 nm für DNA und bei 280 nm für Proteine gemessen. Die Konzentrationsbestimmung erfolgte mit Hilfe des Lambert-Beer'schen Gesetzes:

$$\mathbf{A} = \epsilon * \mathbf{c} * \mathbf{d}$$

A = Absorption

$\epsilon$  = Extinktionskoeffizient

c = Konzentration

d = Dicke der Küvette (beim Nanodrop 0,2 mm bzw. 1 mm)

### 3.4.8 Proteinanalyse mittels MALDI-TOF-Massenspektrometrie

Die MALDI Analyse der Proteine wurde in einer DKFZ-internen Abteilung durchgeführt:

Genomics & Proteomics Core Facilities / MALDI / Dr. Martina Schnölzer

## 4 Experimente und Ergebnisse

Nachfolgend werden im ersten Teil der Ergebnisse der Versuchsaufbau sowie die Entwicklung und Etablierung der Methode anhand des Fusionsproteins MBP-3xMyc-Gli2 und His-Gli2 gezeigt. Im zweiten Teil folgt dann die Anwendung der Methode am Beispiel mehrerer krebsrelevanter Transkriptionsfaktoren.

### 4.1 Versuchsaufbau

#### 4.1.1 *E. coli* - Bakterien und Plasmide

Zur Verfügung standen verschiedene *E. coli* - Stämme und unterschiedliche Plasmide. Zu Beginn der Arbeit wurden die Transkriptionsfaktoren in XL1-Blue Zellen als Fusionsproteine mit dem Maltose Binding Protein (MBP) und einem 3-fachen Myc-Tag exprimiert. Dazu wurden die Vektoren pMAL p2X und pMAL c2X verwendet.

- pMAL c2X: Die in das Plasmid klonierte DNA-Sequenz wird als Fusionsprotein mit MBP in XL1-BlueZellen exprimiert und im Cytoplasma gelagert. Dies kann in vielen Fällen helfen, die Löslichkeit des zu exprimierenden Proteins zu verbessern. Desweiteren verfügt dieses Plasmid über eine DNA-Sequenz, die für eine Schnittstelle der Protease X kodiert und es somit ermöglicht, das Maltose Binding Protein später proteolytisch abzuspalten.
- pMAL p2X: Dieses Plasmid codiert noch für eine zusätzliche, dem MBP vorgeschaltete Aminosäuresequenz, die für das Fusionsprotein als Exportsignal in das Periplasma dient. Das dort vorherrschende Milieu kann für manche Proteine stabilisierend wirken, da Schwefel-Schwefelbindung ausgebildet werden können.

#### 4.1.2 Fusionsproteine

Die in dieser Arbeit verwendeten Fusionsproteine sind nach folgendem Schema aufgebaut:

MBP-3xMyc-TF  
His-TF

- MBP: Maltose bindendes Protein (für Protein-Aufreinigung; N-terminal)
- 3xMyc: 3-fach Myc-Markierung (für Antikörper-Detektion; N-terminal)
- His: Histidin-Markierung (Aufreinigung und Detektion; N-terminal)

- TF: Bindedomäne des Transkriptionsfaktors

Das Maltose bindende Protein wurde für die Aufreinigung der Fusionsproteine benötigt. Durch die sehr gute Löslichkeit des MBP wird zusätzlich die Löslichkeit und Stabilität des Fusionsproteins erhöht (Kapust and Waugh (1999)). Mit Hilfe der 3-fach Myc-Markierung konnte unter der Verwendung eines Anti-Myc-Antikörper und des entsprechenden 2. Antikörper das Fusionsprotein detektiert werden. Der Histidin markierte Transkriptionsfaktor wurde über den His-Tag sowohl aufgereinigt als auch detektiert.

Es wurden für die Experimente bei den meisten Transkriptionsfaktoren nur die Bindeseite verwendet, da in *E. coli* -Bakterien nur Proteine bis 100 KDa exprimiert werden können. Es waren somit nur wenige Vollängenproteine zugänglich.

Von der direkten Kopplung eines Fluoreszenzfarbstoffes an das Protein wurde abgesehen, da die kovalente Bindung die Funktionalität des Proteins zu sehr beeinflussen könnte.

#### 4.1.3 Array-Design

Das Array Design wurde mit Microsoft Excel, dem auf Linux basierenden Programm *Mutscan* und der Geniom Software von Febit angefertigt. Die benötigten Konsensus-Sequenzen und Proteininformation wurden von folgenden Quellen bezogen:

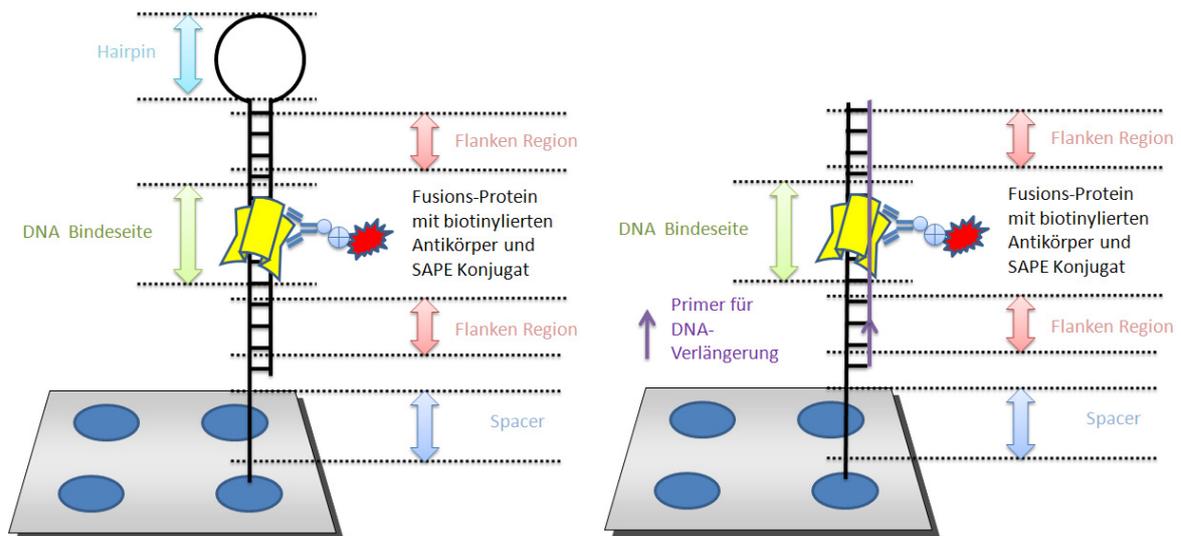
- <http://www.biobase.de/pages/index.php?id=transfacdatabases>
- <http://www.genecards.org/>

Da der zu Beginn der Arbeit benutzte Array jedoch noch nicht über ausreichend Kapazität für alle vorstellbaren Nukleotidsequenzen eines 9 mers ( $4^9 = 262144$ ) verfügt, wurden zuerst nur ein Teil der Sequenzen untersucht. Ausgehend von der neun Basenpaare langen Konsensus-Bindesequenz der Proteine galt es alle möglichen Permutationen von einem bzw. zwei Nukleotiden auf ihre Bindungseigenschaften zu untersuchen. Diese wurden mit dem Programm Mutscan ausgehend von der Konsensus-Sequenz generiert.

Die Erweiterung auf 125.000 Spots pro Array ermöglichte gegen Ende der Doktorarbeit die Untersuchung aller möglichen 9 mer Bindesequenzen. Jede Sequenz eines 9 Basenpaar langen Oligonukleotids ist mindestens einmal auf dem Chip vertreten. Die Berechnung dieser Sequenzen wurde durch die Benutzung einer "doppelsträngigen De Bruijn Sequenz" ermöglicht (Algorithmus von J. Kärkkäinen und P. Rastas, Kaerkaeinen (2003) ). Dabei wurde durch die mehrfach-Belegung eines Spots die Anzahl der Sonden von 262.000 auf 50.000 reduziert. Somit konnten auf einem Chip zwei Replikate synthetisiert werden.

#### 4.1.4 DNA-Synthese

Die Synthese der Oligonukleotide erfolgte im Geniom Gerät durch chemische Synthese nach Standard-Amidit-Chemie aus kommerziell erhältlichen Synthesebausteinen. Die Sequenz der einzelnen Nukleotide ist komplementär, so dass sie durch Bildung eines Hairpins einen Doppelstrang ausbilden können und somit die Bindung der Proteine an die DNA-Helix gewährleisten. Es war ebenfalls möglich, durch Primer-Verlängerung den 2. Strang mit Hilfe einer Polymerase zu erzeugen (siehe Abbildung 4.1).



**Abbildung 4.1:** Protein-Antikörper-DNA-Komplex auf einem Microarray. Hairpin (links) und Primer-Verlängerung (rechts)

#### 4.1.5 Protein-Inkubation

Soweit nicht anders beschrieben wurden die Reaktionsschritte bei Raumtemperatur durchgeführt. Alle Puffer wurden steril filtriert. Die Proteininkubation erfolgte in einer externen Halterung, da hier weniger Protein zur Inkubation benötigt wurde. Der Microarray wurde zu Beginn für eine Stunde mit 2% BSA in 1x PBS geblockt und das Fusionsprotein in 1x Bindepuffer verdünnt. Frisch zugesetzt wurden poly(dI/dC) mit einer Konzentration von 5 µg/ml und AEBSF mit einer Konzentration von 0,5 µg/µl. So sollte zum einen unspezifische DNA-Bindung minimiert und zum anderen das Protein stabilisiert werden. Die verwendeten Konzentrationen der Proteine variierten zwischen 6 - 100 nM. Nach einer 1 stündigen Proteininkubation wurden alle 8 Arrays des Chips mit jeweils 200 µl 1x Bindepuffer gewaschen.

Alle weiteren Inkubationen wurden im Geniom 2X System durchgeführt. Folgende Lösungen wurden dafür verwendet:

- Antikörper-Lösung: 10 ml, 1x Bindepuffer + 10 µl Anti cMyc Antikörper (biotinyliert) + 2% w/v BSA. (Bei Verwendung von Gli2 mit His-Tag wurde ein biotinylierter anti-His Antikörper verwendet.)

- SAPE-Lösung: 9 ml 1x Bindepuffer + 44 µl SAPE + 2% w/v BSA
- Wasch-Lösung: 80 ml 1x Bindepuffer + 2% w/v BSA

#### 4.1.6 Detektion

Die Detektion des Phycoerythrin-Farbstoffes erfolgte mittels eines integrierten Cy3-Filtersets. Typische Belichtungszeiten lagen im Bereich von 1000-5000 ms.

#### 4.1.7 Auswertung

Die Auswertung der Daten erfolgte mittels Microsoft Excel und Sigmaplot von Systat Software. Zuerst wurde der lokale Hintergrund der einzelnen Spots subtrahiert. Nach dem Berechnen der Mittelwerte und des Medians wurden die Werte gegen die aus der Literatur bekannten Konsensus-Sequenz normalisiert. Eine graphische Darstellung erfolgte nach der Methode der gewichteten Positionsmatrix bzw. der Sequenzmusteranalyse.

##### 4.1.7.1 Gewichtete Positionsmatrix

Die Mittelwerte der Signalintensitäten aller Replikate mit einer Permutation wurden gegen die Intensität der Konsensus-Sequenz normalisiert. Das Verhältnis jeder der vier Mutationen eines Nukleotids an einer bestimmten Position (1-9) wird dann durch die Summe aller Werte der selben Position geteilt. Dies wird für jede der 9 möglichen Positionen durchgeführt und die Werte dann auf der x-Achse aufgetragen. Aus diesen Schaubildern lässt sich ableiten, wie kritisch die Substitution eines bestimmten Nukleotids an einer definierten Position in der Bindesequenz für die Affinität der Bindung ist. Die visualisierte Darstellung erfolgte als Sequenzlogo mit Hilfe des von Workman et al. (2005) entwickelten Algorithmus.

##### 4.1.7.2 Sequenzmusteranalyse

Um wie bei der gewichteten Positionsmatrix nicht nur die einfachen Permutationen bei der Auswertung in Betracht zu ziehen, werden bei der Sequenzmusteranalyse alle Mutationen analysiert (Schneider (2002)). Es werden hierbei nur die stärksten 5% aller Signale als Sequenzmuster analysiert. Diese werden entsprechend ihrer höchsten Werte sortiert. Danach wird das Verhältnis der vier verschiedenen Nukleotide an den jeweiligen Position (1-9) der Konsensus-Sequenz bestimmt. Eine einfache Methode dieses Sequenz-Logos visuell darzustellen lieferte Crooks et al. (2004) et al.

## 4.2 Etablierung der Methode anhand des Transkriptionsfaktors Gli2 (Gli-Kruppel family member 2)

Die Wahl Gli2 zur Etablierung der Methode zu benutzen hatte vor allem zwei Gründe. Zum Einen gehört Gli2 zu den bisher etwa 40 mit Krebs in Verbindung gebrachten Transkriptionsfaktoren. Zum Anderen hatte bereits ein Kooperationspartner ergänzende Experimente gemacht und stellte auch einen Teil der verwendeten Konstrukte zur Verfügung.

### 4.2.1 Nachweis von Gli2 mittels Sequenzierung

Die von dem Kooperationspartner zur Verfügung gestellten Plasmide wurden zur Kontrolle noch einmal sequenziert. Dabei wurde nur das Insert mit der codierenden Proteinsequenz kontrolliert.

#### Ergebnis der Sequenzierung:

```
GAGCTCGAACAAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTCAGAATTCGGCTCTGAG
CAAAAGCTCATTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCAACGGTGAGCAGA
AGTTAATATCCGAGGAGGATCTTTCTAGAGTCGACAAGCAGGAGGCTGAGGTGGTCATCTACGAGACCAACTGCCA
CTGGGCAGACTGCACCAAGGAGTATGACACACAGGAGCAGCTGGTGCATCATATCAACAATGAACACATCCACGGG
GAGAAGAAGGAGTTCGTGTGCCGCTGGCAGGCCTGCACGAGAGAGCAGAAGCCCTTCAAGGCCAGTACATGCTGG
TTGTTACATGCGCAGACACACGGGTGAGAAGCCACACAAGTGCACGTTTGAAGGCTGTTCCAAGGCCTACTCTCG
CCTGGAGAACCTGAAGACACACCTGCGTTCACACACAGGAGAGAAGCCATATGTGTGTGAACACGAAGGCTGTAAC
AAAGCCTTCTCCAATGCCTCAGACCGCGCCAAGCACCAGAACCGCACTCACTCCAATGAGAAACCCTACATCTGCA
AGATCCCAGGCTGCACCAAGAGGTACACAGACCCAGCTCACTCCGCAAGCATGTGAAGACTGTCCATGGGCCAGA
CGCCCATGTACCAAGAAACAGCGTAATGATGTGCATGTCCGTGCTCCACTGCTCAAGGAGAATGGGGATAATGAG
GCCAGCGCCGAGCCAGGTGGCCGGGACCTGAGGAGAGTGTGGAGGCCAGTAGCACCAGCCACACTGTGGAGGACT
GCCTACATATCAAAGCCATCAAGACCGAGAGCTCCGGGCTGTAAAAGCTT
```

Diese Sequenz wurde anschließend in die Aminosäuresequenz übersetzt und mit folgender Datenbank abgeglichen: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Der Abgleich lieferte eine 100%ige Übereinstimmung mit der Literatur.

**Proteinsequenz der Sequenzierung:**

Spacer zwischen MBP-Tag und 3xMyc-Tag:

S S N N N N N N N N N N L

3xMyc-Tag:

G I E G R I S E F R S E Q K L I S E E D L N G E Q K L I S E E D L N G E Q K L  
I S E E D L S R V D

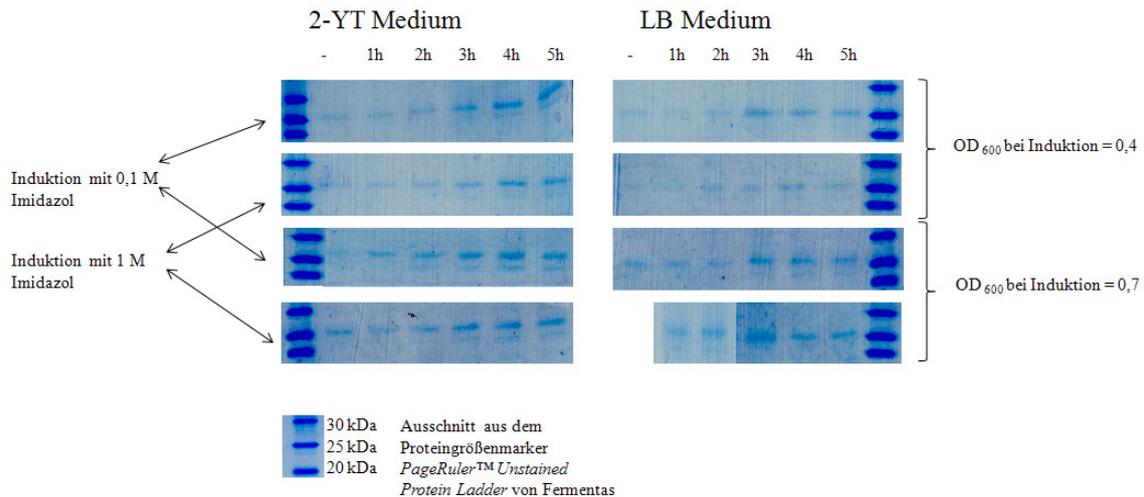
Gli2-Bindedomäne:

K Q E A E V V I Y E T N C H W A D C T K E Y D T Q E Q L V H H I N N E H I H G  
E K K E F V C R W Q A C T R E Q K P F K A Q Y M L V V H M R R H T G E K P H K  
C T F E G C S K A Y S R L E N L K T H L R S H T G E K P Y V C E H E G C N K A  
F S N A S D R A K H Q N R T H S N E K P Y I C K I P G C T K R Y T D P S S L R  
K H V K T V H G P D A H V T K K Q R N D V H V R A P L L K E N G D N E A S A E  
P G G R G P E E S V E A S S T S H T V E D C L H I K A I K T E S S G L**4.2.2 Klonierung der Gli2- Bindedomäne**

Die Gli2-Bindedomäne wurde als Konstrukt mit einem Myc-Tag in den für das Maltose Binding Protein (MBP) kodierenden Vektoren pMAL c2X und pMAL p2X in XL1-Blue-Zellen von einem Kooperationspartner zur Verfügung gestellt. Den Transkriptionsfaktor in den Vektor pQE Trisystem Strep 1 mit einem 8x His-Tag als Fusionsprotein umzuklonieren, hatte zum Ziel, einen möglichen Einfluss des 40 KDa großen MBP auf das Bindevverhalten auszuschließen. Desweiteren sollten eventuell schädigende Aufreinigungsschritte nach proteolytischer Abspaltung des Fusionsprotein umgangen werden. Mittels PCR wurden in die für die Gli2-Bindedomäne kodierende DNA-Sequenz aus dem Vektor pMAL c2X Schnittstellen und der C-terminale His-Tag zur Aufreinigung und zur Detektion eingefügt. Da eine direkte Klonierung des Inserts in den Vektor pQE Trisystem auf Anhieb nicht gelang, wurde es in den Vektor pGEM-T-easy zwischenkloniert. Die sich anschließende Blau-Weiß-Selektion ergab etwa 200 positive Klone, von denen einige Kolonien vereinzelt wurden. Die erfolgreiche Integration des Inserts wurde durch Restriktionsverdau und Gelelektrophorese verifiziert. Da hierzu die selben Enzyme verwendet wurden wie auch später zur Klonierung in den Vektor pQE Trisystem, war gewährleistet, dass intakte Überhänge vorhanden waren. Das betreffende Insert wurde aus dem Gel aufgereinigt und in den neuen Vektor kloniert. Nach der Transformation in M15- und SG13009-Zellen erfolgte die Überprüfung durch Sequenzierung.

**4.2.3 Optimierung der Proteinexpression**

Da bei der Proteinexpression keine Erfahrungen vorlagen, galt es, Parameter wie verwendete Zellen, Medien, Zelldichte und IPTG-Konzentration bei der Induktion sowie die darauf folgende Inkubationsdauer auf ihre Bedeutung hinsichtlich der Proteinausbeute zu untersuchen. Sowohl die Proteinexpression, als auch die Aufreinigung über Ni-NTA-Agarose wurde in einer 96-Loch-Platte mit je 2 ml der Bakterienkultur durchgeführt.



**Abbildung 4.2:** Optimierung der Proteinexpression. SDS-Gele der Proteinexpression in M15 Zellen nach Aufreinigung über Ni-NTA-Agarose. Verglichen wurde die Proteinmenge nicht induzierter Zellen mit der von verschiedenen Varianten der Expression. Der verwendete Marker ist der Page Ruler™ Marker von Fermentas.

Wie in Abbildung 4.2 gezeigt, unterschied sich die erhaltene Proteinmenge in M15-Zellen je nach Ansatz zum Teil erheblich. Mit verlängerter Inkubationsdauer stieg die Ausbeute in allen Ansätzen an, wobei die Zunahme zwischen drei und fünf Stunden nicht sehr groß ausfiel. Es zeigte sich desweiteren, dass eine Induktion bei einer OD 600 von 0,7 etwas bessere Ergebnisse lieferte als die Induktion zu einem früheren Zeitpunkt. Leichte Vorteile brachte auch die Verwendung einer 0,1 M IPTG-Lösung, deutliche hingegen die Wahl von 2 YT als Medium. Der gleiche Versuch wurde auch mit SG13009-Zellen durchgeführt. Da diese jedoch ausnahmslos geringere Proteinmengen produzierten, wurde von ihrer Verwendung abgesehen (SDS-Gele nicht gezeigt). Die Proteinexpression wurde daher unter den folgenden Bedingungen durchgeführt: Verwendet wurden M15-Zellen, die bei einer OD 600 von 0,7 mit 0,1 M IPTG induziert und nach einer Inkubationsdauer von vier Stunden geerntet wurden.

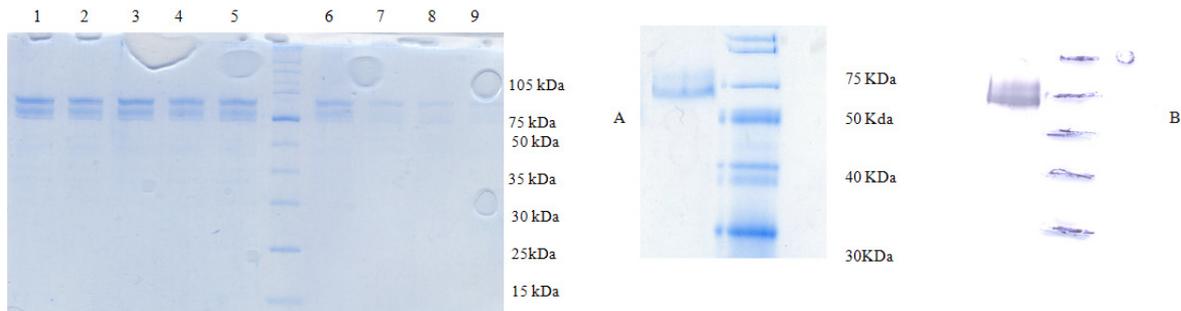
#### 4.2.4 Proteinaufreinigung

Da zur Untersuchung der Bindungseigenschaften von Gli2 durch Microarrayanalyse nur sehr geringe Mengen an Protein benötigt wurden, lag die Priorität auf der Reinheit und besonders auf der Erhaltung der Funktionalität des Proteins.

##### 4.2.4.1 Proteinaufreinigung des MBP-Fusionsproteins über eine Amylose-Harz-Matrix

Das Fusionsprotein aus dem Maltose Binding Protein, dem dreifachen Myc-Tag und der Gli2 Bindedomäne wurde über eine Amylose-Matrix entsprechend dem Handbuch des pMAL Expressionssystem von New England Biolabs aufgereinigt. Nach den Waschschrinen wurde das Protein in 15 ml Säulenpuffer mit 10 mM Maltose eluiert, wobei Fraktionen von je 0,5 ml

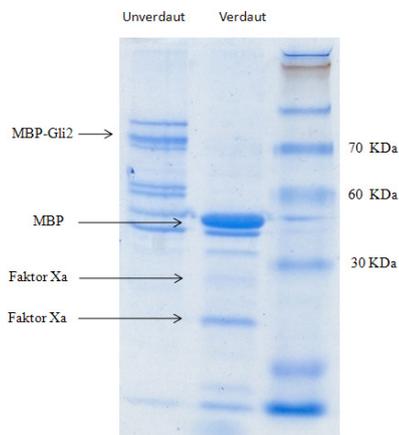
gesammelt wurden. Bei anschließender Analyse der Eluate zeigte sich, dass das Protein bereits nach insgesamt 4 ml Elutionspuffer vollständig und in reiner Form von der Matrix gelöst worden war. Neben der erwarteten Bande des Fusionsproteins bei 76 KDa fällt eine zweite, nicht erwartete Bande. Diese Bande kann durch die MALDI Analyse erklärt werden. Neben dem Volllängenprotein wurde noch ein Protein der gleichen Sequenz exprimiert, das 5-10 KDa kleiner ist. Da *E. coli* - Bakterien sehr seltene Codons nicht exprimieren können, kommt es an diesen Stelle oft zum Abbruch der Expression und somit zu mehreren Produkten (siehe Nachweis von Gli2 mittels MALDI).



**Abbildung 4.3:** Aufreinigung von Gli2 als Fusionsprotein mit MBP im Vektor pMAL c2X. Gezeigt sind die ersten neun Fraktionen der Elution in Säulenpuffer mit 10 mM Maltose nach sämtlichen Waschschritten, sowie der Full-Range Rainbow Proteinmarker von GE Healthcare (links). SDS-Page (A) und Western Blot (B) von MBP-Gli2 (rechts).

#### 4.2.4.2 Abspaltung des MBP-Tags

Es war dank des Vektor-Designs möglich, den für die Aufreinigung der Fusionsproteine nötigen MBP-Tag mittels Faktor XA Protease abzuspalten. Diese Protease schneidet spezifisch die Aminosäuresequenz Ile-Glu/Asp-Gly-Arg. Hierzu wurden die 10 µg Fusionsprotein in 40 µl Abspaltpuffer gelöst und mit 4 U Faktor Xa Protease 3 h bei RT inkubiert. Anschließend wurde der Ansatz auf einem SDS-Gel analysiert. Das Gel ist in Abbildung 4.4 gezeigt.



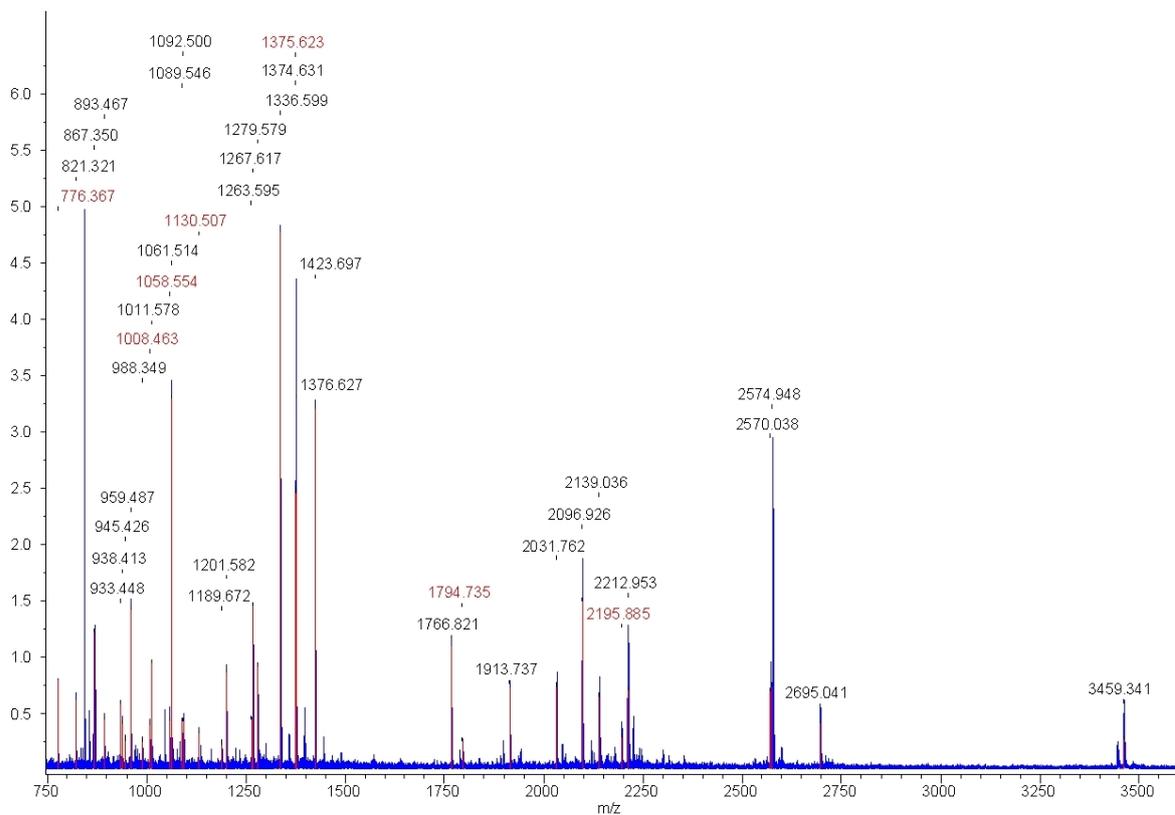
**Abbildung 4.4:** Faktor Xa Verdau von MBP-Gli2. In der ersten Spalte ist das Gesamtprotein mit einigen Nebenprodukten zu sehen. Spalte zwei zeigt das geschnittene Produkt. Es ist nur der MBP-Tag und die zur Abspaltung benutzte Faktor Xa Protease zu sehen. Sie besteht aus zwei Untereinheiten.

Mittels MALDI-Analyse konnten die in Abbildung 4.4 mit Pfeilen markierten Fragmente

eindeutig identifiziert werden. Die Bindedomäne von Gli2 konnte nach dem Verdau auf dem Gel als auch in der MALDI-Analyse nicht detektiert werden. Auch eine Stabilisierung der Gli2-Bindedomäne während und nach der Abspaltung mittels Detergenzien und Polymeren, die die hydrophoben Teile der Bindedomäne stabilisieren sollten, war nicht erfolgreich.

#### 4.2.4.3 MALDI-TOF Analyse

Das exprimierte MBP-3xMyc-Gli2 Fusionsprotein wurde mittels *MALDI-TOF*-Massenspektrometrie auf seine Struktur hin untersucht. Nach einem Trypsinverdau wurden die einzelnen Protein-Fragmente detektiert (siehe Abbildung 4.2.)



**Abbildung 4.5:** MALDI Spektrum von MBP-3xMyc-Gli2

Nach der erfolgreichen Detektion wurden die Fragmente mit einer Massenspektrometrie-Datenbank abgeglichen. Die Tabelle 4.1 und 4.2 zeigen die detektierten und von der Datenbank bestätigten Fragmente des Maltose Binding Proteins (MBP) und des Gli2 Fragmentes.

**Literatursequenz MBP:**

MKLLFAIPLVVPFYSHSADMKIEEGKLVIIWINGDKGYNGLAEVGGKFEKDTGKIKVTV  
 EHPDKLEEKFPQVAATGDGPDIIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWD  
 AVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQE  
 PYFTWPLIAADGGYAFKYENKDYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYS  
 IAEEAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGLSAGINAA  
 SPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKG  
 EIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEG  
 RISEFGSSRVDLQASLALAVVLQRDQWENPGVTQLNRLAAHPPFASWRNSEEARTDRP  
 SSQLRSLNGEWQLGCGGG

**Literatursequenz Gli2:**

TNCHWADCTKEYDTQEQLVHHINNEIHGEKKEFVCRWQACTREQKPFKAQYMLVVHM  
 RRHTGEKPHKCTFEGCSKAYSRLNLKTHLRSHTGEKPYVCEHEGCNKAFSNASDRAK  
 HQNRTHSNEKPYICKIPGCTKRYTDPSSLRKHVKTVHGPDHAV

M/z	SEQUENZ
892.459731	DLLPNPPK
944.418785	DVGVDNAGAK
958.479672	TAVINAASGR
1010.57062	DKPLGAVALK
1091.492561	IAATMENAQK
1188.664472	AGLTFLVDLIK
1200.574526	TWEEIPALDK
1266.609776	LYPFTWDAVR
1335.591703	SYEEELAKDPR
1422.690215	VTVEHPDKLEEK
1765.813963	FGGYAQSGLLAEITPDK
1912.729598	HMNADTDYSIAEEAFNK
2095.918902	EFLNYLLTDEGLEAVNK
2138.028228	GQPSKPFVGLSAGINAASPNK
2211.946179	FPQVAATGDGPDIIIFWAHDR
2573.94121	DAQTNSSSNNNNNNNNNNLGIEGR
3458.3334	QTVDEALKDAQTNSSSNNNNNNNNNNLGIEGR

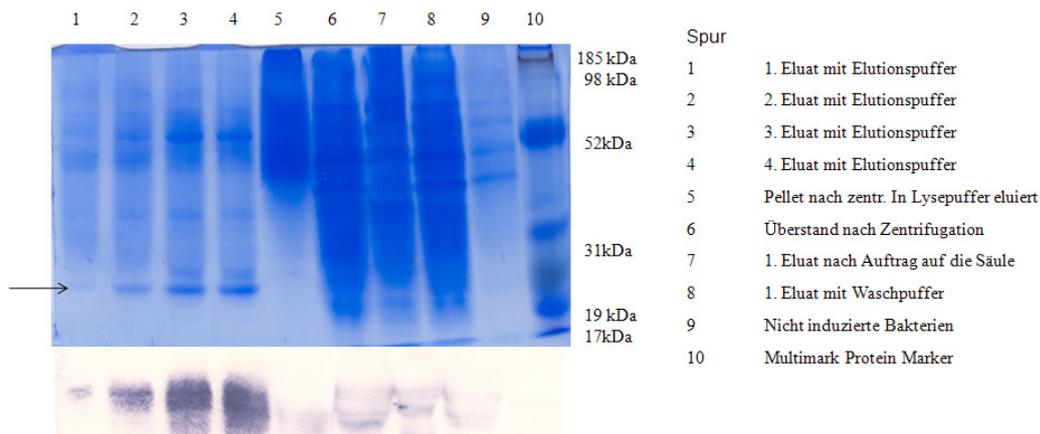
**Tabelle 4.1:** MALDI: Detektierte Fragmente des Maltose bindenden Proteins nach einem Trypsin-Verdau

M/z	SEQUENZ
820.313263	WQACTR
866.342979	AFSNASDR
932.440665	HTGEKPHK
937.405697	YTDPSLR
987.34147	CTFEGCSK
1088.538948	RHTGEKPHK
1262.587543	AQYMLVVHMR
1278.572143	AQYMLVVHMR
1375.620188	THSNEKPYICK
2030.754531	SHTGEKPYVCEHEGCNK
2569.03105	EYDTQEQLVHHINNEHIHGEK

**Tabelle 4.2:** MALDI: Detektierte Fragmente vom Gli2 nach einem Trypsin-Verdau

#### 4.2.4.4 Proteinaufreinigung des His-Fusionsproteins über Ni-NTA-Agarose

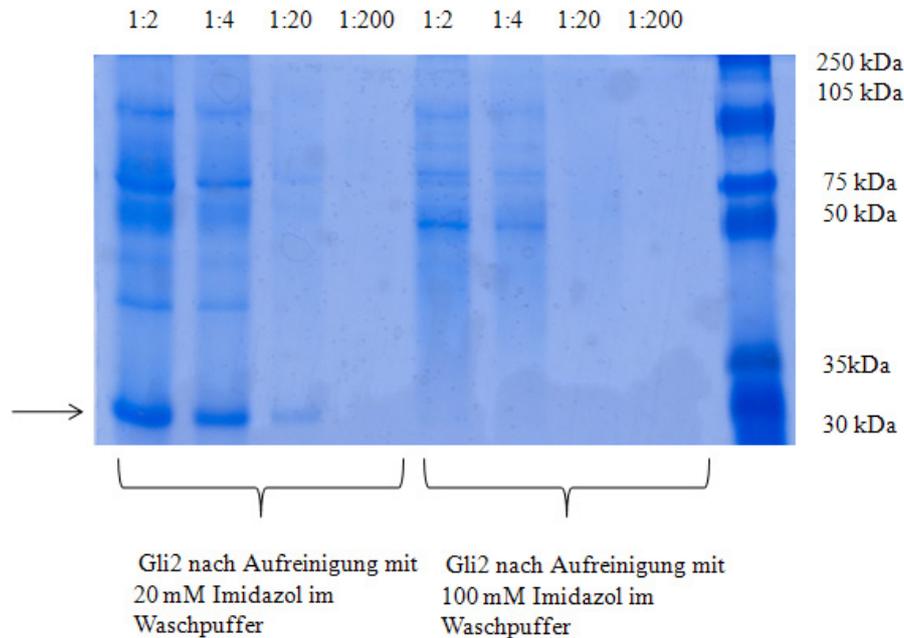
Das Fusionsprotein von Gli2 und dem 8x His-Tag aus dem Vektor pQE Trisystem Strep 1 wurde zunächst über Ni-NTA Agarose aufgereinigt. In Abbildung 4.6 beinhalten die Spuren 1-4 die Elutionsfraktionen. Hier wird deutlich, dass Gli2 nicht unmittelbar aus der Matrix gelöst wurde. Auch waren nach dem Waschen mit 20 mM Imidazol noch erhebliche Verunreinigungen vorhanden und dies obwohl die Spuren 7 und 8 mit den Waschfraktionen überladen waren. Der Western-Blot zeigt die erfolgreiche Proteinexpression, wobei das Protein auch mit His-Tag löslich blieb. Dies wird aus der fehlenden Bande von Gli2 in Spur 5 deutlich.



**Abbildung 4.6:** Aufreinigung von Gli2 als Fusionsprotein mit 8x His-Tag im Vektor pQE Trisystem Strep 1. Gezeigt sind verschiedene Fraktionen der Elution, der Waschschritte mit 20 mM Imidazol und weitere Fraktionen. Zusätzlich der mit einem Pfeil markierte Ausschnitt des Gels als Western – Blot.

In der Folge wurde durch Erhöhen der Imidazolkonzentration im Waschlösung stringenter gewaschen, was jedoch nicht zu befriedigenden Ergebnissen führte. In Abbildung 4.7 ist das

Ergebnis mit Waschpuffer einer Konzentration von 100 mM Imidazol gezeigt. Die Bedingungen waren zu stringent, sodass Gli2 schon vor der eigentlichen Elution aus dem Nickel-Komplex gelöst wurde. Auch Imidazolkonzentrationen von 40 mM und 60 mM gaben keine zufriedenstellende Resultate (Daten nicht gezeigt). Es wurde daher entschieden, die Aufreinigung mittels FPLC zu optimieren.

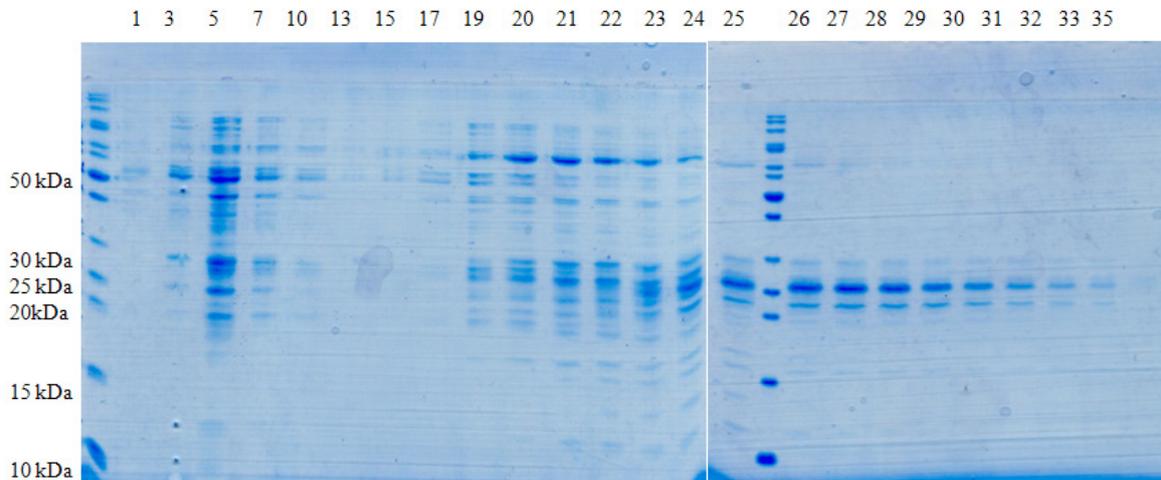


**Abbildung 4.7:** Proteinaufreinigung von Gli2 mit His-Tag unter verschiedenen Waschbedingungen. Gezeigt sind vier Verdünnungen, nachdem das Protein aufgereinigt und konzentriert worden war. Ganz rechts der Full-Range Rainbow Proteinmarker von GE Healthcare.

#### 4.2.4.5 Proteinaufreinigung des His-Fusionsproteins mittels *Fast Protein Liquid Chromatography* (FPLC)

Mit Hilfe der *Fast Protein Liquid Chromatography* (FPLC) wurde das Fusionsprotein ebenfalls über einen Nickel-Chelat-Komplex aufgereinigt. Der entscheidende Vorteil ist jedoch die Möglichkeit, bei der FPLC einen maschinell regulierten, kontinuierlichen Gradienten aus verschiedenen Puffern verwenden zu können. In diesem Fall lief der gleichzeitig zum Waschen und zur Elution verwendete Gradient über ein Volumen von 20 ml. Der Gradient erstreckte sich von 20 mM Imidazol zu Beginn bis zu 250 mM Imidazol zum Ende der Elution. In 4.8 wird deutlich sichtbar, dass mit steigender Imidazolkonzentration zwar auch Teile von Gli2 eluiert werden, vor allem aber die Reinheit des gewünschten Proteins zunimmt. Zum weiteren Gebrauch wurden jeweils die reinsten Fraktionen, in diesem Fall die Fraktionen 30-33, aufkonzentriert und für die Microarrayanalyse verwendet. Wie jedoch zuvor bei Gli2 mit MBP-Tag schon beobachtet, gelang trotz offensichtlich erfolgreicher Aufreinigung die Reduktion auf eine einzelne Bande nicht. Neben der erwarteten Bande bei 27 kDa zeigte sich diesmal eine Weitere bei etwa 22 kDa. Interessanterweise scheinen die Proteine in beiden

Banden dieselbe Affinität zur  $\text{Ni}^{2+}$  Säule zu haben. Die Erklärung hierfür liegt ebenfalls bei den seltenen Codons, die *E. coli* - Bakterien schlecht exprimieren können. Es wurde zwei identische Proteine exprimiert, wobei eines etwa 5 KDa kleiner ist.



**Abbildung 4.8:** Proteinaufreinigung von Gli2 mit His-Tag über FPLC. Gewaschen und eluiert wurde mit einem kontinuierlichen Imidazol-Gradienten. Insgesamt wurden 42 Fraktionen mit je 0,5 ml Volumen gesammelt. Die Zahlen über den Spalten entsprechen der jeweiligen Fraktion. Der verwendete Marker ist der Page Ruler™ Marker von Fermentas.

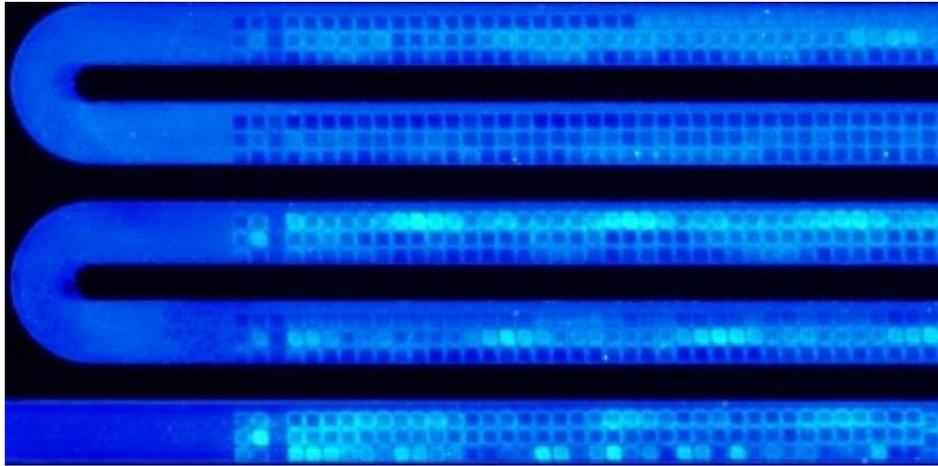
#### 4.2.5 Bestimmung der Protein-DNA Wechselwirkungen

Für die ersten Microarray Versuche wurde das ältere Geniom 1X System benutzt. Dieses System ist auf 4 Arrays mit weniger Spots limitiert, bietet jedoch für die Entwicklung der Methode entscheidende Vorteile. Der wohl Wichtigste war die manuelle Unterbrechung der Synthese. Somit konnten die Chemikalien nachgefüllt und Oligonukleotide bis zu einer Länge von 48 Basenpaaren synthetisiert werden. Mit dem neueren Geniom 2X System konnten zu diesem Zeitpunkt nur Oligonukleotide mit bis zu 30 Basen synthetisiert werden. Ein weiterer Vorteil war die manuelle Bedienung der Waschprogramme sowie die manuelle Einstellung der Flussrate während der Waschprozesse. Bei dem Geniom 2X System werden diese Einstellungen in vorgefertigten Profilen zur Verfügung gestellt.

Die langen Oligonukleotide sind für diese Methode von essentieller Bedeutung. Wie am Anfang des Kapitels bereits erwähnt, erhält man die doppelsträngige DNA durch die Synthese eines *Hairpins*, der sich nach 15 minütigen Erhitzen in 1x PBS und langsamen Abkühlen von selbst ausbildet. Hierfür sind bei einer 9 bp langen Konsensus-Sequenz, 6 Randnukleotide sowie ein 6xT-Hairpin mindestens 48 Basen notwendig. Dies konnte mit dem Geniom 2X zu Beginn der Arbeit nicht erreicht werden.

Die ersten Experimente wurden mit dem Geniom 1X Gerät durchgeführt. Die Synthese eines 48 bp langen Oligonukleotids konnte durch das Stoppen der Synthese und das Auffüllen

der Chemikalien optimiert werden. Das in *E. coli* - Bakterien exprimierte Fusionsprotein MBP-Gli2 wurde aufgereinigt und an die verschiedenen doppelsträngigen DNA Sequenzen gebunden. Dies konnte durch die Markierung mit einem biotinylierten Anti-cMyc-Antikörper (2 h, 0,6 nM) und eines Streptavidin-Phycoerythrin-Konjugates (15 min; 80 nM) nachgewiesen werden. Ein typisches Ergebnis dieser Experimente zeigt Abbildung 4.9



**Abbildung 4.9:** Typisches Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA-Sequenzen auf einem mit dem Geniom 1X synthetisierten Chip.

Die Waschprozesse während der verschiedenen Inkubationen wurden ebenfalls optimiert. 2-facher Austausch des Array-Volumen pro Minute ergab hier die besten Ergebnisse (Daten nicht gezeigt). In Kooperation mit febit wurde danach ein Hard- und Softwareupdate an dem Geniom 2x System durchgeführt. Es war nun auch hier möglich Oligonukleotide bis zu einer Länge von 60 bp zu synthetisieren. Mit den Erkenntnissen aus den Versuchen mit dem 1X System konnten die notwendigen Inkubations- und Waschprofile auf dieses System übertragen werden.

Folgenden Wasch- und Inkubationsschritte wurden für die folgenden Experimente verwendet:

- Waschen (geringer Fluss: 2x Arrayvolumen pro Minute)
- Antikörperinkubation (60 min)
- Waschen (geringer Fluss: 2x Arrayvolumen pro Minute)
- SAPE Inkubation (15 min)
- Waschen (geringer Fluss: 2x Arrayvolumen pro Minute)
- Detektion

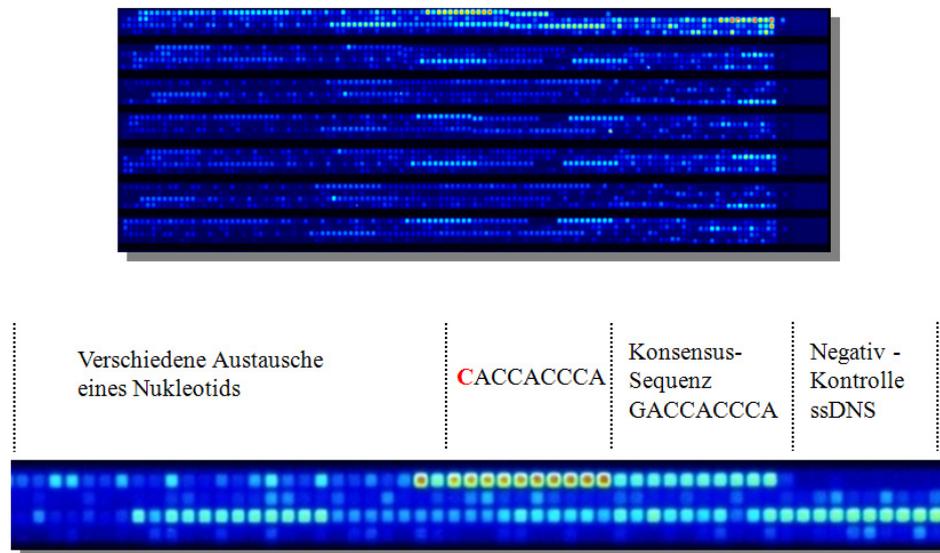
Nachfolgend sind die Ergebnisse gezeigt, die mit dem Geniom 2X System erzielt wurden.

#### 4.2.5.1 Etablierung und Optimierung der Microarrayexperimente

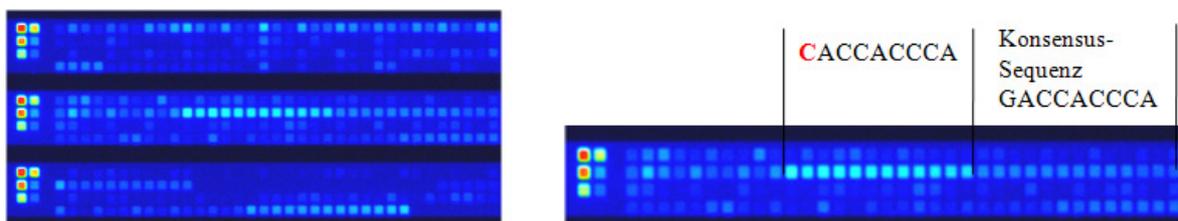
##### Protein-Inkubation von His-Gli2 und MBP-3xMyc-Gli2

Ziel war es, ausgehend von der neun Basenpaare langen Konsensus-Bindesequenz von Gli2 alle möglichen Permutationen von einem bzw. zwei Nukleotiden auf ihre Bindungseigenschaften zu untersuchen. Der Nachweis des gebundenen Proteins erfolgte bei Gli2 mit His-Tag mithilfe eines biotinylierten Penta-His-Antikörper. Bei Gli2 mit MBP-Tag wurde ein gegen das 3xMyc-Epitop des Fusionsproteins gerichteter biotinylierter Antikörper verwendet. Ein detektierbares Fluoreszenzsignal wurde durch Bindung von Streptavidin-Phycoerythrin (SAPE) an das Biotin erhalten.

Die Abbildungen 4.10 und 4.11 zeigen typische Ergebnisse einer erfolgreichen Detektion der Bindung des Transkriptionsfaktors an die Haarnadelstruktur der DNA.



**Abbildung 4.10:** Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA-Sequenzen.



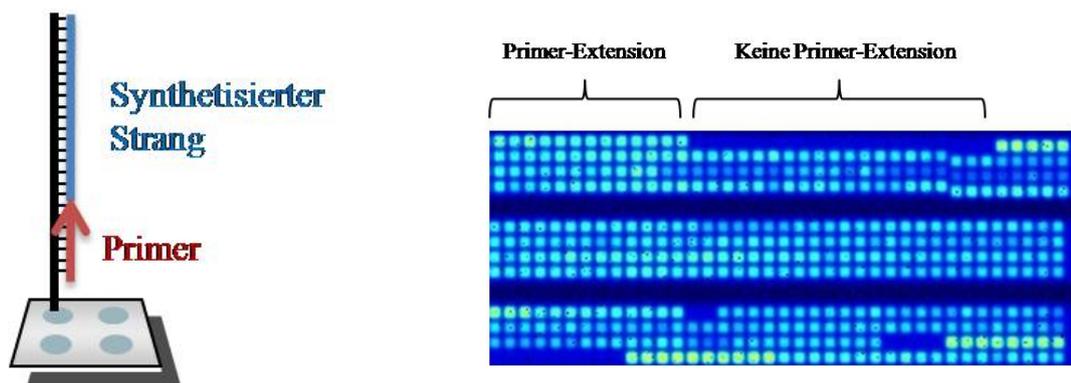
**Abbildung 4.11:** Ergebnis der Bindung von Gli2 mit His-Tag an verschiedene DNA-Sequenzen

Der Vergleich beider Ausschnitte zeigt deutliche Parallelen auf, obwohl es sich um verschiedene Arraydesigns handelt. So kam es in beiden Fällen zu einer Bindung an die aus der Literatur bekannte Konsensus-Sequenz GACCACCCA. Als außergewöhnlich stark fiel die Bindung an die Sequenz CACCACCCA auf. Sie wurde deshalb, wie die Konsensus-Sequenz, mit 10 Replikaten

synthetisiert. Aber auch andere Variationen der Konsensus-Sequenz erlaubten noch immer eine relativ starke Bindung des Zinkfinger-Motivs von Gli2 an den synthetisierten Doppelstrang. Es gab aber auch eine Vielzahl von Variationen der Konsensus-Sequenz, deren veränderte Basenabfolge eine Bindung nicht mehr zuließ. Hierfür stehen Spots, deren Signalintensität nicht oder nur kaum höher ist als der unspezifische Hintergrund. In Abbildung 4.10 ist auch eine einzelsträngige Konsensus-Sequenz als Negativkontrolle gezeigt, an die keine Bindung erfolgte.

#### 4.2.5.2 *Primer-Extension*

Eine weitere Methode zur Herstellung doppelsträngiger DNA ist die Synthese des 2. Stranges mittels *Primer-Extension*. Hierzu wird ein universeller Primer gewählt, der komplementär zum unteren Ende des auf der Oberflächen kovalent gebundenen Nukleotides ist. Durch Zugabe einer Polymerase und dNTP's kann dann der 2. Strang synthetisiert werden (siehe Abbildung 4.12). Durch den Einbau markierter Cy5-dCTP's kann die Verlängerung des Primers detektiert werden (siehe Abbildung 4.12).



**Abbildung 4.12:** Schematische Darstellung der Primer-Extension auf dem Chip (links). Ergebnis einer PCR auf dem Chip mit Cy5 markierten dCTP's. An den dunklen Stellen war der Primer nicht komplementär, somit konnte der Strang nicht verlängert werden (rechts).

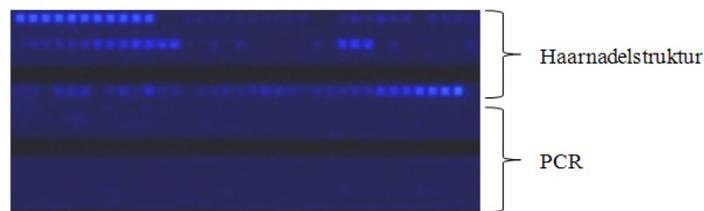
Folgendes Protokoll wurde für die Primer-Verlängerung auf dem Chip verwendet. Die Reaktion wurde in der externen Hybridisierungshalterung durchgeführt.

- Denaturierung bei 85°C.
- Inkubation bei 47°C für 30 Minuten
- Waschen mit 6x SSPE (RT); 0,5x SSPE (45°C)

VOLUMEN	REAGENZIEN
4µl	10 x Thermo Sequenase reaction buffer (Amersham Pharmacia Biotech.Inc.)
8,35 µl	dH <sub>2</sub> O
0,15 µl	RG-3'Primer (1 pmol/µl endconc.)
0,75 µl	dTTP, dGTP, dATP (1 mM: 1000 pmol/µl)
0,5 µl	dCTP (1 mM)
0,25 µl	Cy5 dCTP (1 mM)

**Tabelle 4.3:** Protokoll des *Primer-Extension* Ansatzes.

Ein Binden der Transkriptionsfaktoren war an die mit PCR hergestellten doppelsträngigen DNA nicht möglich. Abbildung 4.13 zeigt einen Ausschnitt eines Arrays, auf dem Gli2 parallel an Sonden mit Haarnadelstruktur und durch Primer-Extension verlängerte DNA inkubiert und detektiert wurde. Es konnten nur Signale bei der Haarnadelstruktur nachgewiesen werden.



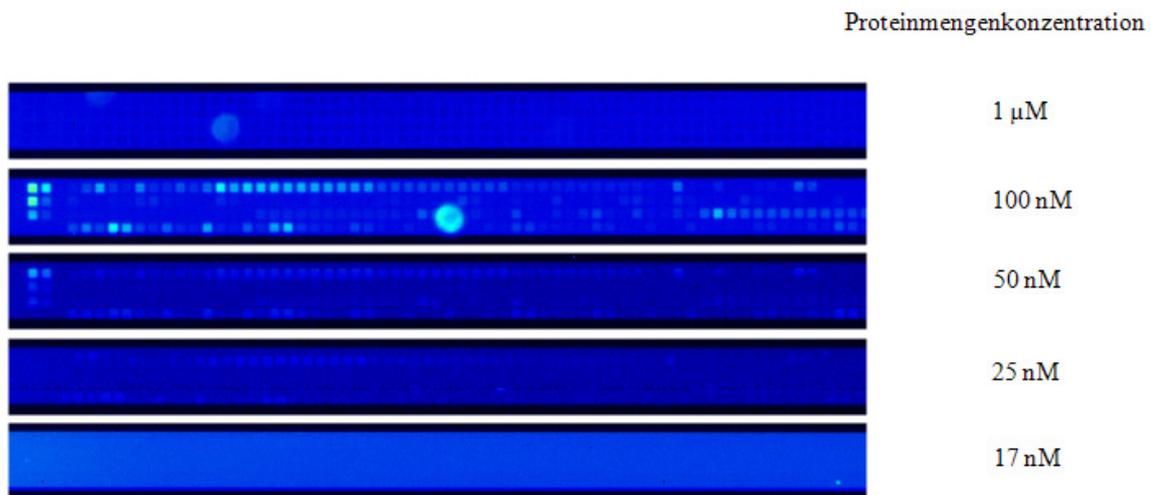
**Abbildung 4.13:** Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA-Sequenzen einer Haarnadelstruktur und an eine mit PCR hergestellte Nukleotidsequenz. Die sichtbaren Spots zeigen die Haarnadelstruktur. An den mit PCR hergestellten Sequenzen konnte keine Proteinbindung detektiert werden.

Da die Haarnadelmethode bessere Ergebnisse liefert, wurde auf die weitere Entwicklung der PCR-Methode verzichtet.

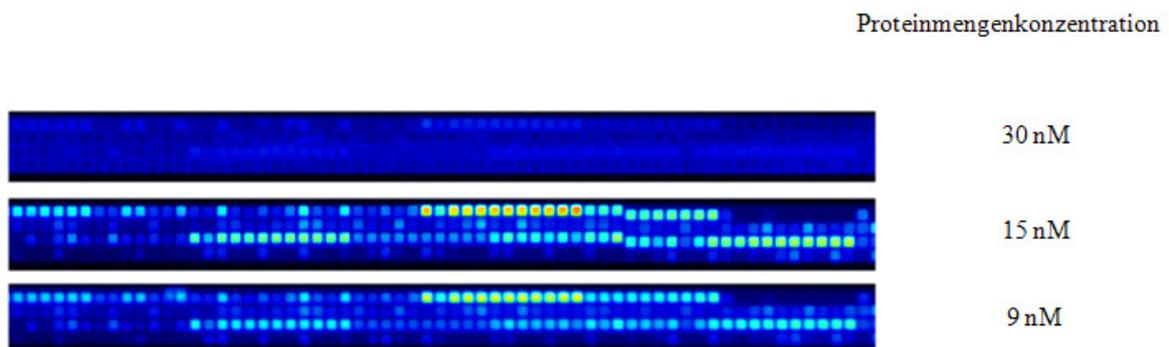
#### 4.2.5.3 Optimierung der Proteinmenge

Es wurden viele Versuche unternommen, die Reproduzierbarkeit der Experimente durch Veränderung der aufgetragenen Proteinmenge zu verbessern. Hierdurch und auch durch Überprüfen der Funktionalität des Proteins beispielsweise durch Electrophoretic Mobility Shift Assays (EMSA) gelang dies zunächst jedoch nur teilweise (Daten nicht gezeigt). Es zeigte sich aber eine deutliche Abhängigkeit der beobachteten Ergebnisse in Bezug auf die eingesetzte Proteinmenge. Wie die Abbildungen 4.14 und 4.15 veranschaulichen, konnte gezeigt werden, dass es eine optimale Menge von eingesetztem Protein gibt. Sowohl höhere als auch geringere Konzentrationen lieferten schlechtere bis gar keine Ergebnisse. Im Fall von Gli2 als Fusionsprotein mit dem 8x His-Tag lag die eingesetzte Stoffmengenkonzentration zwischen 50 und 100 nM, mit MBP als Fusionsprotein zwischen 9-30 nM. Die zur Berechnung notwendige

Proteinkonzentration wurde mit dem Nanodrop bestimmt und das Molekulargewicht aufgrund der Aminosäuresequenz berechnet. Das pro Array aufgetragene Volumen betrug 8  $\mu\text{l}$ .



**Abbildung 4.14:** Abhängigkeit der Signalintensität von der aufgetragenen Proteinmenge im Fall von Gli2-His.



**Abbildung 4.15:** Abhängigkeit der Signalintensität von der aufgetragenen Proteinmenge im Fall von Gli2-MBP.

Unterschiedliche Signalintensitäten zwischen den Abbildungen 4.14 und 4.15 beruhen auf unterschiedlichen Belichtungszeiten. Durch die Normalisierung gegen die Konsensus-Sequenz sind die verschiedenen Chips vergleichbar.

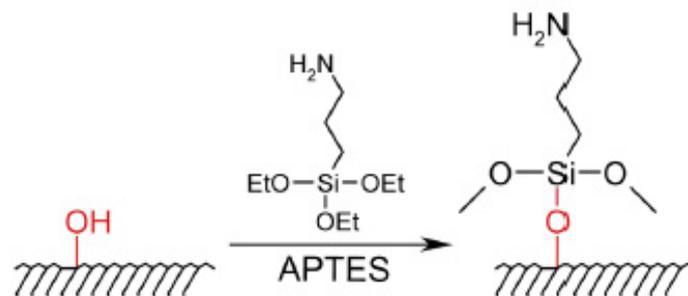
#### 4.2.5.4 Optimierung der DNA-Konzentration

Da die Reproduzierbarkeit der Microarrayexperimente auf Proteinebene alleine nicht zu lösen war, galt es einen anderen Ansatz zu finden. Auffällig waren „negative“ Spots, die bei den Versuchen auftraten und deren Intensität geringer als der lokale Hintergrund war (Abbildung 4.18).

Es wurde vermutet, dass eine zu hohe DNA-Dichte der Spots eine Bindung des Transkriptionsfaktors sterisch verhindert. Die Dichte der auf dem Array synthetisierten Oligonukleotide konnte jedoch nicht ohne Weiteres geändert werden. Daher wurde ein doppelsträngiges Oligonukleotid mit der Sequenz CACCACCCA, in verschiedenen Konzentrationen auf Glas-Objektträger gespottet und hinsichtlich der Bindefähigkeit des Fusionsproteins untersucht.

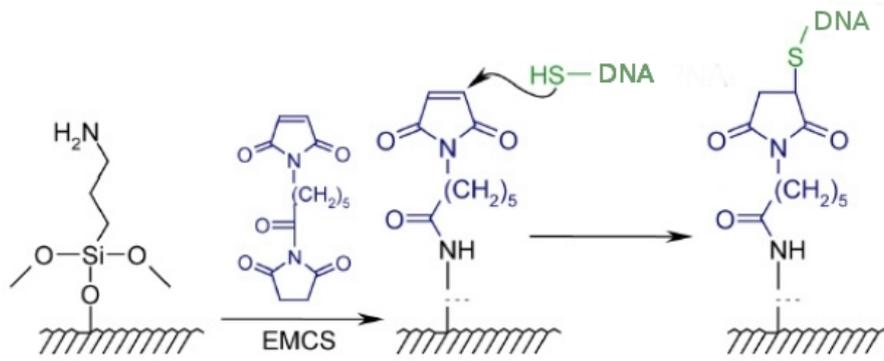
### Microarrays mit dem Gesim Nanoplotter

Da Glasobjektträger gegenüber Aktivierungen mit 6-Maleinimidohexansäure-N-succinimidylester (EMCS) inert sind, wurden sie zuerst mit einer Silanschicht überzogen (siehe Abbildung 4.16). Hierzu wurden die unbehandelten Glasoberflächen mit Dimethylsulfoxid (DMSO), Ethanol und Wasser gründlich gewaschen. Anschließend wurden sie für eine Stunde bei Raumtemperatur in 10% NaOH (w/w), gefolgt von weiteren 15 min im Ultraschallbad, angeätzt. Nach erneutem Waschen mit Wasser und Ethanol wurden die Oberflächen für eine Stunde in die Silanisierungslösung gegeben (1 ml 3-Aminopropyl-triethoxysilan (APTES) in 20 ml 95%igem Ethanol) und in der gleichen Lösung noch zusätzliche 15 min ins Ultraschallbad gestellt. Nach zwei Waschschrritten in Ethanol und einem in Wasser wurden die Oberflächen mit Stickstoff getrocknet und für 20 min bei 110°C gebacken. Die fertig silanisierten Glasobjektträger wurden bei 4°C bis zur weiteren Verwendung aufbewahrt.



**Abbildung 4.16:** Silanisierung von Silizium- oder Glasoberflächen mit 3-Aminopropyltriethoxysilan (APTES)

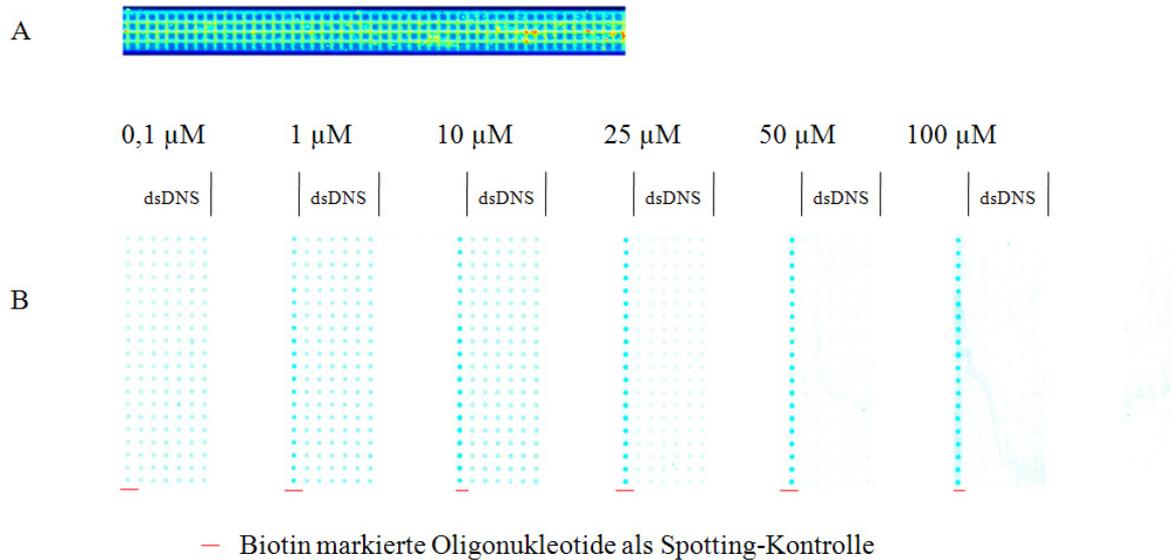
Vor dem Spotten mussten die mit Aminosilan beschichteten Objektträger aktiviert werden (siehe Abbildung 4.17). Dies wurde unmittelbar vor dem Spotten mit 60  $\mu$ l einer EMCS-Lösung (20 mM EMCS in 10% Diisopropylethylamin in DMF) durchgeführt. Mit einem Deckglas geschützt, inkubierten sie für 3 Stunden bei Raumtemperatur und wurden danach zweimal mit wasserfreiem DMF und einmal mit wasserfreiem Aceton gewaschen und schließlich mit Stickstoff getrocknet. Gespottet wurde ein Oligonukleotid mit Thiol-Modifikation am 5'-Ende, was die Bindung an die zuvor aktivierte Oberfläche erlaubte.



**Abbildung 4.17:** Die Aktivierung der silanisierten Oberflächen mit 6-Maleinimido-hexansäure-N-succinimidylester (EMCS, blau) erlebte die Kopplung der DNA über die Thiolgruppe (grün).

Die gespotteten DNA-Konzentrationen variierten von 0,1 bis 100  $\mu\text{M}$ , wobei 2 M Betain in Wasser zum Verdünnen benutzt wurde. Betain soll das Verdunsten und Eintrocknen der DNA nach dem Spotten verhindern (Diehl et al. (2001)). Die Objektträger wurden über Nacht bei Raumtemperatur zum abreagieren inkubiert. Noch freie Aminogruppen auf der Oberfläche wurden am nächsten Tag zum Deaktivieren in 1 mM 3-Mercaptopropionsäure in Ethanol für 30 Minuten inkubiert. Die Objektträger wurden dreimal in Ethanol, einmal in DMF und erneut in Ethanol gründlich gewaschen und in 50 mM Succinimidylanhydrid mit 150 mM 1-Methylimidazol in Dichlorethan für 2 Stunden auf dem Schüttler geblockt. Nachdem sie zweimal mit Dichlorethan gewaschen und die Oberfläche weitere zweimal mit 90°C heißer 5 mM Natriumphosphatlösung mit 1% SDS gespült wurden, folgte der letzte Inkubationsschritt für 10 min in 100°C heißem Wasser. Die Ausbildung der Haarnadelstruktur erreicht man durch 5 minütiges Erhitzen auf 95°C in PBS-Puffer und anschließendes langsames Abkühlen auf RT. Nach dem Trocknen waren die Objektträger bereit zur Weiterverarbeitung. Hierzu wurde die Glasoberfläche zunächst mit einer 4% Milchpulverlösung in 1x PBS für 1 Stunde bei Raumtemperatur geblockt. Gli2 mit His-Tag wurde in 1x Bindepuffer mit 4% w/v Milchpulver auf eine Konzentration von etwa 130 nM verdünnt und auf den Array aufgebracht. Diese 30 minütige Inkubation und die weiteren Schritte wurden auf dem Slidebooster (Advalytix) durchgeführt, um eine bessere Durchmischung zu gewährleisten. Anschließend wurde 1 Stunde mit einem biotinylierten Anti-His-Antikörper (6 nM) inkubiert und letztlich für 15 Minuten mit Streptavidin R-Phycoerythrin (SAPE 60 nM) in 1x Bindepuffer inkubiert. Nach kurzem Waschen erfolgte die Detektion mit dem Scan Array 3000/5000. Zur Kontrolle des Spotting-Vorgangs wurde ein biotinyliertes Oligonukleotid ohne die für das Ausbilden der Haarnadelstruktur notwendigen komplementären Basenpaare in der selben Konzentrationen verwendet.

Es stellte sich heraus, dass die Bindung des Transkriptionsfaktors durch zu hohe DNA-Konzentrationen sterisch behindert wird. Bei geringen DNA-Mengen wird nur ein schwaches Signal sichtbar. Bei steigender DNA-Konzentration können mehr Proteinmoleküle binden, das gemessene Signal wird stärker. Ab einer DNA-Konzentration von ca. 10  $\mu\text{M}$  fällt es wieder und verschwindet letztlich (Abbildung 4.18).

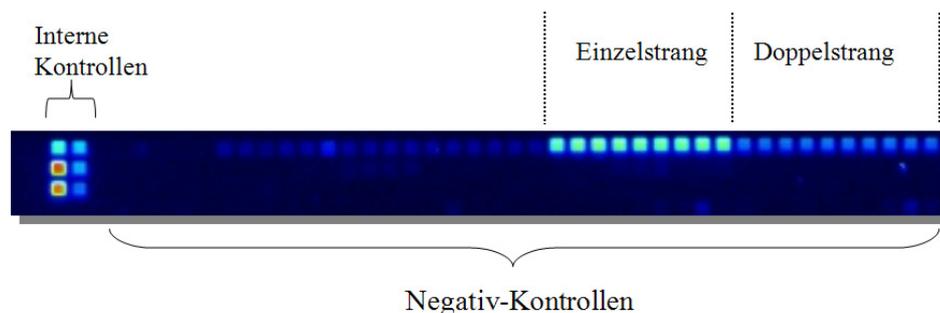


**Abbildung 4.18:** Ausschnitt eines mit Protein überladenen Genom-Chips (A). Falschfarben-Bild des Scans der Bindung von Gli2 an die Sequenz CACCACCA in verschiedenen Konzentrationen (B).

Aufgrund dieser Ergebnisse wurde in Kooperation mit der febit biomed GmbH ein Chip mit 1/10 der ursprünglichen DNA-Dichte entwickelt und für die weiteren Experimente verwendet. Mit diesem Chip wurde das Problem der mangelnden Reproduzierbarkeit behoben.

#### 4.2.5.5 Kontrolle der Haarnadelstruktur

Um zu überprüfen, wie effizient sich die Haarnadelstruktur ausbildet, wurde ein biotinyliertes Oligonukleotid verwendet, das zum 5' Ende der synthetisierten DNA komplementär ist. Bei einer 100%-igen Bildung der Haarnadelstruktur dürfte deshalb keine Hybridisierung und somit kein Signal erkennbar sein. Auf dem Chip war zum Vergleich DNA synthetisiert worden, die keinen Doppelstrang ausbilden konnte, wohl aber die Bindung des Biotin-markierten Oligonukleotids erlaubte. Wie in Abbildung 4.19 gezeigt, kam es auch zu einer Bindung an die Sequenzen der Haarnadelstruktur. Dies verdeutlichte, dass nicht alle Oligonukleotide besagte Struktur ausbilden. Die Berechnung der Effizienz der Haarnadelbildung aus den Signalintensitäten von Doppel- und Einzelstrang ergab einen Wert von ungefähr 80%.



**Abbildung 4.19:** Kontrolle zur Effizienz der Ausbildung der Haarnadelstruktur.

#### 4.2.5.6 Optimierung des Arraydesigns

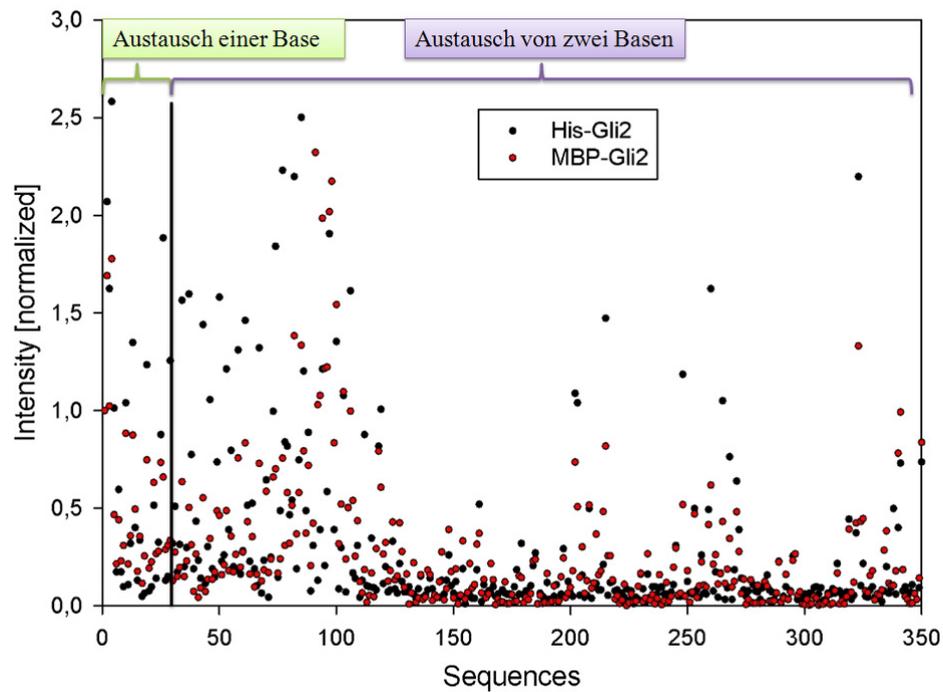
Es wurden unten gezeigte Variationen der gespotteten Oligonukleotide auf mögliche sterische Hindernisse bei der Bindung des Transkriptionsfaktors getestet. Dies betraf den Abstand der Gli2 Bindedomäne zur Oberfläche des Chips und zur Haarnadelstruktur. In roter Schrift ist die Bindedomäne gezeigt, in Grün die Haarnadel-Schleife aus Thymidin-Nukleotiden, daneben auf beiden Seiten entweder 4 oder 6 Basen und am 3'Ende unterschiedliche Längen des T-Spacers (Abbildung 4.20).



**Abbildung 4.20:** Arraydesign des Geniom Chips für das Fusionsprotein Gli2.

#### 4.2.5.7 Analyse der Microarraydaten

Die Verarbeitung der Rohdaten erfolgte durch eine im Geniom integrierte Software. Dabei wurde der lokale unspezifische Hintergrund subtrahiert. Die so erhaltenen Signalintensitäten wurden in Excel weiter verarbeitet. Normalisiert wurde gegen die aus der Literatur bekannte Konsensus-Sequenz. Innerhalb eines Arrays gab es mehrere Replikate aus Konsensus-Sequenz und permutierten Varianten. Der Vergleich mit diesen Werten diente zur Berechnung der Standardabweichung. Das Protein verteilte sich innerhalb eines Arrays nicht völlig gleichmäßig, sondern es bildete sich ein Gradient aus, der die Signalintensität gegen Ende des Kanals abnehmen ließ. Deshalb wurde auf das Signal der jeweils nächstgelegenen Replikate der Konsensus-Sequenz normalisiert. Der Gradient konnte auf diese Art umgangen werden. Werte größer als 1,0 stehen für eine stärkere Bindung als an die Konsensus-Sequenz, Werte kleiner als 1,0 hingegen für eine Schwächere. Abbildung 4.21 zeigt den Mittelwert der Signalintensität der fünf Replikate eines Arrays. Jeder Punkt steht für eine bestimmte Sequenz. Es wird deutlich, dass es bei den Permutationen der Basen drastische Unterschiede in der Signalintensität gibt. Manche Permutationen erlauben eine stärkere Bindung als die Konsensus-Sequenz, ein Großteil der modifizierten Sequenzen jedoch keine oder nur eine deutlich verminderte. Ein Vergleich der Signale in Abbildung 4.21 zeigt die Ähnlichkeit der beiden Tags im Bindeverhalten mit der Konsensus-Sequenz und den Permutationen. Beim Vergleich der Intensitäten zeigt sich, dass eine Basenaustausch in der Mitte der Konsensus-Sequenz kritischer ist, als zum Rand hin. Sequenzen mit hohem Cytosin- und Adenin-Anteil in der Sequenz binden deutlich besser.

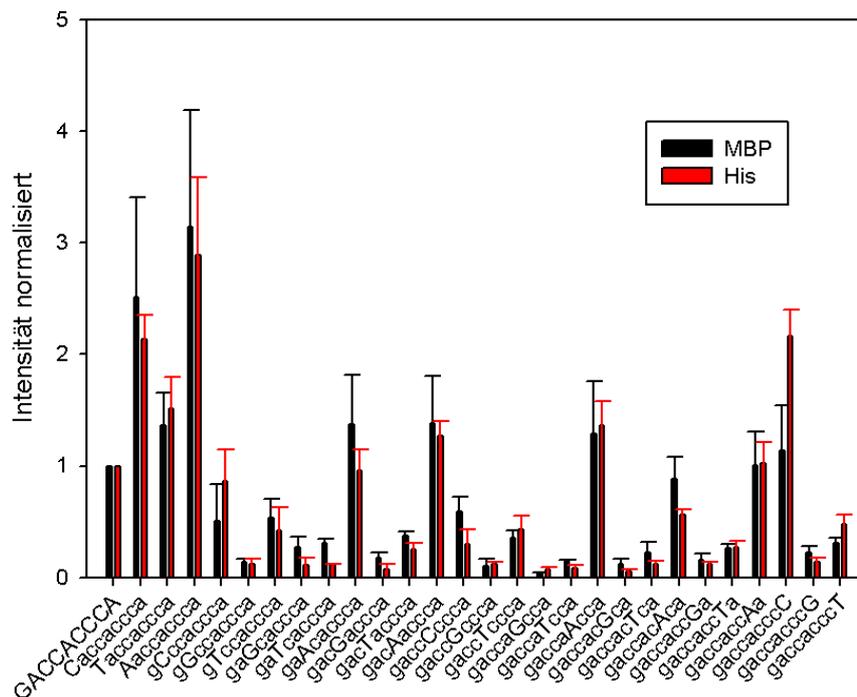


**Abbildung 4.21:** Dargestellt sind die gegen die Konsensus-Sequenz normalisierten Signalintensitäten der Bindung von Gli2 mit His-Tag und MBP-Tag an die permutierten Sequenzen.

### Einflüsse verschiedener DNA-Konzentrationen

Wie im Kapitel “Optimierung der DNA-Konzentration” gezeigt, hat die DNA Konzentration auf dem Microarray einen Einfluss auf die Reproduzierbarkeit der Experimenten. Den Einfluss auf die DNA-Protein-Wechselwirkung zeigt Abbildung 4.22.

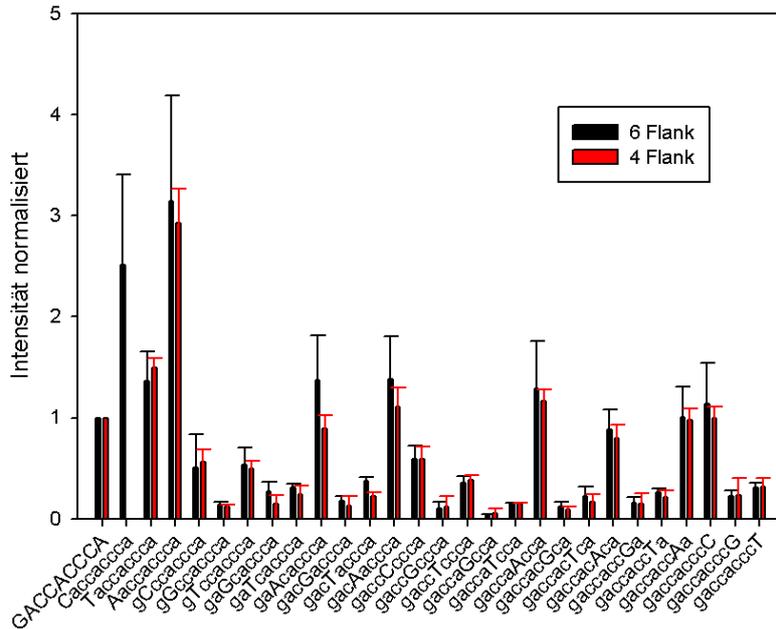




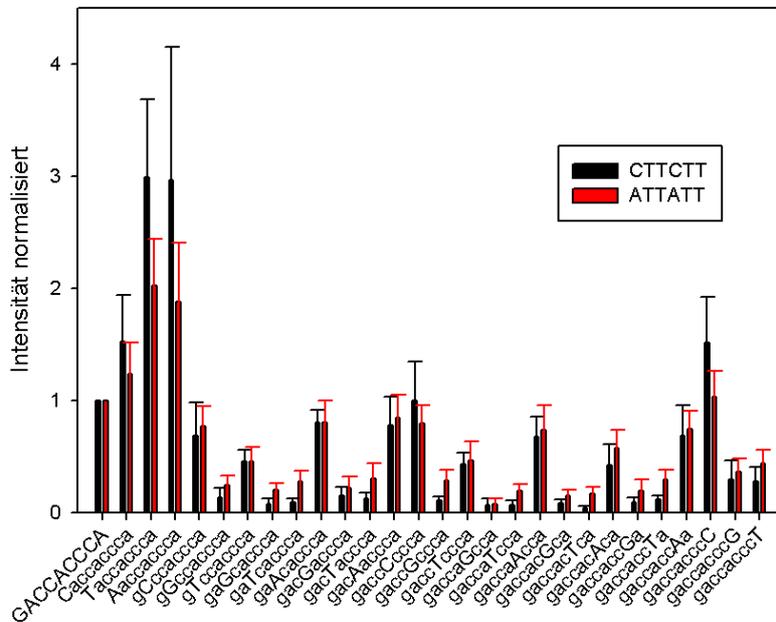
**Abbildung 4.23:** Vergleich der Bindungsstärke von Gli2 His- und MBP-Tag an alle möglichen 1-Basenpaar-Permutationen der Konsensus-Sequenz. Aus den Replikaten des Arrays wurde die Standardabweichung bestimmt. In Großbuchstaben ist die permutierte Base angegeben.

### Einflüsse verschiedener Randnukleotide sowie deren Länge

Es ist aus der Literatur bekannt, dass sowohl die Anzahl als auch die Art der an die Konsensus-Sequenz angrenzenden Nukleotide das Binden des Transkriptionsfaktor an die DNA beeinflussen kann (Sinha et al. (2007)). Es wurde deshalb untersucht, ob die Anzahl und die Variation der an der Bindesequenz angrenzenden Nukleotide einen Einfluss auf das Bindeverhalten hat. Abbildung 4.25 zeigt die Substitution des ATTATT Motivs durch ein CTTCTT Motiv. Desweiteren wurde das Bindeverhalten eines 4 bp gegenüber eines 6 bp Randnukleotids untersucht (Abbildung 4.24).



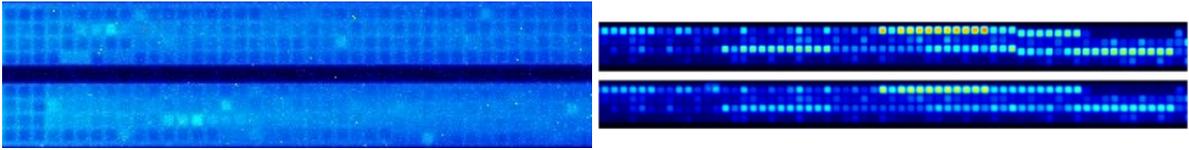
**Abbildung 4.24:** Vergleich der Bindungsstärke von Gli2 an alle möglichen 1-Basenpaar-Permutationen der Konsensus-Sequenz mit verschiedenen langen endständigen Nukleotiden. In Großbuchstaben ist die permutierte Base angegeben.



**Abbildung 4.25:** Vergleich der Bindungsstärke von Gli2 an alle möglichen 1-Basenpaar-Permutationen der Konsensus-Sequenz mit verschiedenen endständigen Nukleotiden. In Großbuchstaben ist die permutierte Base angegeben.





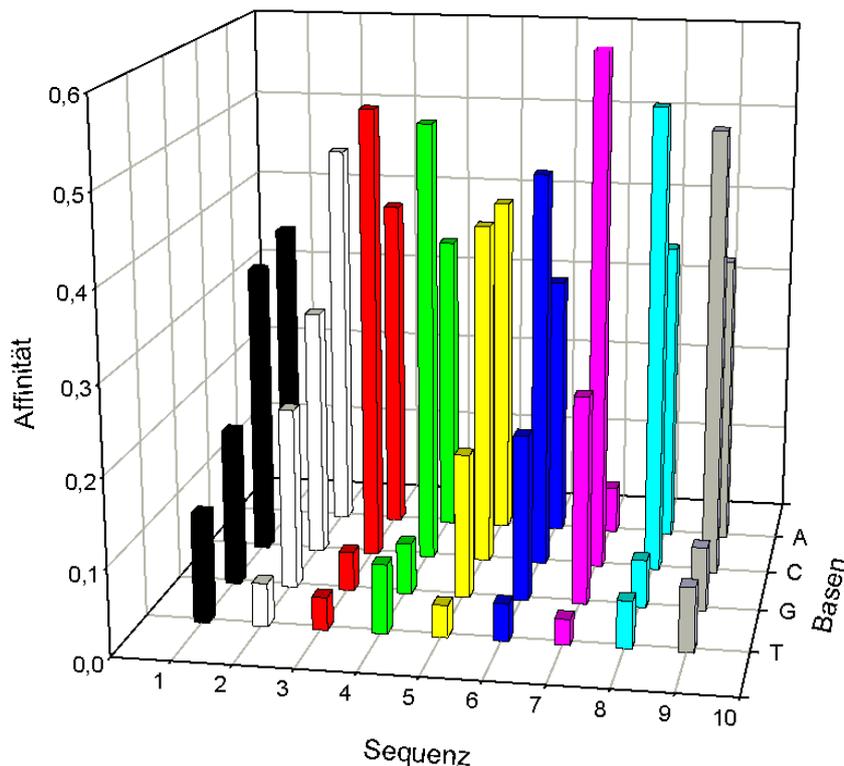


**Abbildung 4.28:** Links: Gli2 markiert mit Alexa555. Rechts: Gli2 markiert mit Streptavidin Phycoerythrin.

Dieser Vergleich zeigt, dass SAPE bei ähnlicher Belichtungszeit (3-4 Sekunden) bessere Signale und einen niedrigeren Hintergrund liefert als der Alexa555 Farbstoff.

#### 4.2.5.9 Gewichtete Positionsmatrix

Eine andere Art der Darstellung der Abhängigkeit der Bindungsaffinität ist die gewichtete Positionsmatrix. Sie ist in Abbildung 4.29 für das Gli2-Fusionsprotein gezeigt. Es wird deutlich, dass die Bindung durch Cytosin- und Adenosin- Nukleotide begünstigt und durch Guanosin- und Thymidin-Nukleotide, nahezu vollständig unterbunden wird.



**Abbildung 4.29:** Gewichtete Positionsmatrix von MBP-Gli2

Eine graphisch etwas bessere Darstellung der gewichteten Positionsmatrix ist das sogenannte Sequenzlogo. Hierbei werden die Verhältnisse der jeweiligen Basen an ihrer jeweiligen Position

ins Verhältnis gesetzt und in Buchstaben ausgegeben. Die Höhe dieser Buchstaben ist somit direkt proportional zu der Bindungsaffinität des Transkriptionsfaktors relativ zu der Konsensus Sequenz.

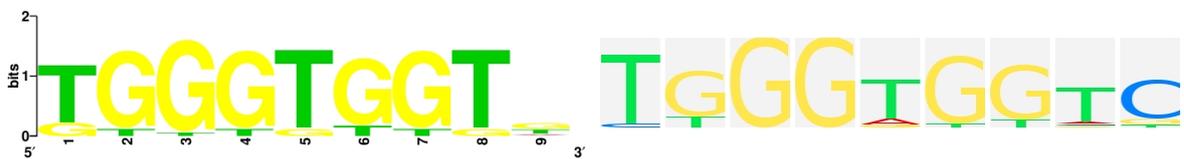


**Abbildung 4.30:** Gewichtete Positionsmatrix aller einfachen Permutationen von Gli2 ausgehend von der Konsensus-Sequenz GACCACCCA (links). Sequenzlogo aus der Literatur von Gli2 (rechts).

Abbildung 4.30 zeigt eine Vergleich der Microarray-Ergebnisse mit den aus der Literatur bekannten Daten.

#### 4.2.5.10 Sequenzmuster-Analyse

Um wie bei der gewichteten Positionsmatrix nicht nur die einfachen permutierten Sequenzen bei der Auswertung in Betracht zu ziehen, werden bei der Sequenzmusteranalyse 5 % der stärksten Signale analysiert, in diesem Fall alle einfach und zweifach mutierten Sequenzen. Abbildung 4.31 zeigt das Ergebnis solch einer Analyse. Es werden die prozentualen Anteile der 4 verschiedenen Basen an den bestimmten Stellen der Bindesequenz berechnet. Als Vergleich ist die aus der Literatur bekannte Konsensus-Sequenz gezeigt.



**Abbildung 4.31:** Sequenzmusteranalyse aller einfach permutierten Sequenzen von Gli2 ausgehend von der Konsensus-Sequenz GACCACCCA (links). Aus der Literatur bekanntes Sequenzlogo von Gli2 (rechts).

Eine Zusammenfassung der oben gezeigten Experimente ist noch einmal in Abbildung 4.32 als *color coded list* zu sehen. Diese Tabelle fasst noch einmal die wesentlichen Aspekte zusammen und gibt durch ihre Farbgebung einen Überblick über die Einflüsse der verschiedenen Konditionen der Experimente.

Folgende Abkürzungen wurden für die Tabelle verwendet:

- 9xT ATT: 9xT Spacer mit einer ATTATT Seitengruppe
- 9xT CTT: 9xT Spacer mit einer CTTCTT Seitengruppe
- 2xT ATT: 2xT Spacer mit einer ATTATT Seitengruppe
- 2xT ATT: 2xT Spacer mit einer CTTCTT Seitengruppe

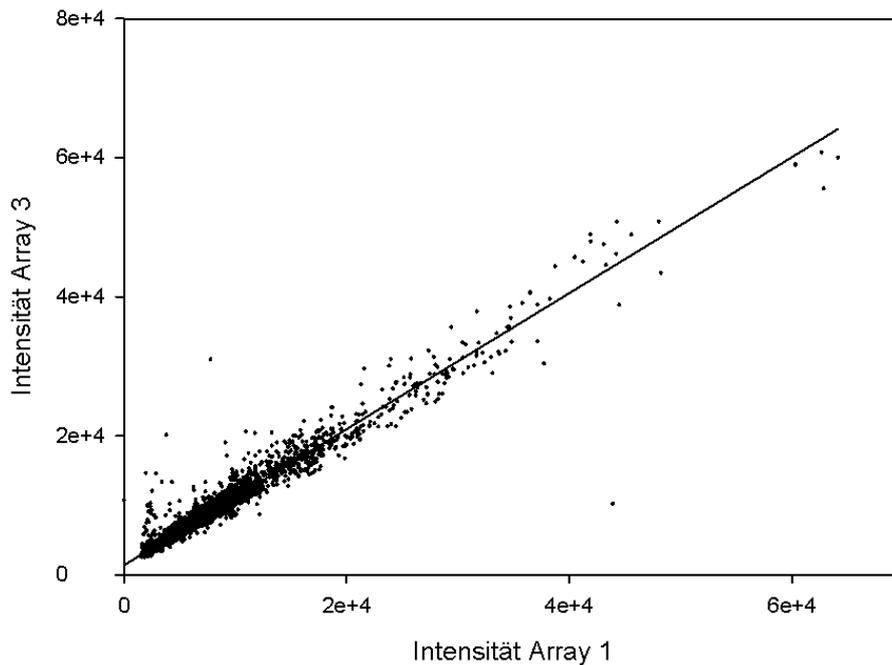
- 4xT ATT: 4xT Spacer mit einer ATTATT Seitengruppe

GLI2	Low DNA Density Chip (MBP)				High DNA Density Chip (HIS)			High DNA Density Chip (MBP)		
	9xT ATT	9xT CTT	2xT ATT	2xT CTT	9xT ATT	2xT ATT	4xT ATT	9xT ATT	2xT ATT	4xT ATT
GACCACCCA	1.07	0.87	0.93	0.83	1.00	1.00	1.00	1.00	1.00	1.00
Caccaccca	2.87	1.84	2.47	1.70	2.38	2.08		2.67	3.07	
Taccaccca	1.54	1.14	0.82	1.00	1.72	1.34	1.21	1.42	1.46	1.54
Aaccaccca	3.03	1.93	2.47	2.05	3.69	2.44	2.40	3.07	3.11	3.08
gCccaccca	0.68	0.73	0.60	0.54	1.01	0.91	0.85	0.47	0.35	0.55
gGccaccca	0.12	0.24	0.16	0.20	0.11	0.13	0.13	0.19	0.08	0.11
gTccaccca	0.49	0.43	0.21	0.30	0.60	0.57	0.39	0.45	0.41	0.33
gaGcaccca	0.08	0.19	0.12	0.15	0.13	0.21	0.09	0.20	0.11	0.12
gaTcaccca	0.09	0.28	0.09	0.18	0.12	0.20	0.10	0.31	0.23	0.24
gaAcaccca	0.79	0.79	0.63	0.68	1.09	0.87	0.52	1.31	1.03	0.82
gacGaccca	0.16	0.22	0.22	0.15	0.09	0.12	0.09	0.21	0.12	0.11
gacTaccca	0.12	0.30	0.20	0.22	0.24	0.33	0.14	0.42	0.32	0.24
gacAaccca	0.72	0.86	0.84	0.61	1.35	1.36	0.77	1.47	1.11	1.02
gaccCccca	1.06	0.77	0.61	0.46	0.35	0.43	0.30	0.57	0.42	0.44
gaccGccca	0.10	0.29	0.07	0.03	0.11	0.14	0.06	0.13	0.03	0.13
gaccTccca	0.45	0.47	0.15	0.07	0.41	0.16	0.18	0.47	0.36	0.31
gaccaGcca	0.07	0.09	0.16	0.03	0.07	0.04	0.10	0.11	0.03	0.09
gaccaTcca	0.05	0.18	0.11	0.02	0.08	0.12	0.07	0.24	0.07	0.25
gaccaAcca	0.71	0.68	0.57	0.39	1.23	0.85	0.83	1.11	1.03	0.96
gaccacGca	0.09	0.14	0.11	0.02	0.06	0.11	0.08	0.13	0.09	0.12
gaccacTca	0.05	0.17	0.12	0.05	0.13	0.12	0.07	0.22	0.06	0.24
gaccacAca	0.38	0.57	0.22	0.26	0.57	0.22	0.30	0.84	0.78	0.53
gaccaccGa	0.11	0.19	0.09	0.03	0.11	0.04	0.10	0.24	0.05	0.22
gaccaccTa	0.11	0.29	0.11	0.02	0.23	0.12	0.17	0.28	0.19	0.31
gaccaccAa	0.69	0.76	0.51	0.44	0.94	0.71	0.80	0.98	0.90	1.10
gaccaccC	1.62	1.02	1.46	0.84	2.28	1.03	1.10	1.04	0.77	0.91
gaccaccG	0.23	0.37	0.24	0.03	0.13	0.09	0.11	0.27	0.04	0.35
gaccaccT	0.23	0.44	0.36	0.07	0.51	0.20	0.34	0.36	0.09	0.38

**Abbildung 4.32:** Vergleich der verschiedenen Experimente mit Gli2 als “color coded list”. Rot: Das Protein bindet gut an die vorgegebene DNA-Sequenz; blau: Protein bindet schlecht an die vorgegebene DNA-Sequenz. In Großbuchstaben ist die permutierte Base angegeben.

#### 4.2.5.11 Reproduzierbarkeit

Die Reproduzierbarkeit der Ergebnisse kann mit Hilfe eines *Scatter Plots* bestimmt werden. Hierbei werden die Rohdaten von zwei verschiedenen Experimenten gegeneinander aufgetragen und der Korrelationskoeffizient berechnet. Ist der Korrelationskoeffizient gleich eins, sind die Ergebnisse identisch. In Abbildung 4.33 ist der *Scatter Plot* für MBP-Gli2 gezeigt. Der berechnete Korrelationskoeffizient beträgt:  $R = 0.9865$ .

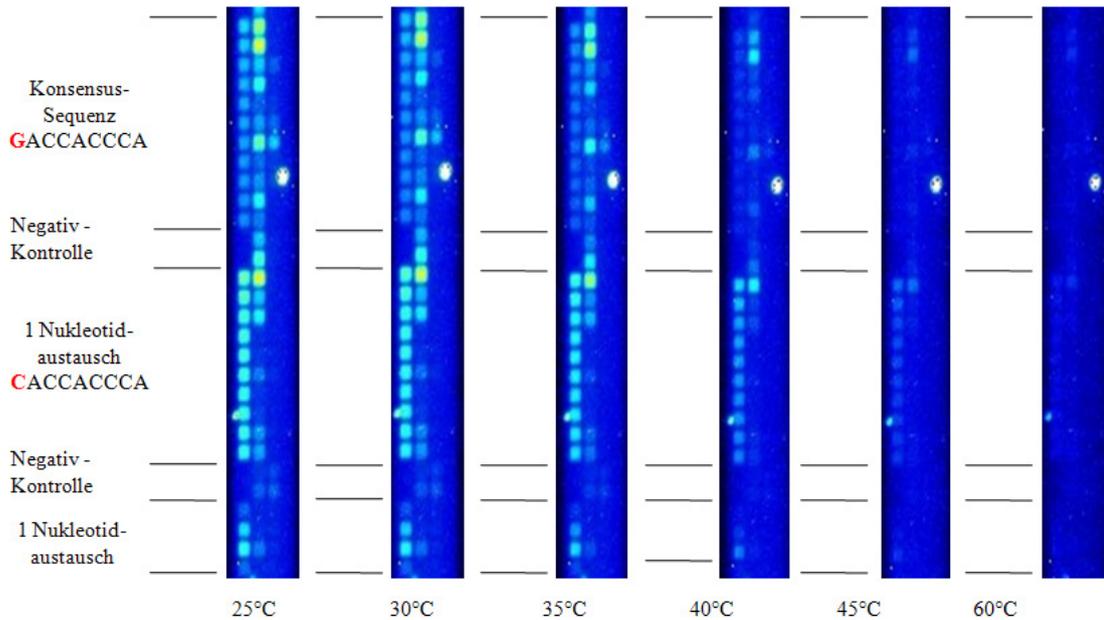


**Abbildung 4.33:** *Scatter Plot* der Intensitäten von Gli2 auf den Arrays 1 und 3. Der berechnete Korrelationskoeffizient beträgt:  $R = 0.9865$

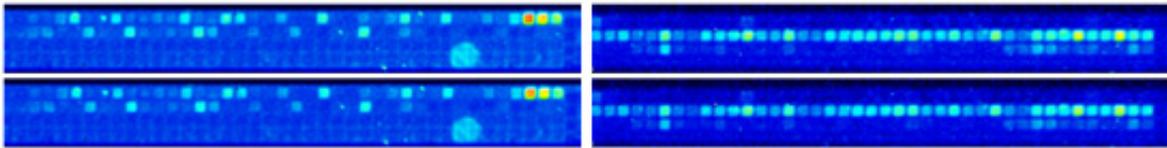
#### 4.2.6 Dynamische Detektion

Das Geniom bietet neben der Detektion nach Inkubation von DNA und Protein auch die Möglichkeit, bis zu zehn zusätzliche Temperatur- und Waschschrte zu integrieren und diese separat zu detektieren. Somit konnte die Stabilität der Bindung des Transkriptionsfaktors an verschiedene Nukleotidsequenzen in Abhängigkeit von der Temperatur untersucht werden.

Wie in Abbildung 4.34 gezeigt, nimmt die Signalintensität mit zunehmender Temperatur ab. Um ein Ausbleichen des Farbstoffes auszuschließen, wurde dreimal ein Foto nach jeweils 10-sekündiger Belichtungszeit aufgenommen. Abbildung 4.35 verdeutlicht, dass es keinen derartigen Effekt gab. Um den Einfluss der Waschschrte auf die Signalintensität zu untersuchen, wurden insgesamt 10 Bilder nach je 2 Sekunden Waschen bei gleichbleibender Temperatur aufgenommen. Auch hier ist keine Abnahme der Signalintensität zu erkennen. Auswertungen ergaben eine Reduktion des Signals durch Wasch- und Ausbleicheffekte von etwa 5%.

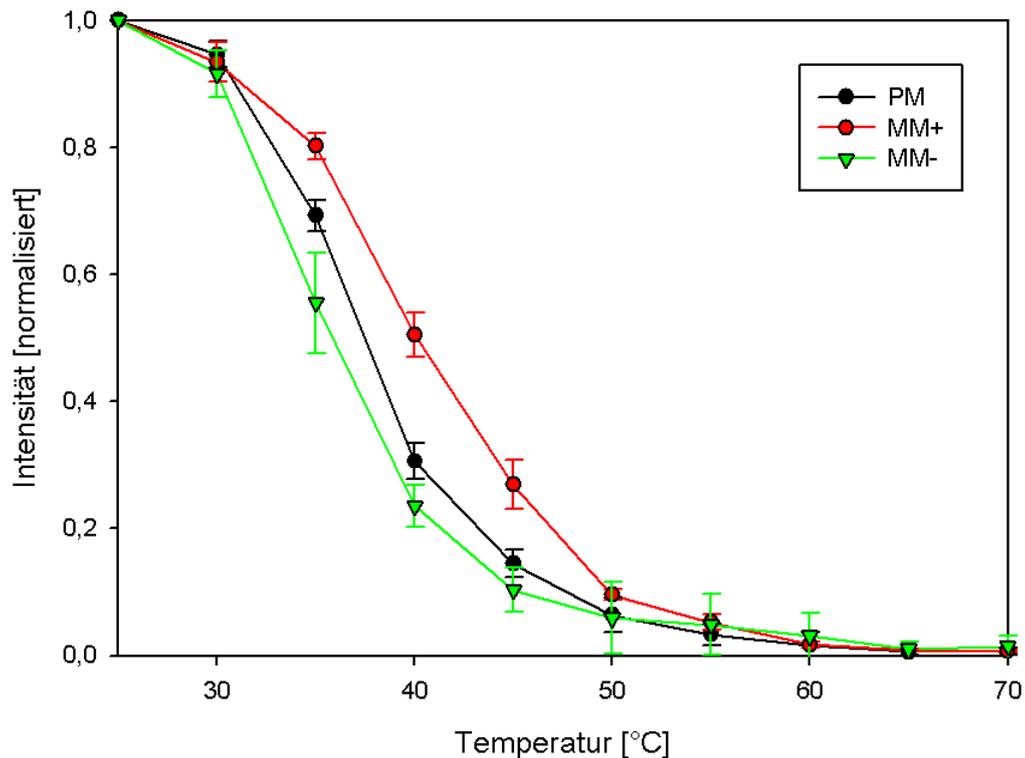


**Abbildung 4.34:** Ausschnitte von Bildern der dynamischen Detektion.



**Abbildung 4.35:** Ausschnitte von Bildern der bleaching Kontrolle (links) sowie Waschkontrolle (rechts).

Abbildung 4.36 stellt den Verlauf der dynamischen Detektion von Gli2 mit MBP-Tag an die Konsensus-Sequenz GACCACCCA, die stärker bindende Sequenz CACCACCCA und die schlechter bindende Sequenz GATCACCCA dar. Es wurden 5°C Schritte gewählt, um einen Überblick über das Dissoziationsverhalten des Proteins zu erhalten. Beginnend bei etwas unter 35°C fallen die Signalintensitäten mit zunehmender Temperatur ab. Bei zirka 45°C hat sich mehr als die Hälfte des Proteins bereits von der DNA gelöst. Normalisiert wurde gegen den Mittelwert der Signalintensität der betreffenden Sequenz bei 25°C. Die Fehlerbalken zeigen die Standardabweichung, der in 5 Blöcken angeordneten Replikate innerhalb eines Arrays.

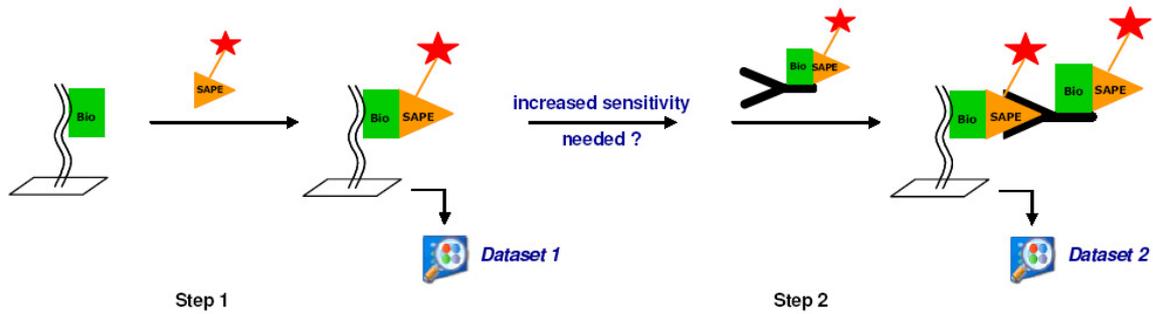


**Abbildung 4.36:** Ergebnis der dynamischen Detektion von Gli2. Vergleich der Signalintensität von Gli2 mit MBP-Tag an die Konsensus-Sequenz GACCACCCA (PM) sowie die stärker bindende Sequenz CACCACCCA (MM+) und die schlechter bindende Sequenz GATCACCCA (MM-) in Abhängigkeit von der Temperatur.

#### 4.2.7 Universeller 9mer Chip

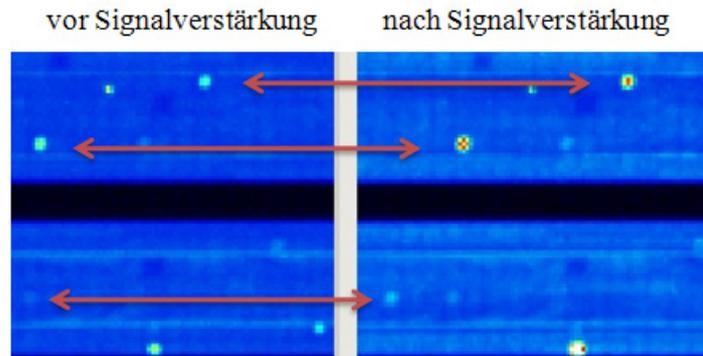
Durch das Ändern der Spiegelkonfiguration konnte gegen Ende der Doktorarbeit ein Chip mit 125 000 Spots synthetisiert werden. Mit diesem Chip war es möglich, alle möglichen mit der de Bruin Sequenz bestimmten 9 mer Sequenzen auf einem Chip als Replikate zu synthetisieren. Die Synthese sowie die Inkubationsbedingungen des Transkriptionsfaktors entsprachen den in den vorigen Kapitel erwähnten Standardbedingungen.

Die Erhöhung der Spotdichte wurde durch die Reduktion der Spotdurchmesser von 34  $\mu\text{m}$  auf 17  $\mu\text{m}$  erreicht. Die Verringerung des Spotdurchmessers erniedrigt gleichzeitig die Bindekapazität der einzelnen Spots. Dies hatte zur Folge, dass nur noch die stark bindenden Sequenzen detektiert werden konnten. Schwächer bindende Sequenzen fielen unter die Detektionsgrenze. Um diese Sequenzen trotzdem detektieren zu können, wurde eine Signalamplifikation durchgeführt. Durch die Verwendung eines biotinylierten Anti-Streptavidin-Antikörpers sowie einer nachfolgenden SAPE Inkubation konnte das Signal verstärkt werden. Ein Schema dieser Signalverstärkung zeigt Abbildung 4.37.



**Abbildung 4.37:** Schema der Signalverstärkung durch die Verwendung eines biotinylierten Anti-Streptavidin-Antikörpers.

Durch die Verwendung dieser Methode wird das Signal im Schnitt verfünffacht, während sich der Hintergrund verdoppelt. Bei einer Belichtungszeit von 2,7 Sekunden konnte diese auf 370 Millisekunden reduziert werden. Ein Beispiel hierfür zeigt Abbildung 4.38.



**Abbildung 4.38:** Signalverstärkung durch die Verwendung eines biotinylierten Anti-Streptavidin-Antikörpers. Oben: Array ohne Signalverstärkung (Belichtungszeit 2,7 Sekunden). Unten: Array mit Signalverstärkung (Belichtungszeit 370 Millisekunden).

#### Protokoll der Signalamplifikation:

- 1. Detektion
- Inkubation des Anti-Streptavidin-Antikörpers für eine Stunde (Konzentration: 0,03 mg/ml)
- Waschen 2x Arrayvolumen
- Inkubation SAPE für 15 Minuten (Konzentration: 80 nM)
- Waschen 2x Arrayvolumen
- 2. Detektion

#### Ergebnisse:

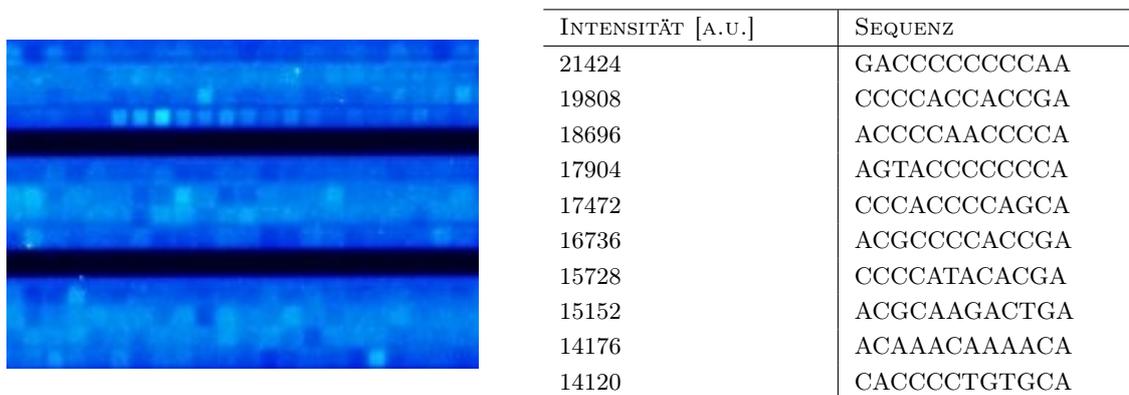
Folgende 10 Sequenzen konnten als die Stärksten identifiziert werden:

- CCCCCCCTCCG
- ACCCCGCTAAGA
- CCATGCCCCCG
- CCCCCGTGGCT
- AACCGCCCACCA
- ATTTGCCCCCCC
- CCCAACAGCGGA
- CCCCCCCAGAT
- CCCCCGTAACCT
- AAGTAGCCCCCC

Es ist deutlich zu erkennen, dass das Fusionsprotein an Teilsequenzen, die nur aus der Base Cytosin bestehen am besten bindet. Aufgrund des gewählten Designs und der Verwendung des “de Bruin Algorithmus” ist es jedoch nicht möglich, die genaue 9 bp lange Bindesequenz zu bestimmen. Sie ist vielmehr in der angegebenen 12 bp Sequenz eingebettet.

Als Vergleich wurde ein *low density chip*, der 54.000 Spots beinhaltet, mit allen 9 mer Sequenzen synthetisiert und mit MBP-Gli2 inkubiert. Da die durch den “de Bruin Algorithmus” berechneten 9 mer Sequenzen 48.000 Plätze auf dem Chip belegen, konnten keine Replikate synthetisiert werden. Dies ist für eine genaue Auswertung leider von Nachteil.

Einen kleinen Ausschnitt dieses Chips sowie die 10 stärksten Signale zeigt Abbildung 4.39.



**Abbildung 4.39:** Intensitäten aller untersuchten 9 mere (links) inklusive der 10 stärksten Signale (rechts).

Es sind deutliche Gemeinsamkeiten bei den 10 stärksten Signalen zu erkennen. Gli2 scheint Cytosin reiche Sequenzen für die Bindung zu bevorzugen.

### 4.3 Evaluierung der Ergebnisse mit anderen Methoden

Um die aus den Microarray-Experimenten gewonnenen Daten zu validieren, wurden zwei weitere Methoden verwendet. Mit Hilfe der Plasmonenresonanzspektroskopie wurde die Kinetik der Protein-DNA-Wechselwirkung bestimmt und mit Angaben aus der Literatur verglichen. Durch die Verwendung der Quarzkristallmikrowaage konnten die Stöchiometrie der Bindung sowie die Schichtdicken der DNA/Proteinbelegung bestimmt werden.

#### 4.3.1 Plasmonenresonanzspektroskopie (SPR)

Die Biacore T100 wurde zur Bestimmung der Bindungskinetiken der drei verschiedenen DNA-Sequenzen benutzt. Es wurde die Kinetik der Protein-DNA-Interaktion für verschiedene Proteinkonzentrationen bei konstanten DNA-Konzentrationen bestimmt. Untersucht wurden die aus der Literatur bekannte Konsensus-Sequenz, eine Permutation an die Gli2 besser bindet (1 MM+) sowie eine Permutation mit schlechteren Bindungseigenschaften (1 MM-). Die Auswahl der beiden Permutationen erfolgte auf Grund der Microarray-Ergebnisse.

##### 4.3.1.1 Immobilisierung der DNA

Als Ligand wurden selbst komplementäre und am 5'-Ende mit Biotin gekoppelte Oligonukleotide verwendet. Als Träger wurde ein mit Streptavidin beschichteter Biochip verwendet.

Die Immobilisierung des Liganden wurde bei einem konstanten Fluss von 30  $\mu\text{l}/\text{min}$  so lange durchgeführt, bis ein Resonanzsignal von 200 RU erreicht wurde. Die Menge der zu beladenden DNA wurde nach folgender Gleichung berechnet:

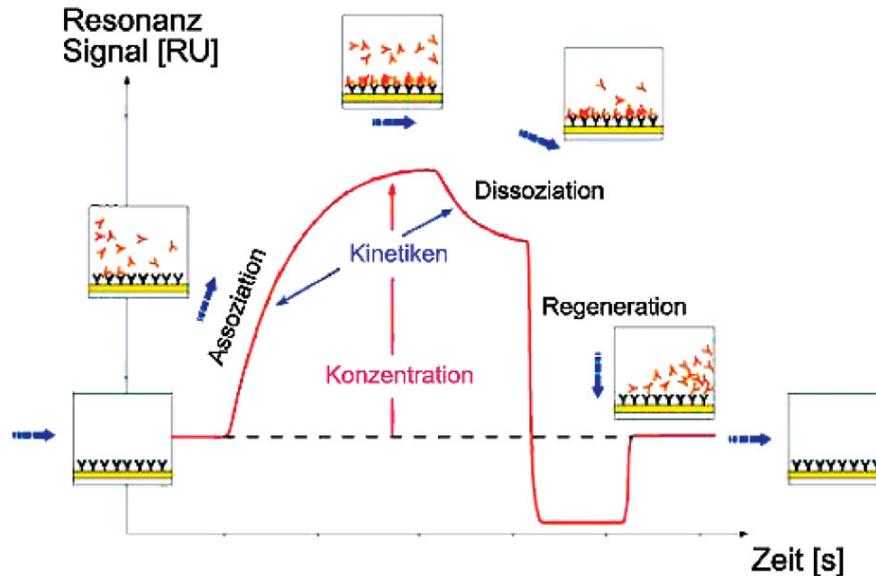
$$R_{max} = \frac{MW_{Analyt}}{MW_{Ligand}} * Resonanzsignal_{DNA} * Stoichiometrie$$

Als Laufpuffer wurde 1x Bindepuffer verwendet.

##### 4.3.1.2 Analyse

Für die SPR-Messung wurde wiederum ein Fluss von 30  $\mu\text{l}/\text{min}$  gewählt, das Probenvolumen für die Injektion betrug 250  $\mu\text{l}$ . Während der Injektion der Probe kann im Sensorgramm die Assoziationskinetik abgelesen werden. Nach dem Ende der Injektion wird der Chip mit Laufpuffer gespült. Die Dissoziation des Analyten vom Liganden kann dadurch bestimmt werden. Aus dieser Phase der Analyse ist es möglich, den  $K_D$ -Wert einer Interaktion zwischen Proteinen und DNA zu bestimmen (siehe Abbildung 4.40). Die Konzentration der Protein-Proben lag im Bereich von 25-6000 nM. Jede Analyse wurde mit 6 verschiedenen Konzentrationen durchgeführt. Für die Bindungskinetik des Analyten mit einem immobilisierten Liganden sind für den Signal-Aufbau im BIAcore-System folgende Parameter von Bedeutung:

- die Konzentration des Analyten
- das Molekulargewicht des Analyten
- die Dauer der Analyt-Kontakt-Zeit



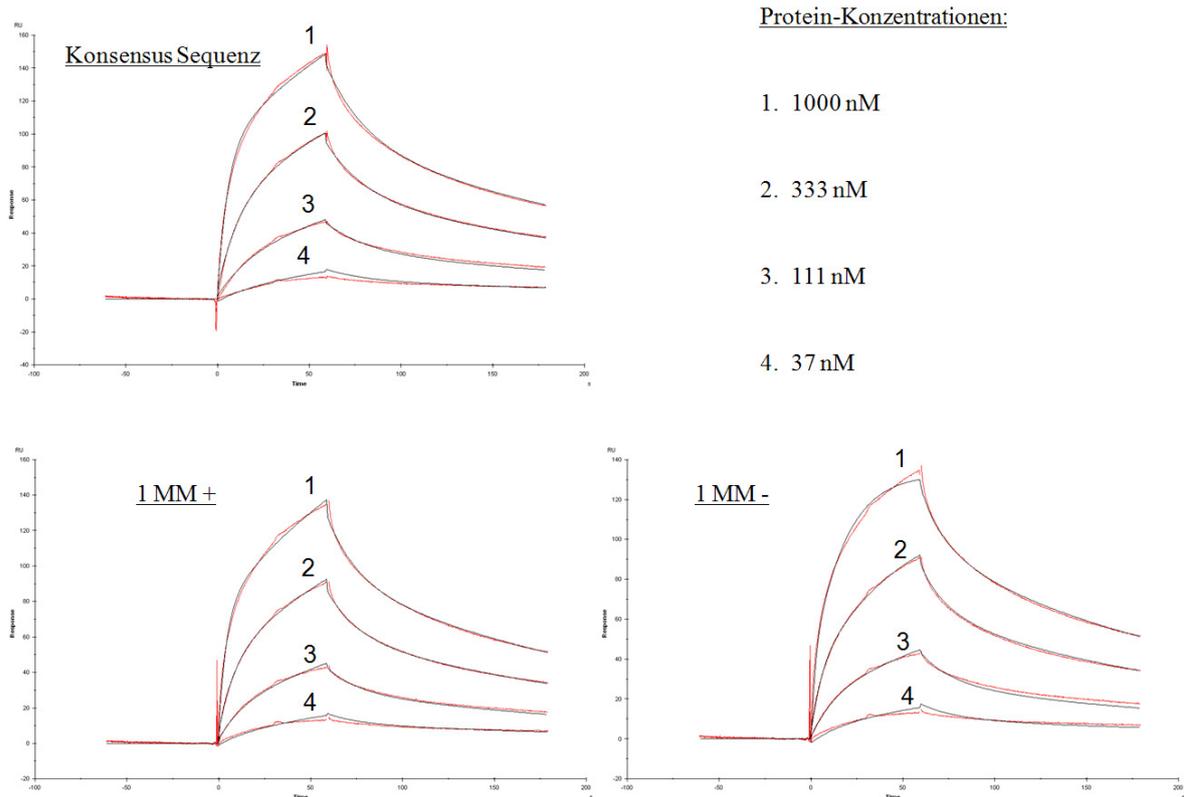
**Abbildung 4.40:** Idealisierte Darstellung der Bindungskinetik einer Protein-DNA -Wechselwirkung in einem BIAcore-T100-System

#### 4.3.1.3 Auswertung

Alle Berechnungen wurden unter Benutzung der BIAevaluation 3.0 Software (BIAcore) durchgeführt.

Die Sensorgramme der drei verschiedenen DNA-Sequenzen mit den jeweiligen Konzentrationen sind in Abbildung 4.41 gezeigt.

Da die Protein-DNA-Wechselwirkung nicht einem 1:1- Bindungsmodell entspricht, konnten die Bindungskinetiken nicht nach einer Reaktion pseudo erster Ordnung bestimmt werden. Die Ausbildung eines DNA Doppelstranges verläuft bei der Hairpin-Bildung nie zu 100%. Somit stehen nicht alle Oligonukleotide als Bindungspartner zur Verfügung. Deshalb wurde für die Auswertung das Modell eines heterogenen Liganden verwendet. Die Assoziations- bzw. Dissoziationskonstanten wurden mit der Option „*lokal fit*“ berechnet, wodurch eine Korrelation der theoretischen - aus den Geschwindigkeits-Konstanten errechneten - Bindungskurven auch im Grenzbereich erreicht werden konnte.



**Abbildung 4.41:** Sensorgrammen von Gli2 aufgenommen mit dem Biacore T100. Es sind die drei DNA Sequenzen mit den verschiedenen Proteinkonzentrationen gezeigt (Konsensus-Sequenz, eine Permutation die besser bindet (MM+) sowie eine mit schwächerer Bindung(MM-)).

Tabelle 4.4 zeigt die aus den verschiedenen Sensorgrammen berechneten Daten:

SEQUENZ	$K_a$	$K_d$	$K_D$
Konsensus	$6,12 * 10^4$ 1/Ms	0,00566 1/s	$9,25 * 10^{-8}$ M
1 MM +	$5,96 * 10^4$ 1/Ms	0,00526 1/s	$9,24 * 10^{-8}$ M
1 MM -	$1,16 * 10^4$ 1/Ms	0,00658 1/s	$5,77 * 10^{-7}$ M

**Tabelle 4.4:** Ergebnisse der Biacore-Messung berechnet aus den verschiedenen Sensorgrammen mit Hilfe der Biacore Auswerte-Software

Vergleicht man die aus der Biacore Messung erhalten Daten mit den Werten aus der Literatur (Yang et al. (1995), Sakai-Kato et al. (2008)), kann man Übereinstimmungen erkennen (siehe Tabelle 4.5).

SEQUENZ	$K_a$	$K_d$	$K_D$
Gli2	-	-	$8,3 * 10^{-9}$ M
Zif268 WT +	$3 * 10^4$ 1/Ms	0,0002 1/s	$6,5 * 10^{-8}$ M

**Tabelle 4.5:** Literaturwerte der Kinetik-Messung von Zink-Finger-Proteinen.

Die Gleichgewichtskonstante von MBP-Gli2 liegt mit  $10^{-8}$  M im Bereich der Literatur  $10^{-9}$  M. Der um Faktor 10 verschiedene Wert kann über den kooperativen Effekt des Gesamtproteins erklärt werden. Da bei den in dieser Arbeit verwendeten Proteinen nur die Bindeseite des Proteins und nicht das Gesamtprotein verwendet wurde, können diese Werte leicht variieren.

Somit lagen die gemessenen Gleichgewichtskonstanten im nanomolaren Bereich, was einer mittelstarken Bindung entspricht.  $K_D$ -Werte im nanomolaren bis subnanomolaren Bereich stehen für starke Wechselwirkungen, während Wechselwirkungen im mikromolaren Bereich eher schwach sind.

Verglichen mit dem C2H2-Zinkfinger Proteinen Zif268 liegen die hier bestimmten Werte in den gleichen Bereichen.

### 4.3.2 Quartz-Kristall-Mikrowaage

Mit Hilfe der Quartz-Kristall-Mikrowaage wurden die Stöchiometrie der Protein-DNA-Bindung sowie die Dicken der DNA- und Proteinschichten bestimmt. Auch die Viskosität der verschiedenen Oberflächen wurde bestimmt.

Sämtliche Messungen wurden mit einem Q-Sense E4 Messgerät durchgeführt. Für die Messung wurde das Programm Qsoft401 und für die Auswertung die Software QTools verwendet.

#### Material:

- E4 Schwingquarzmikrowaage von Q-Sense
- QSX303 SiO<sub>2</sub>-Kristalle von Q-Sense
- DOPC/DOPE-Cap-Biotin SLB mit SA monolayer auf SiO<sub>2</sub>
- Bio-DNAs (1  $\mu$ M (1:100) in BB)

Tabelle 4.6 zeigt den Ablauf des QCMD-Experimentes mit den zugehörigen Injektionszeiten.

INJEKTIONSZEITEN [MIN]	PROBEN
00:00:00	50 $\mu$ l/min Binding Buffer
00:02:00	50 $\mu$ l/min DNA
00:12:00	50 $\mu$ l/min Binding Buffer
00:17:00	50 $\mu$ l/min Protein (Gli2, 10 $\mu$ g/ml in BB)
00:25:00	50 $\mu$ l/min Protein (Gli2, 100 $\mu$ g/ml in BB)
00:37:00	50 $\mu$ l/min Binding Buffer
01:02:00	150 $\mu$ l/min Binding Buffer

**Tabelle 4.6:** Ablauf des QCMD-Experiments

#### 4.3.2.1 Herstellung der Streptavidinoberfläche

Die Reinigung der Quarzkristalle erfolgte für 30 min in Sauerstoffplasma. Sie wurden somit von allen organischen Lösungsmitteln befreit. Danach wurden die Kristalle in den Durchflusskammern fixiert und in das Gerät eingesetzt. Nach dem Anschließen der Zulaufschläuche wurden diese in die Pumpe eingesetzt. Die Flussrate wurde bei allen Versuchen konstant bei 50  $\mu\text{l}/\text{min}$  gehalten. Zuerst wurde der Streptavidin-Monolayer erzeugt. Dies erfolgte durch die Verwendung von negativ geladenen Vesikeln bestehend aus 80% DOPC, 10% DOPS and 10% DOPE-Biotin.

#### 4.3.2.2 Immobilisierung der DNA

Die Immobilisierung der biotinylierten Hairpin-DNA erfolgte bei konstantem Fluss bei Konzentration von 1  $\mu\text{M}$  in 1x Bindepuffer. Es wurden 3 verschiedene Oligonukleotide in 3 verschiedenen Kammer verwendet:

- Konsensus-Sequenz
- Permutation die besser an die DNA bindet als die Konsensus-Sequenz (1MM+)
- Permutation die schlechter an die DNA bindet als die Konsensus-Sequenz (1MM-)

Danach wurde so lange mit Puffer gespült, bis ein konstantes Signal erreicht wurde.

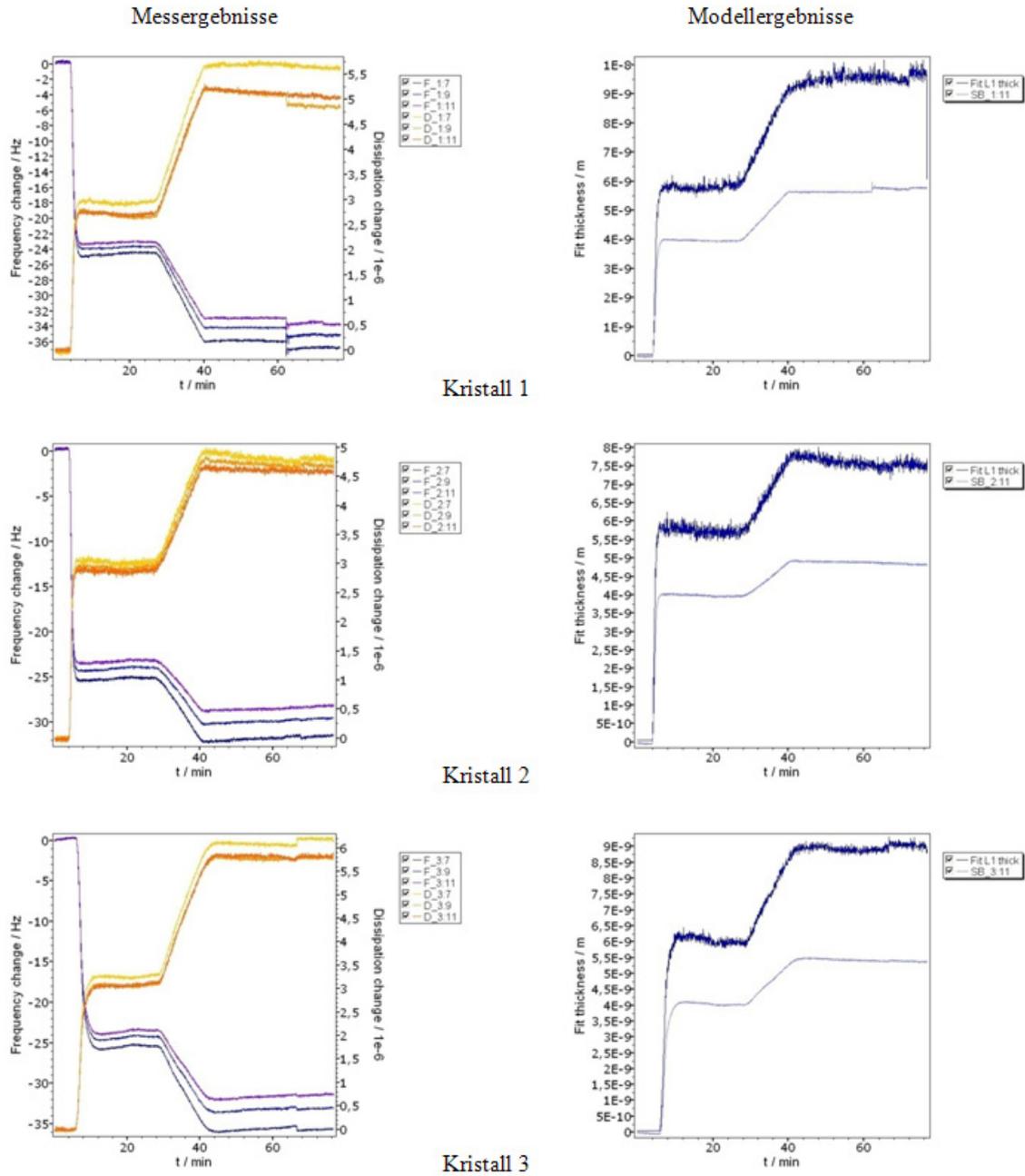
#### 4.3.2.3 Analyse

Für die QCMD-Messung wurde wiederum ein Fluss von 50  $\mu\text{l}/\text{min}$  gewählt, das Probenvolumen für die Injektion betrug 600  $\mu\text{l}$ . Während der 12 minütigen Injektion der Probe kann im Sensorgramm die Frequenzänderung und die Dissipationsänderung direkt ablesen werden. Nach dem Ende der Injektion wurde der Chip mit Puffer gespült.

#### 4.3.2.4 Auswertung

Alle Berechnungen wurden unter Benutzung der QTools Software (Q-Sense) durchgeführt. Folgende Einstellungen wurden hierfür gewählt:

- 1 Layer
- Lösungsmitteldichte [ $\text{kg}/\text{m}^3$ ] : 1020
- Lösungsmittelviskosität [ $\text{kg}/\text{ms}$ ] : 0,001
- Layerdichte [ $\text{kg}/\text{m}^3$ ] : 1050



**Abbildung 4.42:** Links: Messergebnisse für die einzelnen Kristalle: Frequenz- und Dissipationswerte für die Obertöne 7, 9 und 11. Rechts: Modellierte Dicke nach dem Voigt-Modell für visko-elastische Systeme und nach der Sauerbrey-Gleichung für starre Schichten.

Nach der Wahl der Obertöne konnten nun die Kurven gefittet werden. Die Auswertung der Ergebnisse erfolgte nach der Methode von Sauerbrey (Sauerbrey (1959)). Die Modell und Messergebnisse für die einzelnen Kristalle sind in Abbildung 4.42 dargestellt.

### Schichtdickenbestimmung der DNA

Die Schichtdickenbestimmung wurde für die drei auf der Oberflächen gebunden Oligonukleotide nach der Methode von Sauerbrey und Voigt bestimmt. Die bestimmten Dicken sind in Tabelle 4.7 gezeigt.

SEQUENZ	VOIGT; T=17 MIN	SAUERBREY; T=17 MIN
Konsensus	4,0 nm	4,2 nm
1 MM +	4,3 nm	4,2 nm
1 MM -	4,0 nm	4,3 nm

**Tabelle 4.7:** Ergebnisse der QCMD-Messung. Schichtdickenbestimmung der DNA

Da die Oligonukleotide in ihre Länge nicht variieren, unterscheiden sie sich auch in ihrer Schichtdicke nicht wesentlich. Eine kleine Differenz ist im Vergleich der zwei verwendeten Methoden zu erkennen.

### Schichtdickenbestimmung des Proteins

Die Bestimmung der Schichtdicke der Proteinbelegung wurde ebenfalls nach der Methode von Voigt und Sauerbrey bestimmt. Die Ergebnisse sind in Tabelle 4.8 zusammengefasst.

SEQUENZ	VOIGT; T=60 MIN	SAUERBREY; T=60 MIN
Konsensus	1,7 nm	1,9 nm
1 MM +	1,6 nm	1,6 nm
1 MM -	1,0 nm	1,1 nm

**Tabelle 4.8:** Ergebnisse der QCMD-Messung. Schichtdickenbestimmung des Proteins nach Voigt und Sauerbrey.

Die Ergebnisse der Schichtdickenbestimmung des Proteins zeigt deutliche Unterschiede zwischen den immobilisierten Oligonukleotiden. Die Konsensus-Sequenz sowie die permutierte besser bindende Sequenz zeigen eine größere Schichtdicke als die permutierte schlechter bindende Sequenz.

### Massenbestimmung

Die Massenbestimmung der immobilisierten DNA sowie des Proteins wurde nach der Methode von Sauerbrey berechnet. Die Ergebnisse sind in Tabelle 4.9 zusammengefasst.

SEQUENZ	DNA; T=17 MIN	PROTEIN; T=17 MIN
Konsensus	420 ng/cm <sup>2</sup>	183 ng/cm <sup>2</sup>
1 MM +	426 ng/cm <sup>2</sup>	164 ng/cm <sup>2</sup>
1 MM -	423 ng/cm <sup>2</sup>	104 ng/cm <sup>2</sup>

**Tabelle 4.9:** Ergebnisse der QCMD-Messung. Massenbestimmung der DNA und des Proteins nach der Methode von Sauerbrey.

Wie bei der Schichtdickenbestimmung unterscheidet sich auch die Masse der gebundenen DNA auf der Oberfläche nicht gravierend. Die Massenbestimmung der Proteinbelegung zeigt die Unterschiede der Konsensus und der permutierten Sequenzen auf.

### Molaritätsbestimmung

Die aus den oben bestimmten Massen berechneten Molaritäten sind in Tabelle 4.10 zusammengefasst.

SEQUENZ	DNA; T=60 MIN	PROTEIN; T=60 MIN
Konsensus	24 pM	2,4 pM
1 MM +	24 pM	2,1 pM
1 MM -	24 pM	1,3 pM

**Tabelle 4.10:** Ergebnisse der QCMD-Messung. Molaritätsbestimmung nach dem Modell Voigt und Sauerbrey.

### Bestimmung der Viskosität der Schichten

Die Ergebnisse der Viskositätsbestimmung sind in Tabelle 4.11 gezeigt. Es sind keine Unterschiede bei der DNA/Protein-Belegung zu erkennen.

SEQUENZ	VISKOSITÄT; T=17 MIN	VISKOSITÄT; T=60 MIN
Konsensus	0,0024 Pa * sek	0,0020 Pa * sek
1 MM +	0,0023 Pa * sek	0,0020 Pa * sek
1 MM -	0,0023 Pa * sek	0,0019 Pa * sek

**Tabelle 4.11:** Ergebnisse der QCMD-Messung. Bestimmung der Viskosität der verschiedenen Schichten.

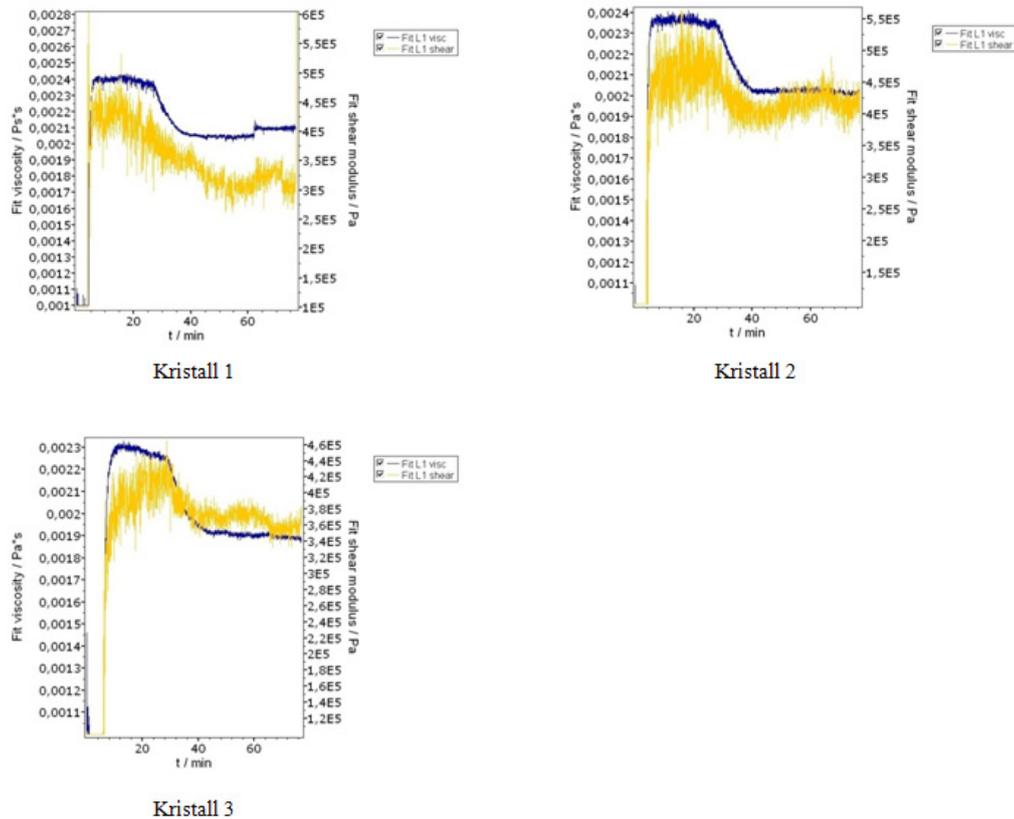


Abbildung 4.43: Werte für die Viskosität und das Schermodul

#### 4.4 Expression der Fusionsproteine in unterschiedlichen *E. coli* - Stämmen.

Für die Proteinexpression in *E. coli* - Bakterien gibt es verschiedene "klassische" Bakterienstämme. Diese sind in der Regel nicht identisch mit Bakterien, die man für zum Klonieren verwendet, da sie speziell für die Proteinexpression modifiziert sind. Für einfache, gut zu exprimierende Proteine allerdings kann man den Stamm XL1-Blue verwenden. Oft empfiehlt sich jedoch, bei der Proteinherstellung Bakterien zu verwenden, aus deren Genom verschiedene Proteasen entfernt wurden, die das hergestellte Protein gefährden würden. Am bekanntesten ist hier der Stamm BL21 (Bsp. Rosetta). Einige Expressionssysteme verwenden T7 Promotoren. Um diese verwenden zu können, sollte man BL21 Bakterien mit DE3 verwenden. Möchte man mit einem T7 Promotor Proteine exprimieren, von denen man annimmt, sie seien toxisch für die Bakterien, kann man durch Verwendung von BL21 DE3 pLYSs eine noch strikere Induzierbarkeit erreichen (Expression eines Inhibitors der T7 Polymerase).

Will man humane Proteine in Bakterien exprimieren, muss man darauf achten, dass in der humanen Sequenz Codons für bestimmte Aminosäuren häufig vorkommen, die in Bakterien extrem selten sind. Daher haben die Bakterien auch nur eine eingeschränkte Menge an den entsprechenden tRNAs. Dies kann zur unvollständigen Synthese und zur schlechten Expression des Proteins führen. Häufig sieht man dann in Kontrollgelen eine starke Bande auf Höhe des

Fusionstags (z.B. GST bei 25 KDa, MBP bei 52 KDa), da dieser Proteinbestandteil frei von *rare codons* und relativ stabil als Protein ist.

Weiterhin gibt es verschiedene Stämme, bei denen sehr selten vorkommende tRNAs supplementiert sind. Ein solcher ist z.B. die in dieser Arbeit verwendeten Stämme Rosetta2 und Rosetta gama (Novagen). Bakterienstämme mit ergänzten *rare codons* findet man bei Stratagene oder Novagen ( z.B Rosetta, Origami etc.). Die tRNAs sind in Plasmide codiert, die gleichzeitig zur Selektion eine Chloramphenicolresistenz tragen. Nachfolgend sind die für *E. coli* - Bakterien selten vorkommenden *rare-codons* aufgezeigt.

- AGG: Arginin. Besser: CGC, CGT
- AGA Arginin.
- CGG Arginin.
- CGA Arginin.
- GGA Glyzin. Besser: GGC, GGT
- AUA Isoleuzin. Besser: ATT, ATC
- CUA Leuzin. Besser: CTG
- CCC Prolin. Besser: CCG

Es sind jedoch noch weitere Nachteile der Proteinexpression in *E. coli* - Bakterien bekannt:

- inkorrekte Faltung
- Instabilität
- Präzipitation (inclusion bodies)
- posttranslationale Modifikationen wie bei Eukaryonten werden nicht gemacht z.B. Glycosylierung, Phosphorylierung, Methylierung, Acetylierung, Ubiquitinierung, etc.

Da außer Gli2 keines der in dieser Arbeit verwendeten Fusionsprotein mit den standard XL1-Blue *E. coli* - Stamm exprimiert werden konnten, wurde zuerst der Rosetta2(DE3)pLysS Stamm getestet. Mit diesem Stamm war es möglich, 8 weitere Transkriptionsfaktoren zu exprimieren.

Der zur Zeit am weitesten weiterentwickelte Stamm Rosetta-gami B(DE3)pLysS wurde gegen Ende der Doktorarbeit für die restlichen nicht zu exprimierenden Transkriptionsfaktoren getestet. Leider konnte hier nur ein weiteres Fusionsprotein exprimiert werden.

Ein Liste der zur Verfügung stehenden Transkriptionsfaktoren, deren Expression in den verschiedenen *E. coli* - Stämmen, sowie das Ergebnis des Microarrays zeigt Tabelle 4.12. Alle diese Fusionsproteine lagen in dem pMAL-p2X Vektor vor.

FAKTOR	<i>E. coli</i>	EXPRESSION ER- FOLGREICH	CHIP ERFOLGREICH
Gli2	Xl1 Blue	Ja	Ja
SMAD4	Rosetta2	Ja	Ja
RBPJK	Rosetta2	Ja	Ja
TP53	Rosetta2	Ja	Ja
TFE3	Rosetta2	Ja	Ja
ATF1	Rosetta2	Ja	Ja
TCF4	Rosetta2	Ja	Ja
PBX1	Rosetta2	Ja	Ja
CBFbeta	Rosetta2	Ja	Ja
DP1	Rosetta2	Ja	Nein
Gli1	Rosetta2	Ja	Nein
WT1	Rosetta gami	Ja	Ja
SMAD3	Rosetta gami	Nein	Nein
MEF2C	Rosetta gami	Nein	Nein
N-MYC	Rosetta gami	Nein	Nein
BCL6	Rosetta gami	Nein	Nein
SOX4	Rosetta gami	Nein	Nein
DP2	Rosetta gami	Nein	Nein
FXRbeta	Rosetta gami	Nein	Nein
E2F1	Rosetta gami	Nein	Nein
E2F6	Rosetta gami	Nein	Nein
TCF7	Rosetta gami	Nein	Nein
Ets1	Rosetta gami	Nein	Nein

**Tabelle 4.12:** Expression der Transkriptionsfaktoren in verschiedenen *E. coli* - Stämmen sowie das Ergebnis der Chip-Experimente.

Da Ets 1 im Fusionsprotein als Vollängenkonstrukt vorlag, konnte es aufgrund seiner Größe (110 KDa) nicht in *E. coli* - Bakterien exprimiert werden.

Es ist außerdem bekannt, dass bei der Verwendung eines MBP-Tags die Expression der Fusionsproteine aus unerklärlichen Gründen nicht funktioniert (Quelle: [www.roche-applied-science.com](http://www.roche-applied-science.com)). Dies ist leider von Protein zu Protein verschieden und kann weder vorhergesagt noch vermieden werden (Kapust and Waugh (1999)).

## 4.5 Bestimmung der Bindungsspezifitäten krebsrelevanter Transkriptionsfaktoren

Expression, Aufreinigung und Konzentrationsbestimmung sowie die Nachweise der einzelnen Proteine wurden alle nach den in Kapitel "Etablierung der Methode anhand des Transkriptionsfaktors Gli2" genannten Protokollen durchgeführt.

Zur Überprüfung des Inserts wurde jedes für die Expression benutzte Plasmid sequenziert. Diese Sequenz wurde anschließend in die Aminosäuresequenz übersetzt und mit folgender Datenbank abgeglichen: (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Jedes in dieser Arbeit untersuchte Protein wurde zu 100% in der Datenbank identifiziert. Alle relevanten Daten der Sequenzierung, der Datenbankabgleiche sowie der MALDI-Analysen sind im Anhang beigefügt.

Die jeweiligen Protein und Antikörperkonzentrationen sind in Tabelle 4.13 zusammengefasst.

ANTIKÖRPER ODER PROTEIN	KONZENTRATION
Fusionsprotein	15 nM
Biotinylierter Anti-cMyc-Antikörper	0,6 nM
Streptavidin-R-Phycoerythrin	80 nM

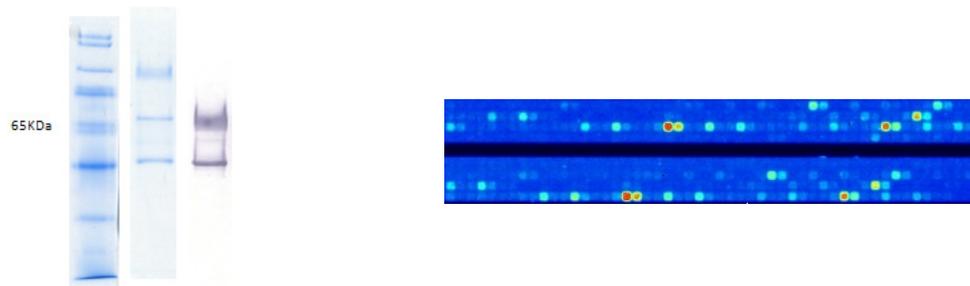
**Tabelle 4.13:** Konzentrationen der eingesetzten Proteine und Antikörper

### 4.5.1 PBX1 (Pre-B-Cell Leukemia Transcription Factor 1)

Der Transkriptionsfaktor PBX1 gehört zu der Klasse der Homöobox-Proteine. Bei dieser Proteinfamilie binden drei aufeinander folgende Alpha-Helices an die DNA, wobei die dritte Helix hauptsächlich an die große Furche der DNA bindet (siehe Abbildung 2.6). Somit ist keine Dimerisierung des Proteins zur DNA-Bindung notwendig. Die Helices eins und zwei sind ähnlich der Helix-Turn-Helix Struktur aufgebaut.

PBX1 bindet laut Literatur bevorzugt an die Sequenz 5'-ATCAATCAA-3' (Lu et al. (1995)). Eine chromosomale Abweichung, bei der auch PBX1 involviert ist, führt u.a. zu B-Zellen-Lymphomen.

Das exprimierte Fusionsprotein MBP-3xMyc-PBX1 besitzt 667 Aminosäuren, hat eine Masse von 67,5 KDa und einen Extinktionskoeffizienten von  $83900 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ . Abbildung 4.44 zeigt das SDS-Gel sowie den Western-Blot des aufgereinigten Fusionsproteins. Die Bande bei 67 KDa zeigt das Vollängenprotein. Desweiteren ist eine Bande bei 52 KDa zu erkennen. Hier handelt es sich wie in Kapitel 4.3 erwähnt um eine Abbruchsequenz während der Proteinexpression. Es konnte durch MALDI-Analyse die Abbruchsequenz als MBP-Tag inklusive Myc-Tag identifiziert werden. Auch das Vollängenprotein bei 67 KDa konnte mit der gleichen Methode nachgewiesen werden.

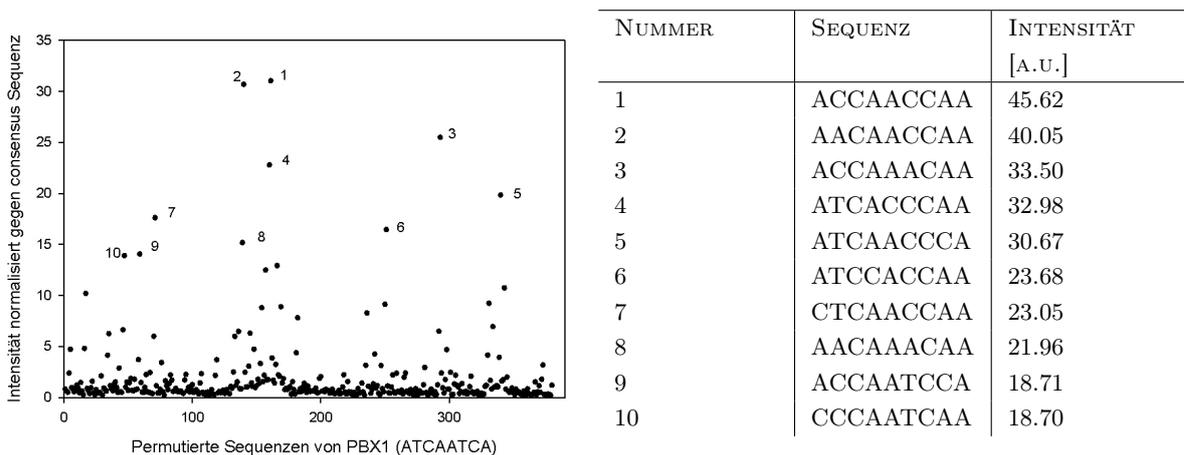


**Abbildung 4.44:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsprotein MBP-3xMyc-PBX1 (links). Ausschnitt eines Arrays mit gebundenen PBX1 Transkriptionsfaktor (rechts).

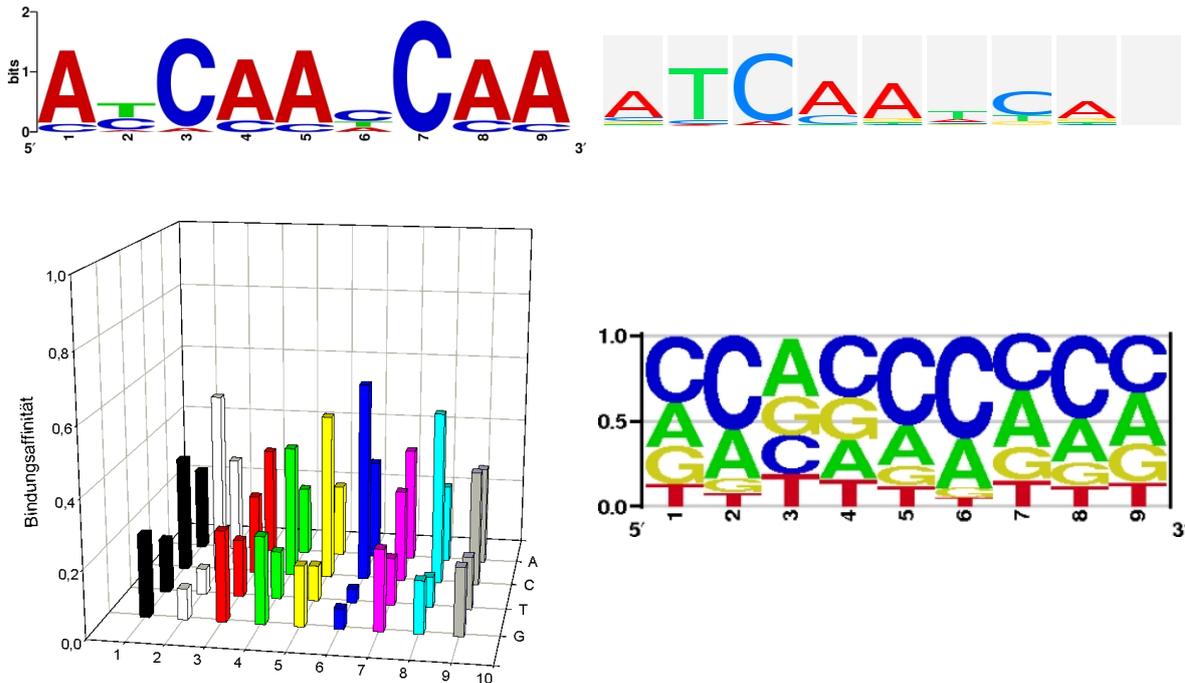
Zieht man bei der Auswertung die stärksten 5% aller gemessenen Signale der ein und zwei Basenpaarmutationen in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.46). Der Transkriptionsfaktor scheint für die DNA-Bindung die Kernsequenz CAA zu benötigen. Kommt diese Sequenz mehrfach in der Bindesequenz vor, verstärkt sich die Bindung.

Die Konsensus-Sequenz aus der Literatur zeigte in der Microarrayanalyse eine sehr schlechte DNA Bindung. Dies kann unter anderem an den fehlenden kooperativen Effekte andere Proteine liegen. Aufgrund dieser Bindung sind auch die durch die gewichtete Positionsmatrix bestimmten Werte nur teilweise auswertbar. Die Matrix sowie das dazugehörige Sequenzlogo sind trotzdem in Abbildung 4.46 gezeigt.

Bei den zweifach permutierten Sequenz konnten jedoch einige gut bindenden Sequenzen detektiert werden.



**Abbildung 4.45:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.46:** Oben: Sequenzlogo von PBX1 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Aus der Darstellung der gewichteten Positionsmatrix und des dazugehörigen Sequenzlogos ist zu erkennen, dass ein Austausch jeder Base mit einem Thymin in der Bindesequenz zu einer Abnahme des Bindeverhaltens führt. Die Kernsequenz CAA ist hier teilweise zu erkennen. Ein Basen-Austausch mit Guanin führt je nach Position zu einer schlechteren Bindung bzw. zu einem Funktionsverlust.

#### 4.5.2 RBPJK (Recombination Signal Binding Protein for Immunoglobulin Kappa J region)

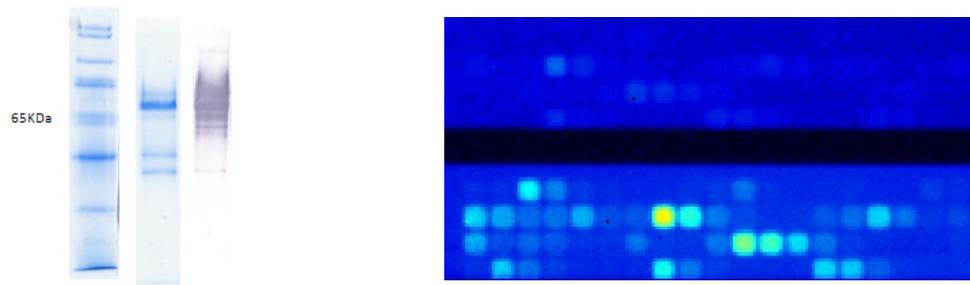
Der Transkriptionsfaktor RBPJK besitzt zur DNA-Bindung eine *Basic Domain*, die in Lösung eine zufällige Struktur besitzt und während der DNA-Bindung eine alpha Helix ausgebildet. Es sind bis jetzt noch keine allgemeine Konsensus-Sequenzen für alle zur *Basic Domain* gehörenden Transkriptionsfaktoren bekannt.

Laut Literatur bindet RBPJK an die Sequenz TTCCCACG (Lecourtois and Schweisguth (1995)). RBPJK reguliert den Notch Signalweg und agiert als Transkriptionsrepressor, wenn er nicht mit Notch Proteinen assoziiert ist. Der Notch Signalweg ist in die Zell-Zell Kommunikation sowie in einigen Apoptose Signalwegen involviert.

#### Daten des Fusionsproteins MBP-3xMyc-RBP:

- Größe: 615 Aminosäuren

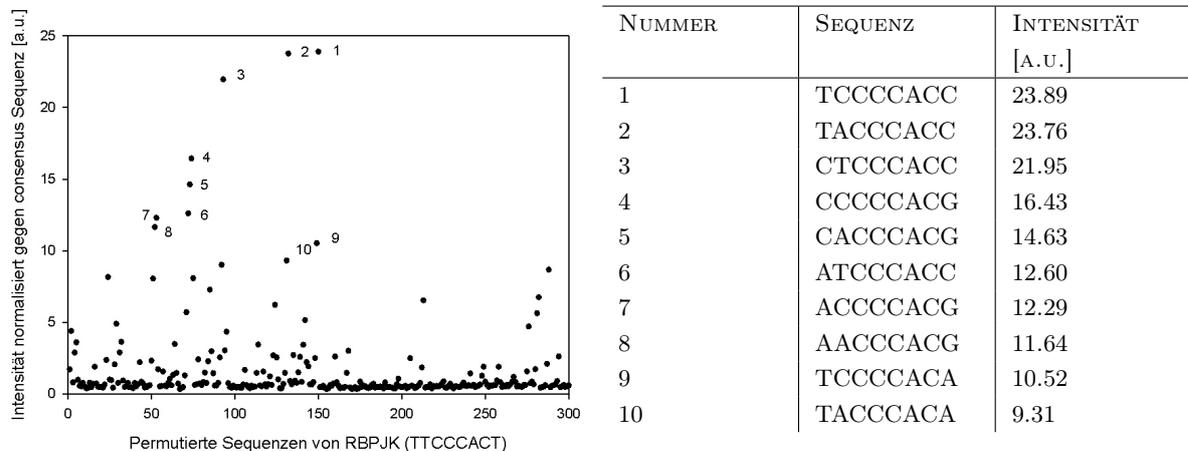
- Masse: 67,6 KDa
- Extinktionskoeffizient:  $93400 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$



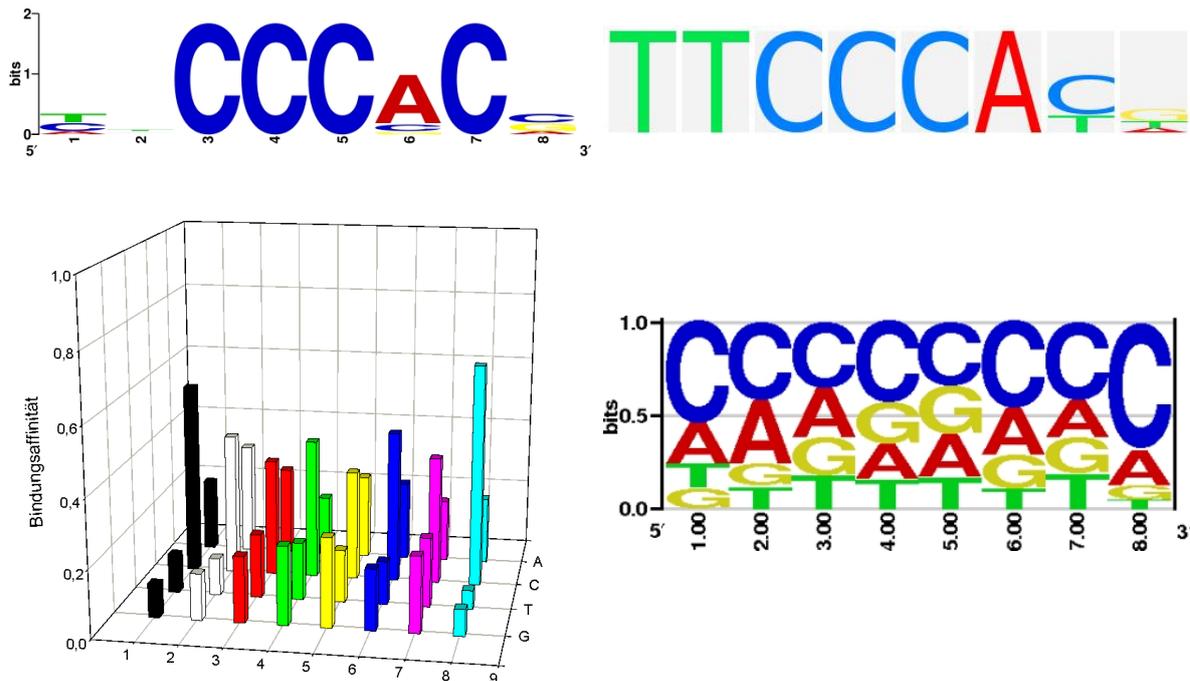
**Abbildung 4.47:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-RBPJK (links). Ausschnitt eines Arrays mit gebundenen RBPJK Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.48 gezeigt.

Das Sequenzlogo der stärksten 5% aller gemessenen Signale zeigt Abbildung 4.49. Verglichen mit der aus der Literatur bekannten Sequenz sind einige Gemeinsamkeiten zu erkennen. Der Transkriptionsfaktor scheint für die DNA-Bindung die Kernsequenz CCCAC zu benötigen. Die umgebenden Nukleotide scheinen für die DNA-Bindung keine große Bedeutung zu haben.



**Abbildung 4.48:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.49:** Oben: Sequenzlogo von RBPJK mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Aus dieser Darstellung ist deutlich zu erkennen, dass ein Austausch jeder Base mit einem Thymin in der Bindesequenz zu einer Abnahme der Bindungsaffinität führt. Es bestätigt sich die Vermutung, dass für die DNA-Bindung die Sequenz CCCAC zuständig ist. Die beiden Thymidine an der Position eins und zwei sowie das Guanin an der achten Stelle scheinen keinen Einfluss auf das Bindeverhalten des Transkriptionsfaktors zu haben.

#### 4.5.3 ATF1 (Cyclic AMP-dependent transcription factor 1)

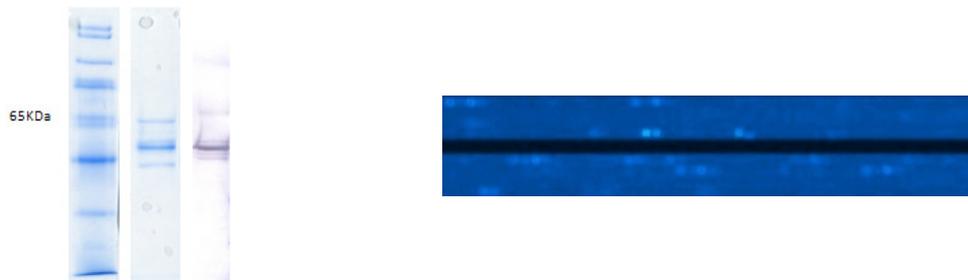
Der Transkriptionsfaktor ATF1 gehört zu der Gruppe von Transkriptionsfaktoren, die eine *Basic Domain* und einen *Leucine-Zipper* besitzen. Der DNA-Bindeseite folgt ein *Leucine-Zipper*, der an jeder 7. Stelle ein Leucin besitzt und die Dimerisierung vermittelt. Die Dimerisierung ist Voraussetzung für das Binden an DNA über die große Furche (Hai et al. (1989)).

Das *cAMP response element (CRE)* (5'-GTGACGTACAG-3') ist die aus der Literatur bekannte Bindesequenz (Hai et al. (1989)).

#### Daten des Fusionsproteins MBP-3xMyc-ATF1:

- Größe: 572 Aminosäuren
- Masse: 63,6 KDa
- Extinktionskoeffizient:  $72400 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

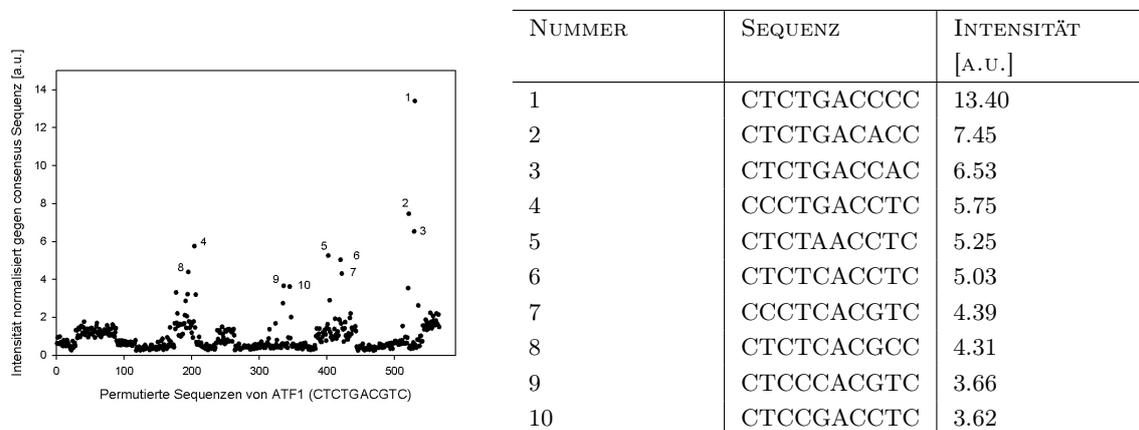
Abbildung 4.50 zeigt die Bande bei 63 KDa zeigt das Volllängenprotein an. Es ist auch ein Teilausschnitt eines Arrays mit gebundenem MBP-3xMyc-ATF1 zu sehen. Die sehr niedrigen Signale zeigen, dass wenige Proteine an die DNA gebunden sind. Dies kann an der geringen Ausgangskonzentration des eingesetzten Transkriptionsfaktors als auch an der Dimerisierung liegen, die für die Bindung vorausgesetzt wird.



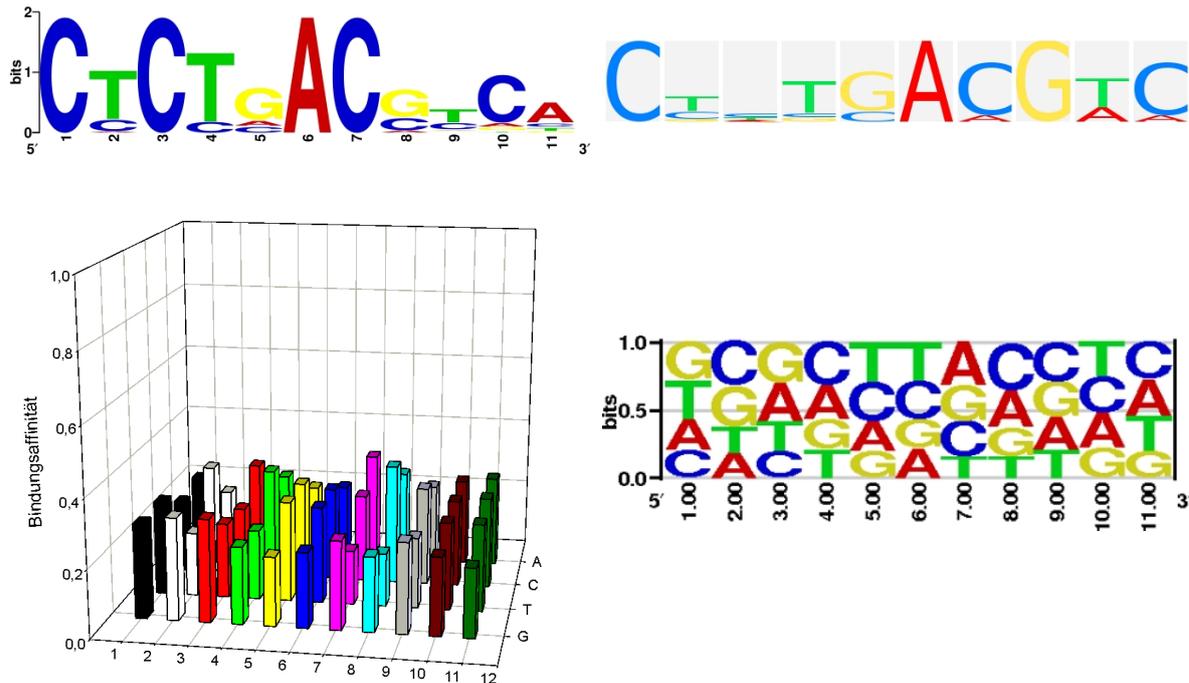
**Abbildung 4.50:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-ATF1 (links). Ausschnitt eines Arrays mit gebundenen ATF1 Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.51 gezeigt.

Bei der Auswertung der stärksten 5% aller gemessenen Signale sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen. Es sind bestimmte Basen an definierten Stellen in der Bindesequenz für die DNA-Bindung notwendig (siehe Abbildung 4.52). Sehr wichtig sind die Cytosine an den Stellen 1, 3, 7, 10 sowie das Adenin an der Position sechs.



**Abbildung 4.51:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.52:** Oben: Sequenzlogo von ATF1 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Aus Darstellung der gewichteten Positionsmatrix sowie in der dazugehörige Sequenzlogo in Abbildung 4.52 sind keine eindeutigen Erkenntnisse zu schließen, da an die Konsensus-Sequenz bzw. an die einfach mutierten Sequenzen keine eindeutige DNA-Bindung statt fand. Es gab jedoch einige zweifach mutierte Sequenzen, die sehr starke Signale lieferten (siehe Abbildung 4.51).

#### 4.5.4 TCF4 (Transcription Factor-4; Immunoglobulin Transcription Factor-2 (ITF2))

Der Transkriptionsfaktor TCF4 besitzt das *Basic helix-turn-helix* Motiv (bHLH). Für ein effizientes Binden an die DNA muss TCF4 mit einem zweiten bHLH-Transkriptionsfaktor dimerisieren. Dies kann als Homo- oder Heterodimer geschehen.

TCF4 erkennt die *Ephrussi-box* ('E-box') Bindeseite ('CANNTG'), wobei N für jede beliebige Base steht. Desweiteren sind noch folgende Bindeseiten bekannt: 5'-ACANNTGT-3' oder 5'-CCANNTGG-3'. Durch die Bindung an das Somatostatin Rezeptor Initiator Element (SSTR2-INR) aktiviert TCF4 die Transkription.

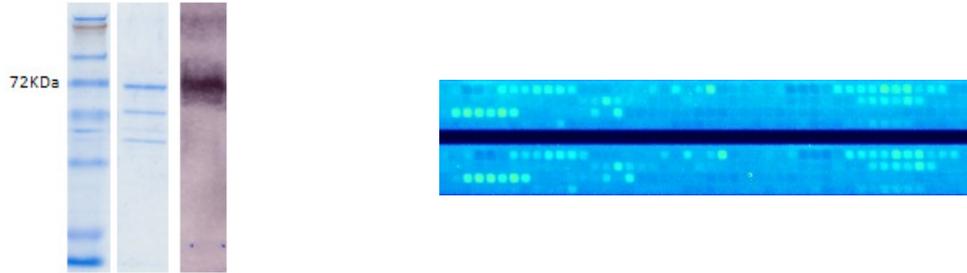
Defekte in TCF4 können das Pitt-Hopkins-Syndrom (PTHS) auslösen.

#### Daten des Fusionsproteins MBP-3xMyc-TCF4:

- Größe: 560 Aminosäuren
- Masse: 67 KDa

- Extinktionskoeffizient:  $69300 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

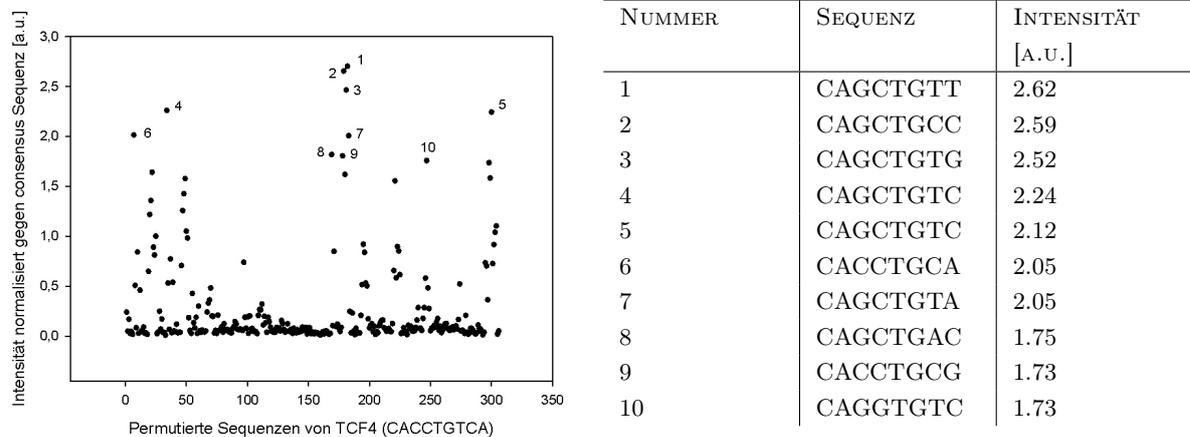
Das SDS-Gel sowie der Western-Blot von TCF4 ist in Abbildung 4.53 gezeigt. Die Bande bei 67 KDa zeigt das Volllängenprotein.



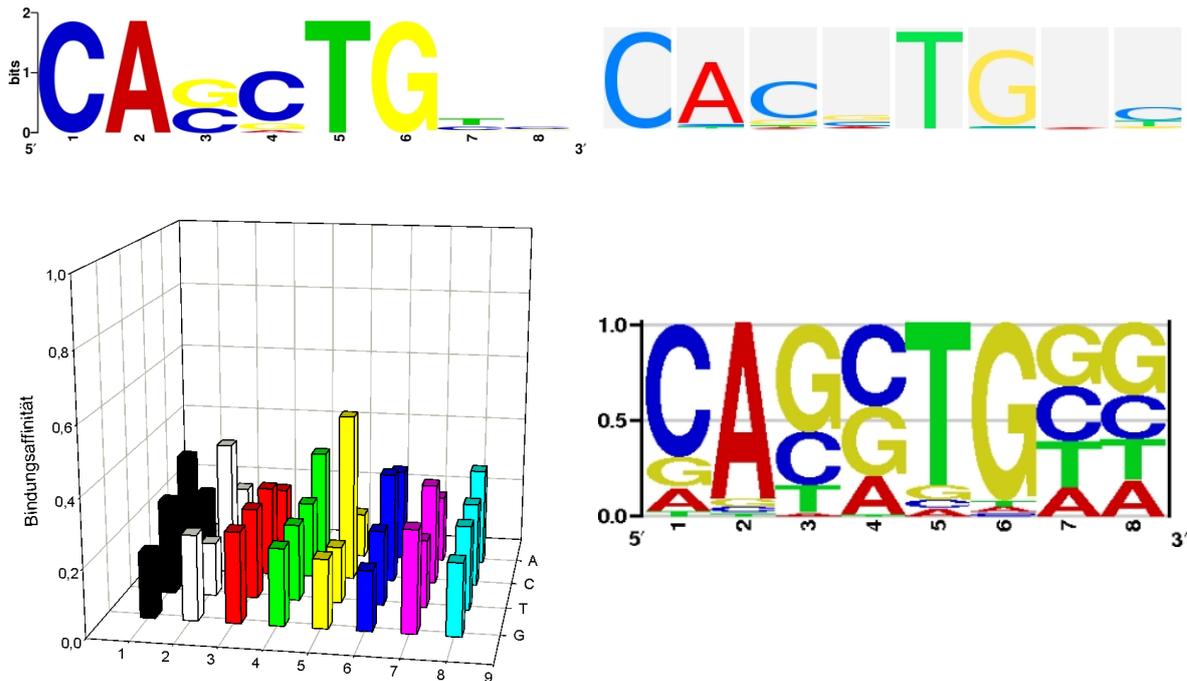
**Abbildung 4.53:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsprotein MBP-3xMyc-TCF4 (links). Ausschnitt eines Arrays mit gebundenen TCF4 Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.54 gezeigt.

Zieht man bei der Auswertung nur die stärksten 5% aller gemessenen Signale in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.55). Hauptverantwortlich für die DNA-Bindung ist das Motiv CA-TG-. Die mit - gekennzeichneten Nukleotide sind für die DNA-Bindung nicht relevant.



**Abbildung 4.54:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.55:** Oben: Sequenzlogo von TCF4 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Die gemessenen Intensitäten können auch hinsichtlich der Konsensus-Sequenz und aller einfachen Permutationen ausgewertet werden. Die gewichtete Positionsmatrix sowie das dazugehörige Sequenzlogo sind in Abbildung 4.55 gezeigt.

Aus dieser Darstellung ist deutlich zu erkennen, dass für die DNA-Bindung nur die Basen 1+2 bzw. 5+6 notwendig sind (CA–TG–). Die Mutation einer dieser Basen führt zu einem völligen Funktionsverlust des Transkriptionsfaktors.

#### 4.5.5 WT1 (Wilms tumor protein 1)

Der Transkriptionsfaktor WT1 gehört zu der Familie der *EGR C2H2-type zinc-finger* Proteine. Die Sequenz beinhaltet vier aufeinanderfolgende C2H2-Zinkfinger sowie eine Prolin und Glutamin haltige DNA Bindedomäne. WT1 erkennt und bindet die DNA Sequenz 5'-CGCCCCGC-3'. Desweiteren bindet es laut Literatur an GC-reiche und sich wiederholende TC-Sequenzen.

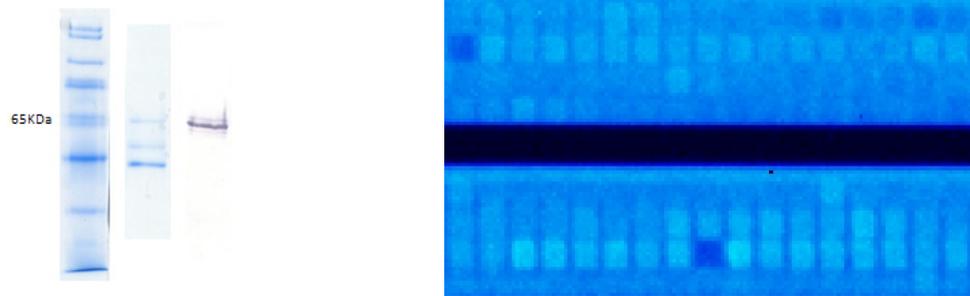
Eine Mutation dieses Transkriptionsfaktors ist u.a. verantwortlich für für den Wilm's Tumor.

#### Daten des Fusionsproteins MBP-3xMyc-WT1:

- Größe: 633 Aminosäuren
- Masse: 70,8 KDa

- Extinktionskoeffizient:  $798020 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

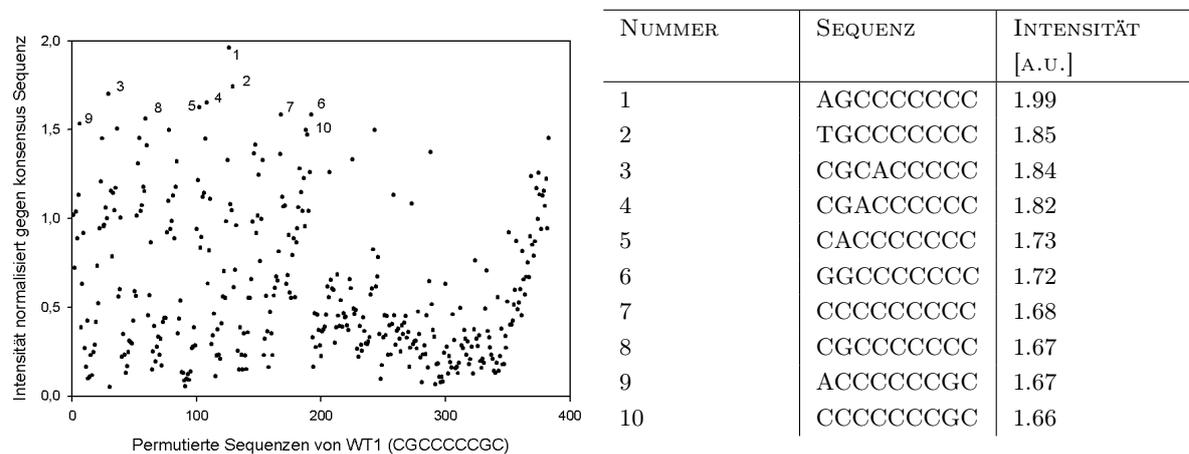
Die Bande bei 70 KDa zeigt in Abbildung 4.56 das Volllängenprotein sowie die Abbruchsequenz bei 52 KDa.



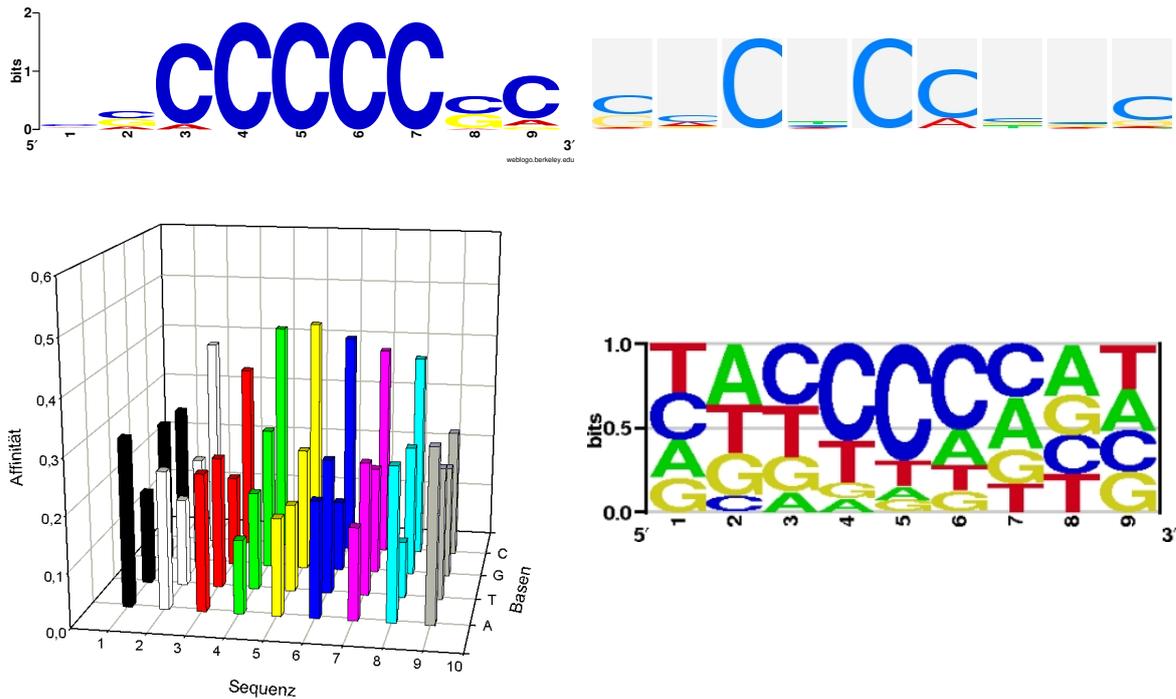
**Abbildung 4.56:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-WT1 (links). Ausschnitt eines Arrays mit gebundenen WT1 Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.57 gezeigt.

Zieht man bei der Auswertung nur die stärksten 5% aller gemessenen Signale in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.58). Fünf aufeinanderfolgende Cytosine sind hierbei für die DNA-Bindung verantwortlich. Die umliegenden Nukleotide sind hierfür nicht notwendig und können beliebig variiert werden.



**Abbildung 4.57:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.58:** Sequenzlogo von WT1 mit 5% der stärksten Signale (links) und das daraus resultierende Sequenzlogo aus der Literatur (rechts; Quelle: Transfac).

Betrachtet man bei der Auswertung das Sequenzlogo sowie die gewichtete Positionsmatrix bestätigt sich Beobachtung, dass vorwiegend die Cytosine für Bindung verantwortlich sind.

#### 4.5.6 CBFB (Core-Binding Factor, beta subunit)

Der Transkriptionsfaktor CBFB gehört zu der Gruppe der *runt homology domain* Proteine, die erstmals in *Drosophila* entdeckt wurden. Die Domäne enthält 12 beta Strände, von denen sieben die Form eines Immunoglobulin-ähnlichen beta Sandwich besitzen. Sie kommt auch in anderen Transkriptionsfaktoren wie NFκB oder p53 vor. Diese Domäne ist sowohl für die DNA-Bindung als auch für die Dimerisierung des Proteins zuständig (Baekstroem et al. (2002), Nagata et al. (1999), Berardi et al. (1999)). Die Dimerisierung führt zusätzlich zu einer Konformationsänderung in der Runt Domäne und stabilisiert somit die DNA-Bindung (Tahirov et al. (2001), Tang et al. (2000)).

Der Transkriptionsfaktor CBFB lag nicht wie die anderen Transkriptionsfaktoren als Fusionsprotein mit der Bindeseite, sondern als kompletter Transkriptionsfaktor vor.

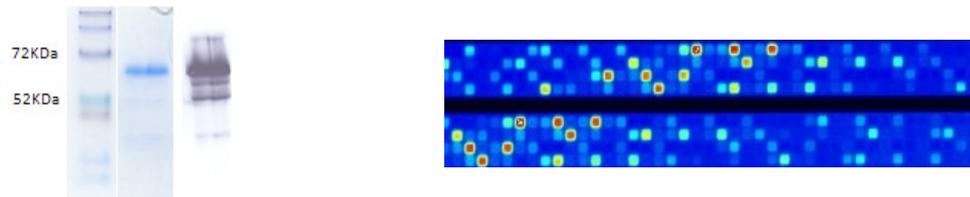
CBFB bindet spezifisch an die Kernsequenz 5'-NYGNYGGT-3', wobei N für A, C; T oder G und Y für C oder T steht.

#### Daten des Fusionsproteins MBP-3xMyc-CBFB:

- Größe: 605 Aminosäuren

- Masse: 69 KDa
- Extinktionskoeffizienten:  $69300 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

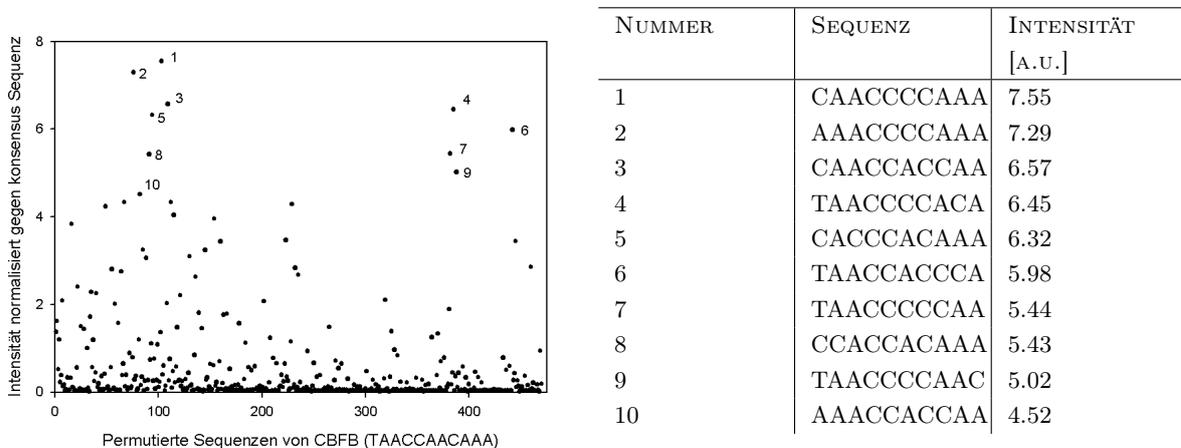
Abbildung 4.59 zeigt das SDS-Gel sowie den Western Blot des MBP-3xMyc-CBFB Fusionsproteins. Die Bande bei 67 KDa zeigt das Volllängenprotein.



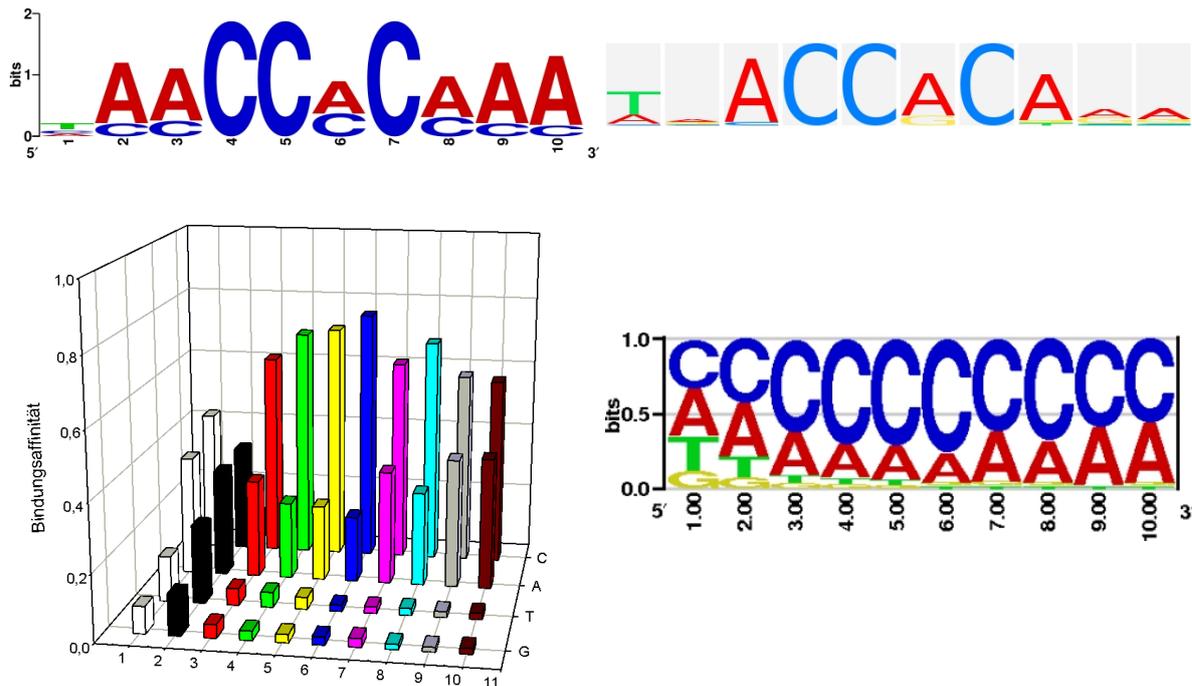
**Abbildung 4.59:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteins MBP-3xMyc-CBFB (links). Ausschnitt eines Arrays mit gebundenen CBFB Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.60 gezeigt.

Das Sequenzlogo der stärksten 5% aller Permutationen ist in Abbildung 4.61 gezeigt. Vergleicht man dies mit der Literatur, ist die gleiche Kernsequenz ACCACA deutlich zu erkennen. Am wichtigsten erscheinen die Cytosine an der Stelle 4, 5 und 7. Desweiteren sind nur die Basen Cytosin und Guanin in der Mitte der Bindesequenz enthalten.



**Abbildung 4.60:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.61:** Oben: Sequenzlogo von CBF $\beta$  mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Die gewichtete Positionsmatrix sowie das dazugehörige Sequenzlogo sind in Abbildung 4.61 gezeigt. Aus dieser Darstellung ist deutlich zu erkennen, dass ein Austausch jeder Base mit einem Thymin oder einen Guanin in der Bindesequenz zu einer reduzierten Bindung des Transkriptionsfaktors führt.

#### 4.5.7 SMAD4 (Mothers Against Decapentaplegic Homolog 4)

Der Transkriptionsfaktor SMAD4 gehört zu der Gruppe *beta-Scaffold Factors with Minor Groove Contacts*. Sie besitzen eine MH2-Domäne, die am einen Ende ein Beta Sandwich mit 3 alpha-Helices verbindet und am Anderen drei große Schleifen mit einer alpha Helix besitzt. Die MH1-Domäne beinhaltet 4  $\alpha$  Helices, 6 kurze  $\beta$  Stränge sowie 5 Schleifen. Die spezifische Erkennung und Bindung der DNA erfolgt mit der MH1-Domäne an das AGAC-Motiv mit Hilfe des 11 Aminosäure langen Beta-Hairpin.

SMAD4 fungiert als allgemeiner Vermittler der Signaltransduktion von TGF- $\beta$ . Es vermittelt die Bindung des SMAD2/SMAD4/FAST-1 Komplexes an die DNA.

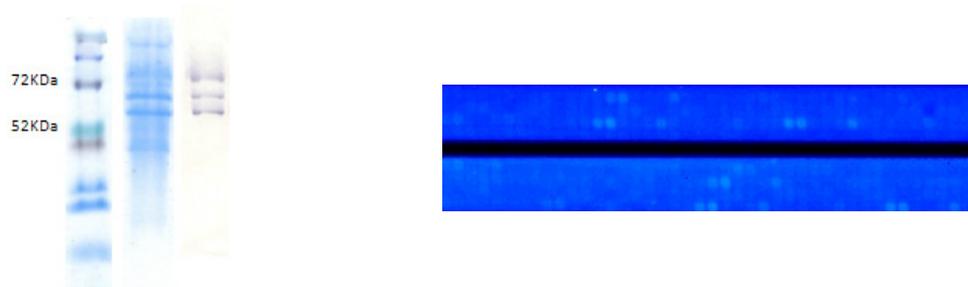
Defekte in SMAD4 können eine Ursache für Pankreas- und Darmkrebs sein.

#### Daten des Fusionsproteins MBP-3xMyc-SMAD4:

- Größe: 617 Aminosäuren
- Masse: 68 KDa

- Extinktionskoeffizienten:  $88100 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

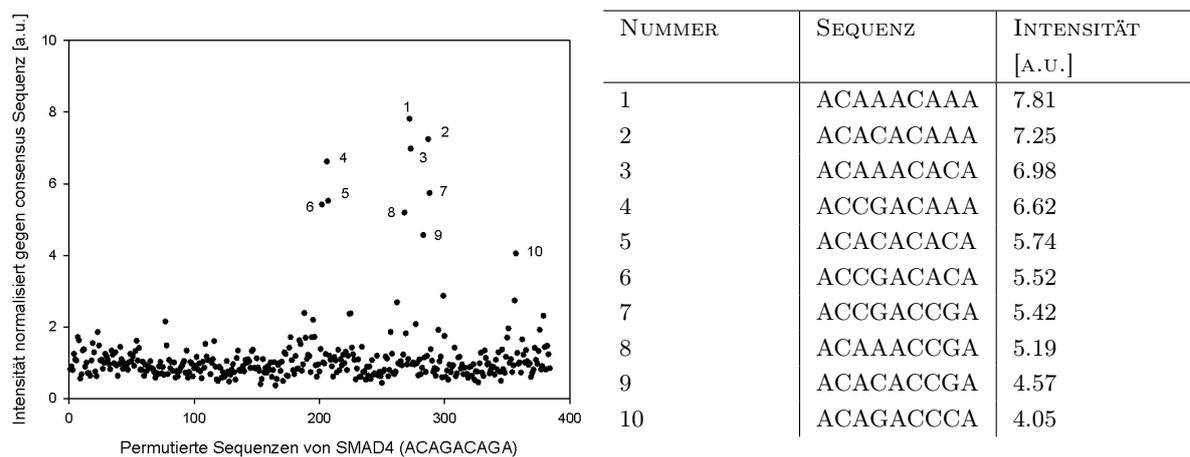
Abbildung 4.62 zeigt das Ergebnis der Western-Blot-Analyse sowie des SDS-Gels. Die Bande bei 67 KDa zeigt das Volllängenprotein. Desweiteren sind zwei Banden bei 50 bzw. 52 KDa zu erkennen. Hier handelt es sich um die Abbruchsequenzen während der Proteinexpression.



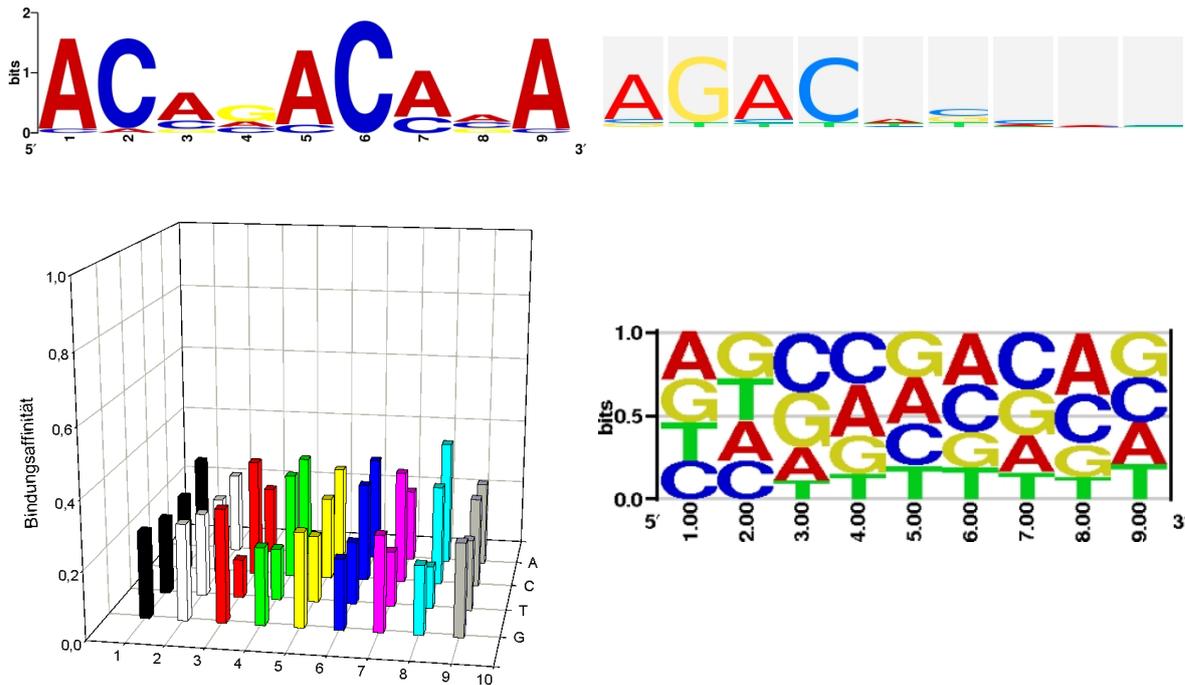
**Abbildung 4.62:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-SMAD4 (links). Ausschnitt eines Arrays mit gebundenen SMAD4 Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.63 gezeigt.

Zieht man bei der Auswertung nur die stärksten 5% aller gemessenen Signale in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.64). Die beiden Basen Adenin und Cytosin sind hauptsächlich für die DNA-Bindung verantwortlich. Im Gegensatz zur Literatur ist die Base Guanin an der 2. Stelle für die Bindung nicht notwendig.



**Abbildung 4.63:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.64:** Oben: Sequenzlogo von SMAD4 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Die gemessenen Intensitäten konnten hinsichtlich ihrer einfach permutierten Sequenzen bei diesem Transkriptionsfaktor aufgrund der mangelnden Intensitäten sowie den wenigen Signalen nicht ausgewertet werden. Die gewichteten Positionsmatrix sowie das dazugehörige Sequenzlogo sind trotzdem in Abbildung 4.64 gezeigt.

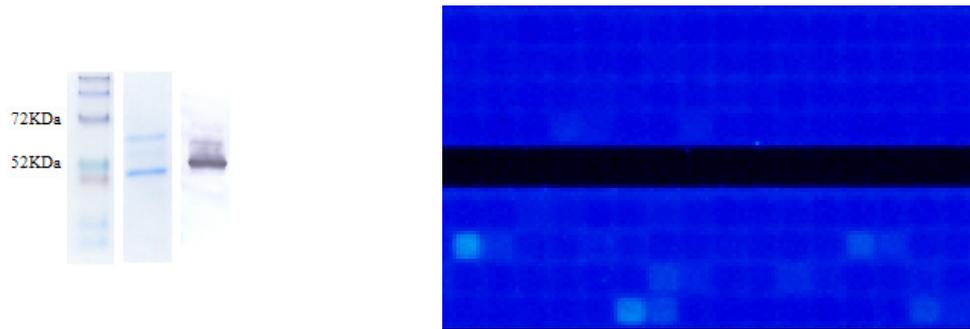
#### 4.5.8 TFE3 (Transcription Factor E3)

TFE3 ist Mitglied der Helix-Loop-Helix Familie (siehe Abbildung 2.8) und bindet an das mu-E3 Motif (CATGTGGC), an das USF1 Motiv sowie an das E-Box Motiv CACCTGCA (Henthorn et al. (1991)). Eine effiziente Bindung an die DNA erfordert eine Dimerisierung mit einem anderen HLH-Protein (Beckmann and Kadesch (1991)).

##### Daten des Fusionsproteins MBP-3xMyc-TFE3:

- Größe: 583 Aminosäuren
- Masse: 65 KDa
- Extinktionskoeffizienten:  $74800 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

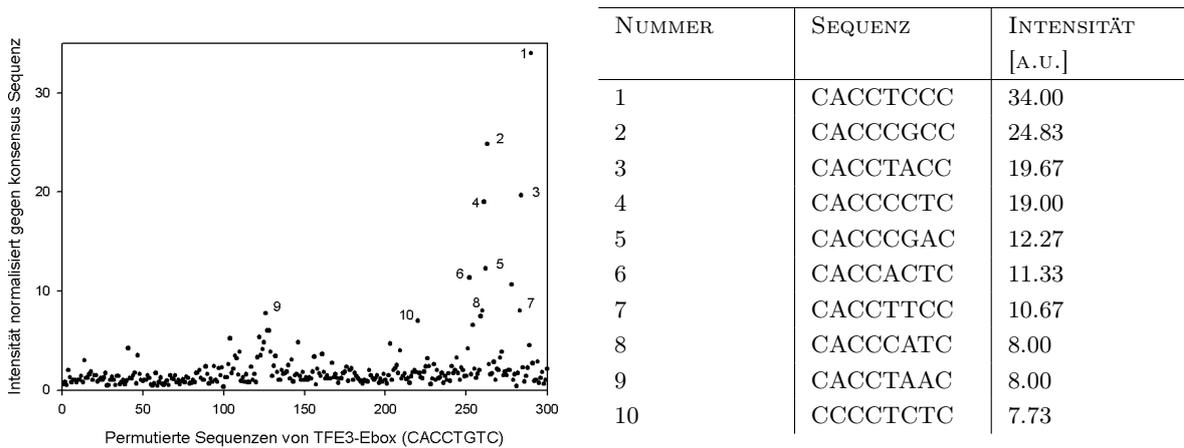
Die Bande in Abbildung 4.65 bei 65 KDa zeigt das Volllängenprotein. Die Bande bei 52 KDa zeigt eine Abbruchsequenz während der Proteinexpression.



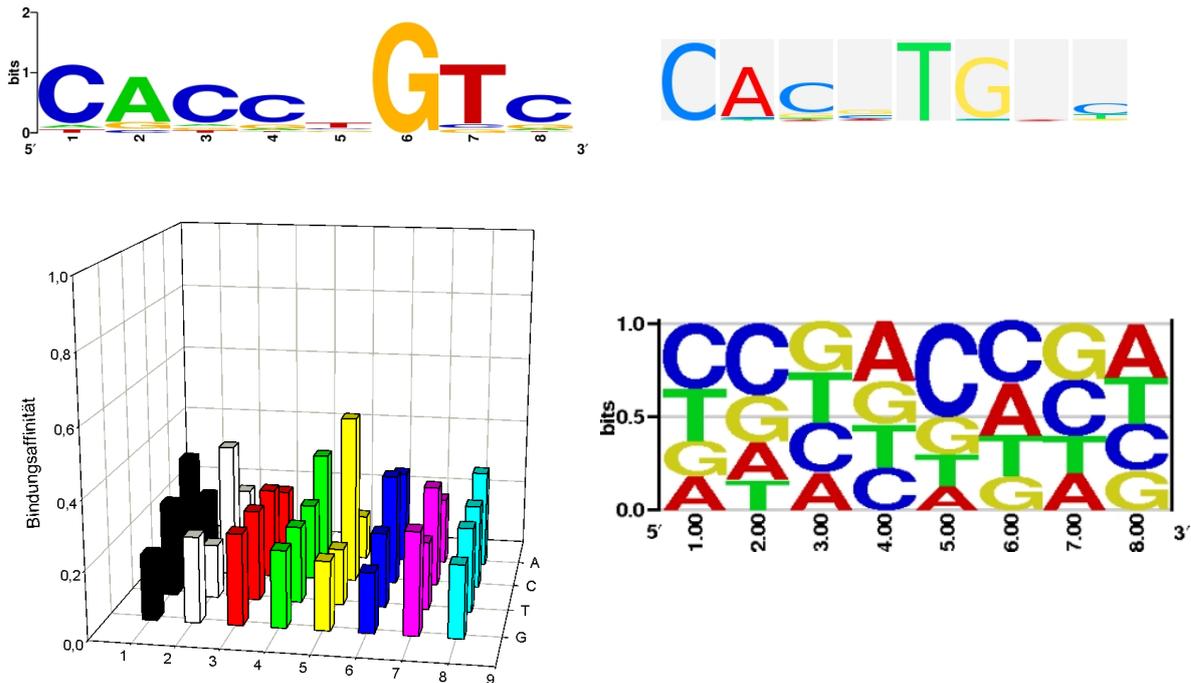
**Abbildung 4.65:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsprotein MBP-3xMyc-TTE3 (links). Ausschnitt eines Arrays mit gebundenen TFE3 Transkriptionsfaktor (rechts).

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.66 gezeigt.

Zieht man bei der Auswertung nur die stärksten 5% aller gemessenen Signale in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.67). Am wichtigsten erscheinen die Basen 1, 2, 3, 6 und 7 der Sequenz CAC-GT-.



**Abbildung 4.66:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.67:** Oben: Sequenzlogo von TFE3-Ebox Motiv (links) mit 5% der stärksten Signale sowie das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts).

Die gemessenen Intensitäten konnten hinsichtlich ihrer einfach permutierten Sequenzen bei diesem Transkriptionsfaktor aufgrund der mangelnden Intensitäten sowie den wenigen Signalen schlecht ausgewertet werden. Die gewichtete Positionsmatrix sowie das dazugehörige Sequenzlogo sind in Abbildung 4.67 gezeigt.

#### 4.5.9 p53 (Cellular tumor antigen p53)

Der Transkriptionsfaktor p53 gehört zu der Gruppe der *beta-Scaffold Factors with Minor Groove Contacts*-Transkriptionsfaktoren. Sie besitzen eine MH2-Domäne, die am einen Ende ein  $\beta$ -Sandwich mit 3  $\alpha$  Helices verbindet und am anderen drei große Schleifen mit einer  $\alpha$  Helix besitzt. In dem Helix-Motiv ist auch ein Zink-Ion eingebettet, welches jedoch keinem Zinkfingermotiv entspricht (Cho et al. (1994)). Der C-terminale Teil des Proteins kann unspezifisch an DNA-Motive binden (Foord et al. (1991)).

p53 kann sowohl als Monomer, Dimer aber bevorzugt als Tetramer an die DNA binden. Das Tetramer wird aus zwei Dimeren gebildet, wobei die Strukturen sehr komplex sind (Friedman et al. (1993), Clore et al. (1994), Jeffrey et al. (1995)).

p53 ist einer der wichtigsten Transkriptionsfaktoren für die Apoptose Signalwege. Seine wichtigsten Funktionen sind unten dargestellt:

- Funktion bei geringen Zellschädigungen (leichte DNA Schäden etc.)

- DNA Reparatur
- Zell-Zyklus Arrest
- Reaktive Sauerstoff Spezies Abbau (ROS)
- Funktion bei irreversible, starke Zellschädigung (Onkogene)
  - Apoptose, Zelltod
  - ca. 1000 Gene werden durch p53 reguliert ! Aktivierung von p53

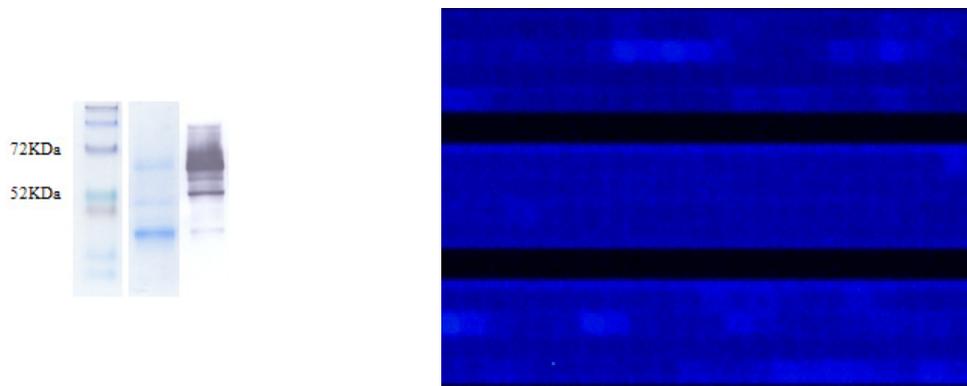
#### Daten des Fusionsproteins MBP-3xMyc-p53:

- Größe: 686 Aminosäuren
- Masse: 75,7 KDa
- Extinktionskoeffizienten:  $91400 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

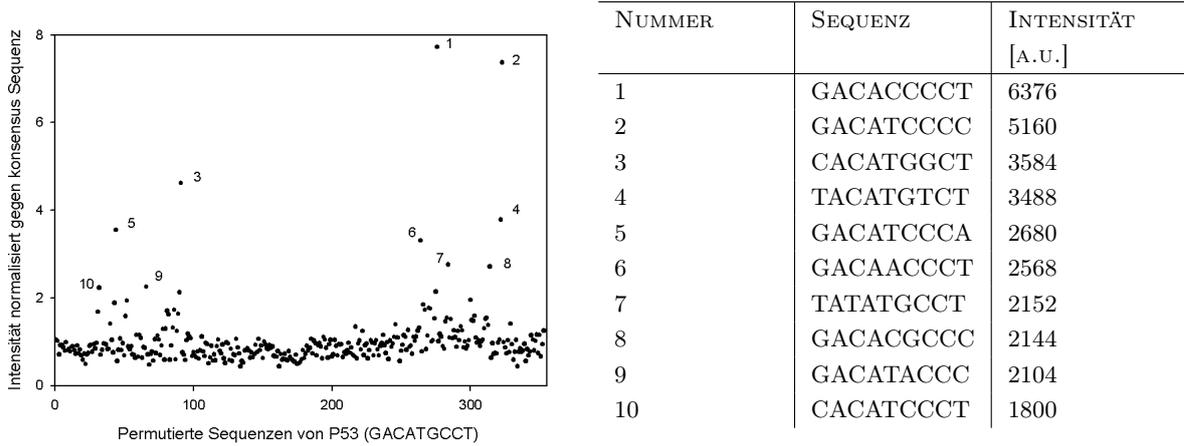
Das Gel und der Western Blot ist in Abbildung 4.68 gezeigt.

Die Intensitäten aller untersuchten ein und zwei Basen-Permutationen sowie die 10 stärksten Signale sind in Abbildung 4.69 gezeigt.

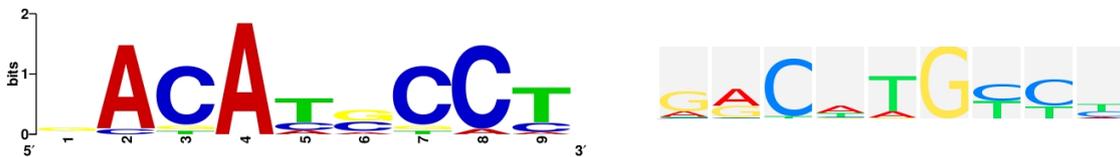
Zieht man bei der Auswertung nur die stärksten 5% aller gemessenen Signale in Betracht und berechnet aus ihnen das Sequenzlogo, sind deutliche Parallelen mit der aus der Literatur bekannten Sequenz zu erkennen (siehe Abbildung 4.70).



**Abbildung 4.68:** SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-p53 (links). Ausschnitt eines Arrays mit gebundenen p53 Transkriptionsfaktor (rechts).



**Abbildung 4.69:** Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts).



**Abbildung 4.70:** Sequenzlogo von p53-Ebox Motiv (links) mit 5% der stärksten Signale sowie das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac).

Die gemessenen Intensitäten konnten hinsichtlich ihrer einfach permutierten Sequenzen bei diesem Transkriptionsfaktor aufgrund der schlechten Intensitäten sowie den wenigen Signalen schlecht ausgewertet werden.

## 5 Diskussion und Ausblick

### Diskussion

Der Mechanismus, mit dem onkogene Transkriptionsfaktoren das Wachstum regulieren, ist bis heute noch nicht ganz verstanden. Nur wenige Ausnahmen, wie z. B. der Östrogen- oder der Androgenrezeptor wurden bis jetzt charakterisiert. Es sind zur Zeit etwa 40 Transkriptionsfaktoren bekannt, bei denen eine Veränderung der Genregulation mit Krebs in Verbindung gebracht werden kann. Darüber hinaus ist der Einfluss von Mutationen in der DNA Bindesequenz, welche das Binden von Transkriptionsfaktoren beeinflusst, noch nicht geklärt. Die Information über die spezifischen Bindestellen dieser Transkriptionsfaktoren sind meist unvollständig oder nur die Sequenz mit der höchsten Affinität ist bekannt. Viele dieser so genannten Konsensus Sequenzen werden als einzelne Sequenz angegeben. Diese Angabe führt jedoch oft zu falschen Ergebnissen. Besitzt eine Position der Bindeseite zu 75% ein Adenin und zu 25% Thymin, gehen bei der Angabe nur einer Sequenz Information verloren. Bezogen auf alle Positionen dieser Bindesequenz kann dies zu falschen Bindesequenzen führen. Desweiteren können theoretische Modelle diese Sequenzen nur unvollständig beschreiben oder vorhersagen. Von den im menschlichen Genom geschätzten 2000 Transkriptionsfaktoren sind bislang nur ein Bruchteil der Bindestellen bekannt, gut charakterisiert und in öffentlichen Datenbanken zugänglich.

Je nach Fragestellung werden zur Zeit verschiedene Methoden zur Untersuchung von Protein-DNA-Wechselwirkungen angewendet. Zu nennen wären hier vor allem SELEX (Systematic Evolution of Ligands by Exponential Enrichment), EMSA (Electrophoretic Mobility Shift Assay), ChIP-chip (chromatin immunoprecipitation-on-chip), DamID (DNA adenine methyltransferase identification) oder auch der von einem Kooperationspartner entwickelte Luziferase Assay. Mit diesen Methoden kann nach verschiedenen Bindesequenzen gesucht werden. Aber auch die Plasmonenresonanzspektroskopie oder die QCMD-Technologie können für diese Untersuchungen genutzt werden. Mit ihnen kann die Kinetik dieser Bindung bestimmt werden. Einen Nachteil haben jedoch alle diese Methoden: sie sind in Bezug auf die zu untersuchenden Sequenzen limitiert. Die Entwicklung einer Methode, die auf DNA Microarrays basiert, würde die Kosten senken und gleichzeitig die Anzahl der zu untersuchenden Proben erhöhen. Dadurch sollte es möglich sein, in naher Zukunft genomweit nach bestimmten Transkriptionsfaktorbindestellen zu suchen.

Im ersten Teil der Arbeit stand die Etablierung der auf Microarray basierenden Methode anhand des Transkriptionsfaktors Gli2 sowie ein Vergleich der Ergebnisse mit anderen Methoden im Vordergrund. Die Wahl von Gli2 zur Etablierung der Methode hatte vor allem zwei Gründe.

Zum einen gehört Gli2 zu den bisher etwa 40 mit Krebs in Verbindung gebrachten Transkriptionsfaktoren. Zum anderen hatte bereits ein Kooperationspartner ergänzende Experimente gemacht und stellte auch einen Teil der verwendeten Konstrukte zur Verfügung.

Im zweiten Teil der Arbeit wurden mit der auf Microarrays basierenden Methode die Bindungsspezifitäten krebsrelevanter Transkriptionsfaktoren bestimmt. Ausgehend von der aus der Literatur bekannten Konsensus Sequenz wurden alle möglichen Ein- und Zweibasennaarmutationen untersucht.

### **Proteinexpression und Aufreinigung**

Um Expression und Aufreinigung zu vereinfachen, wurde statt des ganzen Gli2-Proteins nur die DNA-Bindedomäne mit wenigen zusätzlichen Aminosäuren zu beiden Seiten exprimiert. Es wurde davon ausgegangen, dass sich die Bindeseiten des Proteins auch ohne weitere Domänen und umgebende Sekundärstrukturen korrekt ausbildet und die erzielten Ergebnisse aussagekräftig sind (Hallikas et al. (2006)). Es muss auch beachtet werden, dass auf die DNA-Bindung (de-)stabilisierend wirkende Kofaktoren keine Funktion ausüben können, da für ihre Bindung an Fusionsprotein ebenfalls keine Domäne vorhanden sind. Es kann also grundsätzlich nicht ausgeschlossen werden, dass die untersuchten Transkriptionsfaktoren *in-vivo* eventuell an etwas andere Sequenzen binden würden.

Zur Verfügung standen die Vektoren pMAL c2X und pMAL p2X, die die Transkriptionsfaktoren als Fusionsprotein mit MBP- und Myc-Tag exprimieren. Sie unterscheiden sich in der Lokalisierung des translatierten Proteins. Die im Cytosol oder Periplasma vorherrschenden Milieus unterscheiden sich grundlegend. So kann die oxidierend wirkende Umgebung des Periplasmas für die Expression mancher Proteine von Vorteil sein. Da dies für Gli2 jedoch nicht der Fall war, wurde der Vektor pMAL c2X bevorzugt, da die Ausbeute der Aufreinigung aus der zytosolischen Fraktion etwa zehnmal höher ist. Ein genereller Nachteil der Expression von eukaryotischen Proteinen in Bakterienzellen sind die fehlenden posttranslationalen Modifikationen. Hierzu zählen unter anderem Glykosilierungen oder Phosphorylierungen.

Das Maltose Binding Protein als "Tag" zu benutzen, hat den Vorteil, dass es aufgrund seiner Eigenschaften helfen kann, Proteine in Lösung zu halten. Demgegenüber steht der Nachteil der nicht unerheblichen Größe von etwa 40 KDa. Es war vor Beginn der Experimente nicht abzusehen, dass die Bindungseigenschaften von Gli2 dadurch nicht beeinflusst würden. Gleiches gilt auch für den nur 8 Aminosäuren großen His-Tag. Je nachdem welches Protein aufgereinigt werden soll, kann es zur Optimierung nötig sein, sowohl "Tag" als auch Orientierung (C- oder N-terminal) zu variieren. Auch den "Tag" nach der Aufreinigung proteolytisch abzuspalten kann sinnvoll sein, jedoch muss bedacht werden, dass dieser zusätzliche Schritt vielleicht auch die Funktionalität des Transkriptionsfaktors negativ beeinflussen kann. Die in dieser Arbeit verwendeten Konstrukte hatten einen N-terminalen MBP/Myc-Tag bzw. einen N-terminalen His-Tag. Die Abspaltung des MBP-Tags mittels Faktor Xa Protease war nicht erfolgreich. Die dadurch erzeugten Fragmente konnten mittels MALDI-Analyse als MBP-Tag sowie als

Untereinheiten der zur Abspaltung benutzten Protease identifiziert werden. Die Bindeseite konnte nicht detektiert werden. Dies wurde für mehrere Transkriptionsfaktoren wiederholt und bestätigt.

Eines der Hauptprobleme der Arbeit war neben der Entwicklung der Methode die Expression der funktionellen Fusionsproteine. Mit dem anfänglich verwendeten *E. coli* - Stamm konnten nur wenige Fusionsproteine in voller Länge und funktionell exprimiert werden. Der Rosetta2 *E. coli* - Stamm war in der Lage acht weitere Proteine zu exprimieren. Gegen Ende der Arbeit wurde versucht, die restlichen Fusionsproteine in Rosetta gami *E. coli* - Bakterien zu exprimieren. Dieser Stamm ist der zur Zeit am weitesten entwickelte. Er ist in der Lage Schwefel-Schwefel-Bindungen zu knüpfen und kann sieben *rare codons* lesen. Da manche der *E. coli* - Proteasen deletiert wurden, sind die produzierten Proteine vor enzymatischer Spaltung geschützt. Die Expression aller Proteine war auch mit diesem Stamm leider möglich. Der Grund hierfür ist der Abbruch der Expression des Fusionsproteins nach dem N-terminalen MBP Tag. Dies ist von Protein zu Protein verschieden und kann bis jetzt noch nicht vollständig erklärt werden.

Nachdem die Proteinaufreinigung durch die Optimierung der Waschschriffe gute Ergebnisse lieferte, fiel auf, dass bei beiden verwendeten Techniken neben der erwarteten Bande in SDS-Gelen jeweils eine zusätzliche erschien. Bei der Aufreinigung von Gli2 mit MBP-Tag über eine Amylose-Matrix lag diese Bande bei etwa 95 kDa und somit über der erwarteten von 76 kDa. Nach der Aufreinigung von Gli2 mit His-Tag mittels FPLC wurden Banden bei 27 kDa und 22 kDa detektiert. Unterschiede in der räumlichen Struktur als Erklärung sind ausgeschlossen, da die Proben auf ein denaturierendes SDS-Gel aufgetragen wurden. Beide zusätzliche Banden zeigen die gleiche Affinität wie das gewünschte Produkt zu der jeweiligen Matrix. Dies spricht dafür, dass auch beide den verwendeten Tag beinhalten und es sich um unterschiedlich lange Varianten des Fusionsproteins handelt. Durch die MALDI-Analyse konnte dies bestätigt werden. Beide Banden beinhalten ein und dasselbe Protein, wobei die kleinere Bande eine Abbruchsequenz des Volllängenprotein beinhaltet. Dies kann durch den Abbruch der Expression aufgrund von *rare-codons* erklärt werden. Dieses Problem trat bei fast allen Fusionsproteinen auf und konnte nicht verhindert werden. Würde man zur Aufreinigung einen zweiten C-terminalen His-Tag verwenden, könnten diese Probleme behoben werden.

### **Microarray-Experimente**

Für die ersten Microarray Versuche wurde das etwas ältere Geniom 1X System benutzt. Dieses System ist auf 4 Arrays mit weniger Spots limitiert, bietet jedoch für die Entwicklung der Methode entscheidende Vorteile. Der wohl Wichtigste war die manuelle Unterbrechung der Synthese. Somit konnten die Chemikalien nachgefüllt und Oligonukleotide bis zu einer Länge von 48 Basenpaaren synthetisiert werden. Mit dem neueren Geniom 2X System konnten zu diesem Zeitpunkt nur Oligonukleotide mit bis zu 30 Basen synthetisiert werden. Ein weiterer Vorteil war die manuelle Bedienung der Waschprogramme sowie die manuelle Einstellung der

Flussrate während der Waschprozesse. Bei dem Geniom 2X System werden diese Einstellungen in vorgefertigten Profilen zur Verfügung gestellt.

Mit den Versuchen am Geniom 1X konnte wurde die Methode zur Bestimmung von Protein-DNA-Wechselwirkungen entwickelt und etabliert. Die Synthese eines 48 bp langen Oligonukleotids konnte durch das Stoppen der Synthese und das Auffüllen der Chemikalien erreicht werden. Das in *E. coli* - Bakterien exprimierte Fusionsprotein MBP-Gli2 wurde aufgereinigt und an die verschiedenen doppelsträngigen DNA Sequenzen gebunden. Dies konnte durch die Markierung mit einem biotinylierten Anti-cMyc-Antikörper und eines Streptavidin-Phycoerythrin-Konjugates nachgewiesen werden. Bei der Auswertung der Ergebnisse wurden die Signalintensitäten aller möglichen Sequenzen gegen die Konsensus Sequenz normalisiert. Verwendet wurden dabei mehrere Replikate innerhalb eines Arrays. Die Waschprozesse während der verschiedenen Inkubationen wurden ebenfalls optimiert.

In Kooperation mit febit wurde danach ein Hard- und Softwareupdate an dem Geniom 2x System durchgeführt. Es war nun auch hier möglich Oligonukleotide bis zu einer Länge von 60 bp zu synthetisieren. Mit den Erkenntnissen aus den Versuchen mit dem 1X System konnten die notwendigen Inkubations- und Waschprofile auf dieses System übertragen werden. Die Reproduzierbarkeit dieser Experimente war jedoch anfänglich nicht immer gewährleistet. Dies konnte auch durch die Optimierung der Proteinaufreinigung nicht verbessert werden. Es wurde deshalb ein Glasobjektträger mit verschiedenen Konzentrationen der Konsensus Sequenz von Gli2 (GACCACCA) hergestellt und mit dem Fusionsprotein MBP-3xMyc-Gli2 inkubiert. Hiermit sollte der Einfluss der synthetisierten DNA hinsichtlich der Konzentration auf dem Chip überprüft werden. Da dies bei den kommerziell erhältlichen Chips der Firma febit nicht ohne Weiteres möglich war, wurde dieser Versuchsaufbau gewählt. Es konnte gezeigt werden, dass die Konzentration der DNA auf der Oberfläche einen erheblichen Einfluss auf das Binden des Transkriptionsfaktors hat. Ist die DNA-Konzentration zu hoch, kann aufgrund sterischer Wechselwirkungen keine Protein-DNA-Bindung erfolgen. In Zusammenarbeit mit febit wurde ein Chip mit nur 10% der herkömmlichen DNA-Konzentration hergestellt und getestet. Erreicht wurde dies durch eine Modifizierung der Aminosilan-Oberfläche. Mit diesem Chip konnten die Probleme der mangelnden Reproduzierbarkeit der Experimente behoben werden. Der Erfolg bei den folgenden Versuchen lässt einfach erklären. Bei zu hoher DNA-Dichte ist nicht gewährleistet, dass die Einzelstränge die zum Gelingen der Experimente benötigte Haarnadelstruktur korrekt ausbilden. Wegen der räumlichen Nähe der Moleküle zueinander ist die Bildung von netzwerkartigen Strukturen vorstellbar. Dass diese dann eine Bindung des Proteins verhindern, ist verständlich. Auch sterische Wechselwirkungen des Fusionsproteins können trotz einer korrekt ausgebildeten Haarnadel ein Binden verhindern. Ein Gelingen der Experimente mit dem alten Chip ist beispielsweise durch eine nicht optimal verlaufende Synthese zu erklären.

Eine doppelsträngige DNA kann auf einem Microarray auf zwei verschiedene Methoden hergestellt werden. Die einfache Methode ist die Synthese einer selbstkomplementären Haarnadelstruktur. Nach Erhitzen und langsamen Abkühlen bildet sich diese von selbst aus. Bei der

zweiten Methode wird der komplementäre Strang mittels *Primer Extension* hergestellt. Dies konnte durch den Einbau fluoreszenzmarkierter Cytosine bestätigt werden. Ein erfolgreiches Binden der Fusionsproteine an diese Sequenzen konnte nicht beobachtet werden. Da die erste Methode erfolgreich funktioniert, wurde auf eine Weiterentwicklung verzichtet.

Nach der Optimierung der Optik des Geniom 2X Systems konnte die Anzahl der zu synthetisierenden Proben von 6776 auf 15299 erhöht werden. Dies wurde durch eine Veränderung der Spiegelkonfiguration und einer Reduktion der Spotdurchmesser der einzelnen Sonden von 34  $\mu\text{m}$  auf 17  $\mu\text{m}$  erreicht. Durch diese Reduktion war es möglich, die durch die de Bruijn Methode hergestellten Sequenzen, die alle möglichen 9 mere beinhalten, als 2 Replikate auf dem Chip zu synthetisieren. Mit dem Fusionsprotein MBP-Gli2 wurde diese Weiterentwicklung getestet. Leider waren die Ergebnisse dieser Experimente nicht direkt mit den auf dem normalen Array zu vergleichen. Durch die Reduktion der Spotdurchmesser wurde auch die Sensitivität der Sonden geschwächt. Somit konnten nur noch die Signale der stark bindenden Sequenzen detektiert werden. Durch die Verwendung der Signalamplifikation konnten diese Probleme teilweise verbessert, jedoch nicht komplett beseitigt werden.

Mit Hilfe der dynamischen Detektion konnten die Temperaturprofile für die Protein-DNA-Bindungen bestimmt werden. Es konnte gezeigt werden, dass verschiedene Permutationen unterschiedliche Schmelztemperaturen besitzen. Stärkere Bindungen werden erst bei höheren Temperaturen gelöst. Die Schmelztemperatur der Konsensus Sequenz von Gli2 betrug 37°C. Schlechter bindende Sequenzen wurden bei 36°C, bessere bei 39°C detektiert. Die für den menschlichen Organismus relevante Konsensus Sequenz ist somit bei Körpertemperatur am besten wirksam.

Um die durch die Microarray-Analyse gewonnenen Ergebnisse von Gli2 zu validieren, wurden zwei weitere Methoden verwendet. Mit Hilfe der Plasmonenresonanzspektroskopie konnten die Kinetiken der Gli2-DNA Bindung bestimmt und mit der Literatur verglichen werden. Verwendet wurden hierfür die Konsensus Sequenz sowie zwei Permutationen, die durch die Microarray-Analyse bestimmt wurden. Eine Permutation zeigte bessere Bindungseigenschaften als die Konsensus, eine schlechter. Der Vergleich der Bindungskinetiken lieferte übereinstimmende Ergebnisse mit der Literatur. Somit konnte die Funktionalität des Fusionsproteins sowie eine intakte Protein-DNA-Bindung auch für die Microarray-Experimente vorausgesetzt werden. Mit Hilfe der Quartz-Kristall-Mikrowaage konnte die Protein-DNA-Bindung hinsichtlich ihrer Stöchiometrie untersucht werden. Es war ebenfalls möglich, die Dicke der Schichten sowie die Viskosität zu bestimmen. Der Vergleich der verschiedenen Mutationen in der Bindesequenz lieferte zum Teil die zu erwartenden Ergebnisse. Die schlecht bindende Permutation besaß die geringste Schichtdicke sowie die wenigsten Moleküle pro Oberfläche. Jedoch zeigte die Konsensus Sequenz durchgehend das beste Bindeverhalten, was im Widerspruch zu den Microarrayergebnissen steht. Leider gab es für dieses Experiment keinen direkten Vergleich mit der Literatur. Sie bestätigten jedoch die Funktionalität des Fusionsproteins. Der Unterschied dieser Ergebnisse kann durch nicht korrekt ausgebildete Haarnadelstrukturen erklärt werden. Eine Wiederholung dieser Experimente sollte die Microarray-Daten bestätigen.

Es wurden in dieser Arbeit die Bindungsspezifitäten von 10 Transkriptionsfaktoren bestimmt. Ausgehend von der aus der Literatur bekannten Konsensus Sequenz wurden alle möglichen Ein- und Zweibasenaarmutationen untersucht. Dies wurde für die Transkriptionsfaktoren in dieser Arbeit gezeigt. Die Protein-DNA-Wechselwirkung der verschiedenen Transkriptionsfaktoren war jedoch sehr unterschiedlich. Einige zeigten sehr starke Signale, bei Anderen konnten nur sehr schwache Signale detektiert werden. Dies mag zum einen an den teilweise sehr geringen Konzentrationen der eingesetzten Proteine liegen. Einige Transkriptionsfaktoren konnten nur in geringen Konzentrationen als Vollängenprodukt erhalten werden, da die Expression nach dem MBP-Tag abbrach. Da über diesen "Tag" die Proteine mittels Affinitätschromatographie aufgereinigt wurden, konnten sie nicht voneinander getrennt werden. Die gesamte Proteinkonzentration konnte nicht erhöht werden, da die Abbruchsequenzen den Hintergrund zu stark erhöhten. Auch die Konformation der Proteine trägt sehr stark zu den beobachteten Effekten bei. Die Signalintensitäten der Proteine, die für eine DNA-Bindung dimerisieren müssen, waren im Schnitt geringer. Ein Einfluss des relativ großen MBP-Tags auf die Dimerisierung kann deshalb nicht ausgeschlossen werden. Versuche den Tag mittels Proteasen abzuspalten und nur die Bindeseite für die Experimente zu benutzen schlug fehl, da diese ohne Tag nicht löslich bzw. nicht stabil war.

Die durch die Microarray-Experimente gewonnen Daten konnten mit der Literatur nicht direkt verglichen werden, da nur die Bindeseite anstatt des gesamten Transkriptionsfaktors verwendet wurde. Es sind trotzdem eine Vielzahl von Parallelen zu erkennen. Es wurde sehr oft eine Kernsequenz in der Bindeseite gefunden, die hauptsächlich für die Bindung verantwortlich ist. Diese Kernsequenz ist auch in den Literaturdaten wiederzufinden. Ein sehr gutes Beispiel hierfür ist die Sequenz CA-TG- bei dem Transkriptionsfaktor TCF4 sowie -AACCA bei CBFβ. Das Fusionsprotein CBFβ zeigte in dieser Arbeit mit Abstand das beste Bindevverhalten. Dies könnte unter anderem an der Verwendung des kompletten Transkriptionsfaktors anstatt der Bindedomäne liegen. CBFβ ist relativ klein und konnte somit komplett in das Fusionsprotein integriert werden. Bei beiden Proteinen sind die angegebenen Konsensus Sequenzen maßgeblich für die Bindung verantwortlich, die umliegende Sequenzen können variiert werden. Mutationen dieser Kernsequenz führen zur Abnahme des Bindevhaltens. Hier zeigt sich, dass der Austausch dieser wichtigen Basen als besonders kritisch hinsichtlich der Proteinbindung ist. Dies lässt sich durch strukturelle Veränderung in der DNA erklären. Die Bindung des Proteins an die Bindedomäne funktioniert ähnlich dem Schlüssel-Schloss-Prinzip. Für die Spezifität und Stabilität der Bindung sind Wasserstoffbrücken, Ionenbindungen sowie hydrophobe Wechselwirkungen verantwortlich. Wird dieses perfekt aufeinander abgestimmte System verändert und die Struktur des DNA bindenden Motivs passt nicht mehr exakt zu der Form der DNA ist die Affinität reduziert. Leicht vorstellbar ist dies beispielsweise bei einem Austausch von Pyrimidin gegen Purin. Auch der gegenteilige Fall ist denkbar. Es wurden auch stärker bindende Sequenzen als die Konsensus-Sequenz gefunden wurden. Ein gutes Beispiel hierfür ist der Transkriptionsfaktor Gli2. Ersetzt man in der Konsensus Sequenz GACCACCA die erste Base durch eine Cytosin oder ein Adenin, nimmt die Affinität deutlich zu. Da Gli2 bevorzugt an die Sequenzen mit diese beiden Basen bindet, ist dies durchaus verständlich. Die häufigste

natürlich vorkommende Bindesequenz muss also nicht unbedingt die Stärkste sein. Wichtig ist, dass sie schnell und effektiv regulierbar ist. Dies kann man sich anhand der Experimente bei der dynamischen Detektion gut erklären. Schlecht bindende Permutationen lösten sich bei Temperaturen unter 37°C von der DNA. Bei besser bindende Permutationen konnte dies erst bei 39°C beobachtet werden.

Da diese Arbeit in ein Projekt eingebunden ist, das es zum Ziel hat die Bindungseigenschaften aller mit Krebs in Verbindung stehenden Transkriptionsfaktoren zu untersuchen, sollen hier einige Methoden zur Bestimmung von Protein-DNA Interaktionen erwähnt werden. Diese basieren zum größten Teil auf Gel-Elektrophorese oder Microarray-Technologie. Aber auch Methoden wie Systematische Evolution von Liganden durch exponentielle Anreicherung (SELEX) können hierfür verwendet werden. Bei der SELEX-Methode werden Molekülbibliotheken mit dem zu untersuchenden Zielmolekül (z.B. ein Transkriptionsfaktor) inkubiert. Nach mehreren Wasch- und Selektionsschritten wird der mit dem Zielmolekül interagierende Ligand (z.B. DNA) mittels Sequenzierung nachgewiesen. Roulet et al. (2002) sequenzierten mehr als 1000 Liganden, um das Bindungsverhalten von CTF/NFI mittels SELEX zu untersuchen. Den Kosten von 10 € pro Sequenzierung stehen zusätzlich viele Wochen Laborarbeit gegenüber.

*DNAse footprinting* und *electrophoretic mobility shift assay* (EMSA) sind Methoden, bei denen die Protein-DNA-Komplexe mittels Gelelektrophorese aufgetrennt und analysiert werden. Hierbei ist die Anzahl der zu untersuchenden Proben sehr begrenzt. Da die Interaktionen nicht direkt gemessen werden können, ist die Sensitivität dieser Methoden sehr beschränkt. Auch die Microarray-Technologie wird für diese Untersuchungen verwendet. Man kann hier zwischen der Probengewinnung *in-vivo* und *in-vitro* unterscheiden.

Bei *DNA adenine methyltransferase identification*-Technologie (DamID) wird das zu untersuchende Protein als Fusionsprotein mit Dam *in-vivo* überexprimiert. Nach Binden des Transkriptionsfaktors an seine Zielsequenz werden in der näheren Umgebung Adenine, in GATC-Sequenzen, durch Dam *in-vivo* methyliert. Anschließend können die durch Methyl-spezifischen Restriktionsverdau erhaltenen DNA-Fragmente mit Kontrollen verglichen werden. Dies geschieht nach Amplifizierung und Markierung der DNA ebenfalls durch Microarrayanalyse. Etabliert wurde diese Methode für die Spezies *Drosophila* und *Arabidopsis*.

Bei der CHip-Chip-Technologie werden die Protein-DNA Komplexe *in-vivo* mit einem Crosslinker wie Formaldehyd inkubiert. Nach dem Lysieren der Zellen können diese Komplexe aufgereinigt und untersucht werden. Der größte Vorteil der ChIP-Chip-Methode ist die Analyse posttranslationalen Modifikationen und kooperativer Effekte anderer Bindungspartner. Ein Nachteil der oben genannten Methoden ist die Länge der untersuchten DNA. Sie ist größer als die eigentliche Bindesequenz und limitiert das Auflösungsvermögen dieser Techniken.

Mit der Plasmonenresonanzspektroskopie können Protein-DNA-Wechselwirkungen zur Zeit am genauesten untersucht werden. Es ist möglich, sowohl die Bindungskonstanten als auch die Bindungsenergien der Protein-DNA Bindung zu bestimmen. Der größte Nachteil dieser Methode ist die limitierte Anzahl der Proben. Eine recht neue Methode zur Bestimmung von Protein-DNA-Interaktionen sind die sogenannten Protein bindenden Microarrays (PBMs).

Hierbei werden in *E. coli* - Bakterien oder Insektenzellen exprimierte Fusionsproteine auf DNA Microarrays untersucht. Mukherjee et al. (2004) entwickelten eine Methode basierend auf gespotteten DNA-Chips. Im Gegensatz dazu verwenden Bulyk (2006) und Warren et al. (2006) synthetisierte oder kommerziell erhältliche Microarrays, die alle möglichen Permutationen einer 10 bp langen Sequenz beinhalten. Diese Methoden sind sehr effektiv, wenn keine Informationen über die Bindeseiten vorliegen. Wie bei allen PBMs können leider nicht die Einflüsse von Kofaktoren untersucht werden und auch eine Beeinträchtigung der Bindungseigenschaften durch das Fusionsprotein kann nicht ausgeschlossen werden.

Im Gegensatz zu den kommerziell erhältlichen Chips erlaubt das Geniom als Microarray-Plattform durch seine Flexibilität eine freie Synthese von bis zu 125.000 DNA-Sequenzen pro Chip. Da auch die Inkubation, die Waschschriffe sowie die Detektion in einem Gerät erfolgen, ist man nicht auf vorgefertigte Microarrays oder externe Scanner angewiesen. Durch die Verwendung eines bestimmten Tags benötigt man auch keine spezifischen Antikörper gegen das zu untersuchende Fusionsprotein. Da der Chip in 8 Arrays aufgeteilt ist, können mehrere Proben parallel inkubiert und detektiert werden. Es können einzelne Mutationen als auch eine große Anzahl verschiedener Sequenzen analysiert werden. Sofern die DNA-Bindedomäne nicht größer als 9 bp ist und dadurch die zur Verfügung stehende Anzahl an Spots nicht überschritten wird, ist es theoretisch möglich alle Permutationen dieser Sequenz zu untersuchen und zu analysieren. So kann man auch schwache Bindesequenzen erkennen und erhält auch Informationen über Sequenzen, die ein Binden des Proteins nicht erlauben.

## Ausblick

Im Rahmen dieser Arbeit wurde eine Methode entwickelt, die es erlaubt Protein-DNA-Interaktionen mit Hilfe der Microarray-Technologie zu untersuchen. Voraussetzung sind sauber aufgereinigte, funktionelle und mit einer Markierung versehene Proteine. Es ist möglich, mit einem Biochip in 8 Arrays verschiedene Proteine parallel zu analysieren. Mit dem *low density chip* (34  $\mu\text{m}$  Spots) können 6776 verschiedene Sequenzen pro Array untersucht werden, während mit dem *high density chip* (17  $\mu\text{m}$  Spots) 15299 möglich sind. Die Auflösung ist jedoch bei dem *high density chip* limitiert. Dies kann durch eine Signalamplifikation verbessert werden. Die aus der Literatur bekannten Konsensus-Sequenzen konnten bei vielen Transkriptionsfaktoren als Sequenzen mit hoher Affinität identifiziert werden. Die Ergebnisse zeigen weiterhin, dass der Austausch einiger Basen hinsichtlich der Proteinbindung besonders kritisch ist. Bis jetzt sind jedoch nur wenige Mutation in diesen Regionen bekannt, die Krankheiten wie Krebs verursachen können. Mit den aus dieser Arbeit gewonnenen Daten können Vorhersagen getroffen werden, wie sich verschiedene Mutationen in den Promoterregionen hinsichtlich der Proteinbindung und somit auch auf die Expression auswirken können.

Mit der dynamischen Detektion ist es möglich, Temperaturprofile der Protein-DNA-Bindung zu untersuchen. Es ist ebenso möglich, nicht die Temperatur sondern den Fluss zu variieren. Somit könnten in Zukunft Dissoziationskonstanten mit der Microarray-Technologie bestimmt werden. Durch die Kooperation mit der febit biomed GmbH sollte es bald möglich sein, markierte Oligonukleotide während der Nukleotid-Synthese einzubauen. Werden diese direkt nach dem Spacer und am Ende der Nukleotide eingefügt, kann durch das Intensitätsverhältnis die Syntheseausbeute bestimmt werden. Da die Konzentration der Nukleotide auf der Oberfläche zu Beginn der Synthese bekannt ist, kann somit auch die Konzentration der einzelnen Oligonukleotide abgeschätzt werden. Verwendet man für die Markierung der Proteine den gleichen Farbstoff, kann auch diese Konzentration anhand dieser Kontrollen bestimmt werden. Somit sollte es in naher Zukunft möglich sein, mit Hilfe der dynamischen Detektion Dissoziationskonstanten mit Hilfe der Microarray-Technologie näherungsweise zu bestimmen.

## 6 Literaturverzeichnis

- Baekstroem, S., Wolf-Watz, M., Grundstroem, C., Haerd, T., Grundstroem, T., and Sauer, U. H. (2002). The runx1 runt domain at 1.25a resolution: a structural switch and specifically bound chloride ions modulate dna binding. *J Mol Biol*, 322(2):259–272.
- Beckmann, H. and Kadesch, T. (1991). The leucine zipper of tfe3 dictates helix-loop-helix dimerization specificity. *Genes Dev*, 5(6):1057–1066.
- Beier, M. and Hoheisel, J. D. (2000). Production by quantitative photolithographic synthesis of individually quality checked dna microarrays. *Nucleic Acids Res*, 28(4):E11.
- Beier, M. and Hoheisel, J. D. (2005). Dna microarray preparation by light-controlled in situ synthesis. *Curr Protoc Nucleic Acid Chem*, Chapter 12:Unit 12.5.
- Berardi, M. J., Sun, C., Zehr, M., Abildgaard, F., Peng, J., Speck, N. A., and Bushweller, J. H. (1999). The ig fold of the core binding factor alpha runt domain is a member of a family of structurally and functionally related ig-fold dna-binding domains. *Structure*, 7(10):1247–1256.
- Boccaccio, G. and Quesada-Allue, L. (1989). Interference of sodium dodecyl sulfate in the bradford assay for protein quantitation. *An. Asoc. Quim.Argent.*, 77:79–88.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72:248–254.
- Brandt, O., Feldner, J., Stephan, A., Schroeder, M., Schnoelzer, M., Arlinghaus, H. F., Hoheisel, J. D., and Jacob, A. (2003). Pna microarrays for hybridisation of unlabelled dna samples. *Nucleic Acids Res*, 31(19):e119.
- Bulyk, M. L. (2006). Analysis of sequence specificities of dna-binding proteins with protein binding microarrays. *Methods Enzymol*, 410:279–299.
- Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-dna complex: understanding tumorigenic mutations. *Science*, 265(5170):346–355.
- Clore, G. M., Omichinski, J. G., Sakaguchi, K., Zambrano, N., Sakamoto, H., Appella, E., and Gronenborn, A. M. (1994). High-resolution structure of the oligomerization domain of p53 by multidimensional nmr. *Science*, 265(5170):386–391.
- Consortium, S. G., Consortium, C. S. G., and Consortium, N. S. G. (2008). Protein production and purification. *Nat Methods*, 5(2):135–146.
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). Weblogo: a sequence logo generator. *Genome Res*, 14(6):1188–1190.

- Diehl, F., Grahlmann, S., Beier, M., and Hoheisel, J. D. (2001). Manufacturing dna microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res*, 29(7):E38.
- Ellington, A. D. and Szostak, J. W. (1990). In vitro selection of rna molecules that bind specific ligands. *Nature*, 346(6287):818–822.
- Foord, O. S., Bhattacharya, P., Reich, Z., and Rotter, V. (1991). A dna binding domain is contained in the c-terminus of wild type p53 protein. *Nucleic Acids Res*, 19(19):5191–5198.
- Friedman, P. N., Chen, X., Bargonetti, J., and Prives, C. (1993). The p53 protein is an unusually shaped tetramer that binds directly to dna. *Proc Natl Acad Sci U S A*, 90(8):3319–3323.
- Garbassi, F., Morra, M., and Occhiello, E. (1998). *From Physics to Technology*. Polymer Surfaces.
- Garner, M. M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific dna regions: application to components of the escherichia coli lactose operon regulatory system. *Nucleic Acids Res*, 9(13):3047–3060.
- Hai, T. W., Liu, F., Coukos, W. J., and Green, M. R. (1989). Transcription factor atf cdna clones: an extensive family of leucine zipper proteins able to selectively form dna-binding heterodimers. *Genes Dev*, 3(12B):2083–2090.
- Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, 124(1):47–59.
- Hallikas, O. and Taipale, J. (2006). High-throughput assay for determining specificity and affinity of protein-dna binding interactions. *Nat Protoc*, 1(1):215–222.
- Halperin, A., Buhot, A., and Zhulina, E. B. (2006). Hybridization at a surface: the role of spacers in dna microarrays. *Langmuir*, 22(26):11290–11304.
- Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1):57–70.
- Henthorn, P. S., Stewart, C. C., Kadesch, T., and Puck, J. M. (1991). The gene encoding human tfe3, a transcription factor that binds the immunoglobulin heavy-chain enhancer, maps to xp11.22. *Genomics*, 11(2):374–378.
- Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem*, 63(24):1193A–1203A.
- Hoheisel, J. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet*, 7(3):200–210.
- Hook, F. (1997). *Development of a novel QCM technique for protein adsorption studies*. PhD thesis, Chalmers and Gothenburg University.
- Howe, J. R., Mitros, F. A., and Summers, R. W. (1998). The risk of gastrointestinal carcinoma in familial juvenile polyposis. *Ann Surg Oncol*, 5(8):751–756.
- Jacob, A., Brandt, O., Stephan, A., and Hoheisel, J. D. (2004). Peptide nucleic acid microarrays. *Methods Mol Biol*, 283:283–293.

- Jacobs, J. W. and Fodor, S. P. (1994). Combinatorial chemistry—applications of light-directed chemical synthesis. *Trends Biotechnol*, 12(1):19–26.
- Jeffrey, P. D., Gorina, S., and Pavletich, N. P. (1995). Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science*, 267(5203):1498–1502.
- Joensson, U., Faegerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Loefas, S., Persson, B., Roos, H., and Roennberg, I. (1991). Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Biotechniques*, 11(5):620–627.
- Johnsson, B., Loefas, S., and Lindquist, G. (1991). Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal Biochem*, 198(2):268–277.
- Kaerkkäinen, J.; Sanders, P. (2003). *Simple Linear Work Suffix Array Construction*. Proceedings of the 13th International Conference on Automata, Languages and Programming.
- Kapust, R. B. and Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 8(8):1668–1674.
- Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, 60(20):2299–2301.
- Knueppel, R., Dietze, P., Lehnberg, W., Frech, K., and Wingender, E. (1994). Transfac retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol*, 1(3):191–198.
- Kretschmann, E. (1971). The determination of the optical constants of metals by excitation of surface plasmons. *Z Physics*, 241:313–324.
- Kusnezow, W. and Hoheisel, J. D. (2002). Antibody microarrays: promises and problems. *Biotechniques*, Suppl:14–23.
- Laity, J. H., Lee, B. M., and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol*, 11(1):39–46.
- Lecourtois, M. and Schweisguth, F. (1995). The neurogenic suppressor of hairless dna-binding protein mediates the transcriptional activation of the enhancer of split complex genes triggered by notch signaling. *Genes Dev*, 9(21):2598–2608.
- Lieb, J. D., Liu, X., Botstein, D., and Brown, P. O. (2001). Promoter-specific binding of rap1 revealed by genome-wide maps of protein-dna association. *Nat Genet*, 28(4):327–334.
- Lu, Q., Knoepfler, P. S., Scheele, J., Wright, D. D., and Kamps, M. P. (1995). Both pbx1 and e2a-pbx1 bind the dna motif atcaatcaa cooperatively with the products of multiple murine hox genes, some of which are themselves oncogenes. *Mol Cell Biol*, 15(7):3786–3795.
- Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. (2006). Transfac and its module transcompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, 34(Database issue):D108–D110.

- Mukherjee, S., Berger, M. F., Jona, G., Wang, X. S., Muzzey, D., Snyder, M., Young, R. A., and Bulyk, M. L. (2004). Rapid analysis of the dna-binding specificities of transcription factors with dna microarrays. *Nat Genet*, 36(12):1331–1339.
- Nagata, T., Gupta, V., Sorce, D., Kim, W. Y., Sali, A., Chait, B. T., Shigesada, K., Ito, Y., and Werner, M. H. (1999). Immunoglobulin motif dna recognition and heterodimerization of the pepb2/cbf runt domain. *Nat Struct Biol*, 6(7):615–619.
- Nomura, T. and Hattori, O. (1980). Determination of micromolecular concentrations of cyanide in solution with a piezoelectric detector. *Analytica chimica acta*, 115:323.
- Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P., and Fodor, S. P. (1994). Light-generated oligonucleotide arrays for rapid dna sequence analysis. *Proc Natl Acad Sci U S A*, 91(11):5022–5026.
- Polevoda, B. and Sherman, F. (2007). Methylation of proteins involved in translation. *Mol Microbiol*, 65(3):590–606.
- Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A. (2000). Genome-wide location and function of dna binding proteins. *Science*, 290(5500):2306–2309.
- Riobo, N. A., Lu, K., and Emerson, C. P. (2006). Hedgehog signal transduction: signal integration and cross talk in development and cancer. *Cell Cycle*, 5(15):1612–1615.
- Rodahl, M. (1995). *On the frequency and Q factor response of the quartz crystal microbalance to liquid overlays*. PhD thesis, Chalmers.
- Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J. G., Mermod, N., and Bucher, P. (2002). High-throughput selex sage method for quantitative modeling of transcription-factor binding sites. *Nat Biotechnol*, 20(8):831–835.
- Sakai-Kato, K., Ishiguro, A., Mikoshiba, K., Aruga, J., and Utsunomiya-Tate, N. (2008). Cd spectra show the relational style between zic-, gli-, glis-zinc finger protein and dna. *Biochim Biophys Acta*, 1784(7-8):1011–1019.
- Sauerbrey, G. (1959). Use of vibrating quartz for thin films weighing and microweighing. *Z. Phys.*, 155:206.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science*, 270(5235):467–470.
- Schneider, T. D. (2002). Consensus sequence zen. *Appl Bioinformatics*, 1(3):111–119.
- Senear, D. F., Brenowitz, M., Shea, M. A., and Ackers, G. K. (1986). Energetics of cooperative protein-dna interactions: comparison between quantitative deoxyribonuclease footprint titration and filter binding. *Biochemistry*, 25(23):7344–7354.
- Sinha, A. U., Phatak, M., Bhatnagar, R., and Jegga, A. G. (2007). *Identifying Functional Binding Motifs of Tumor Protein p53 Using Support Vector Machines*. IEEE Computer Society.

- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Anal Biochem*, 150(1):76–85.
- Southern, E. M. (1975). Detection of specific sequences among dna fragments separated by gel electrophoresis. *J Mol Biol*, 98(3):503–517.
- Stormo, G. D. (2000). Dna binding sites: representation and discovery. *Bioinformatics*, 16(1):16–23.
- Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2001). Structural analyses of dna recognition by the aml1/runx-1 runt domain and its allosteric control by cbfbeta. *Cell*, 104(5):755–767.
- Taipale, J. and Beachy, P. A. (2001). The hedgehog and wnt signalling pathways in cancer. *Nature*, 411(6835):349–354.
- Tal, M., Silberstein, A., and Nusser, E. (1985). Why does coomassie brilliant blue r interact differently with different proteins? a partial answer. *J Biol Chem*, 260(18):9976–9980.
- Tanaka, K., Waki, H., Ido, Y., Akit, S., Yoshida, Y., and Yoshida, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 20:151–153.
- Tang, Y. Y., Crute, B. E., Kelley, J. J., Huang, X., Yan, J., Shi, J., Hartman, K. L., Laue, T. M., Speck, N. A., and Bushweller, J. H. (2000). Biophysical characterization of interactions between the core binding factor alpha and beta subunits and dna. *FEBS Lett*, 470(2):167–172.
- Tompa, R., McCallum, C. M., Delrow, J., Henikoff, J. G., van Steensel, B., and Henikoff, S. (2002). Genome-wide profiling of dna methylation reveals transposon targets of chromomethylase3. *Curr Biol*, 12(1):65–68.
- van Steensel, B., Delrow, J., and Henikoff, S. (2001). Chromatin profiling using targeted dna adenine methyltransferase. *Nat Genet*, 27(3):304–308.
- Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N., and Ansari, A. Z. (2006). Defining the sequence-recognition profile of dna-binding molecules. *Proc Natl Acad Sci U S A*, 103(4):867–872.
- Wasserman, W. W. and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*, 5(4):276–287.
- Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738.
- Wiechelmann, K. J., Braun, R. D., and Fitzpatrick, J. D. (1988). Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. *Anal Biochem*, 175(1):231–237.
- Wilson, W. J., Strout, C. L., DeSantis, T. Z., Stilwell, J. L., Carrano, A. V., and Andersen, G. L. (2002). Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes*, 16(2):119–127.

- Workman, C. T., Yin, Y., Corcoran, D. L., Ideker, T., Stormo, G. D., and Benos, P. V. (2005). enologos: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res*, 33(Web Server issue):W389–W392.
- Yang, W. P., Wu, H., and Barbas, C. F. (1995). Surface plasmon resonance based kinetic studies of zinc finger-dna interactions. *J Immunol Methods*, 183(1):175–182.

## 7 Abkürzungverzeichnis

A	Adenin
a	antisense
AK	Antikörper
AMP	Adenosinmonophosphat
APS	Ammoniumpersulfat
AS	Aminosäure
ATP	Adenosintriphosphat
A.U.	Arbitrary Units
BCA	bicinchoninic acid
BCIP / NBT	5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt / Nitro-Blue Tetrazolium Chloride
bp	Basenpaar(e)
BSA	Bovine Serum Albumin, (Rinderserumalbumin)
C	Cytosin
cm	centimeter
Da	Dalton
dATP	Desoxyadenosin-5'-triphosphat
ddH <sub>2</sub> O	doppelt destilliertes Wasser
DMF	2,5-Dimethylformamid
DMSO	Dimethylsulfoxid
DNS	deoxyribonucleic acid
DNA	Desoxyribonukleinsäure
dNTP	Deoxynukleosid-Triphosphat
DOPC	1,2-dioleoyl-sn-glycero-phosphocholine
DOPE	Dioleoyl-phosphatidylethanolamin
DOPS	Dioleoylphosphatidylserin
dsDNS	Doppelsträngige DNS
DTT	Dithiothreitol
E. coli	Escherichia coli
ECL	Enzymatische Chemolumineszenz
EDTA	Ethylendiamintetraessigsäure
EMCS	6-maleimidohexansäure-N- hydroxysuccinimid-ester

FPLC	Fast protein liquid chromatography
g	Erdbeschleunigung (9,81 m/s <sup>2</sup> )
g	Gramm
G	Guanin
GFP	Green fluorescent protein
h	Stunde
H <sub>2</sub> O	Wasser
His	Histidin
HLH	Helix-Loop-Helix
HTH	Helix-turn-Helix
hs	homo sapiens
IgG	Immunglobulin G
IPTG	Isopropyl-1-thio-β-D-galactosid
k	kilo
kb	Kilobasen
kDa	kiloDalton
LB-Medium	Luria Broth Medium
M	Mol
MBP	Maltose Binding Protein
ME	Mercaptoethanol
mg	Milligramm
MG	Monoglyceride
min	Minute
ml	Milliliter
mm	Millimeter
mM	millimolar
mRNA	messenger RNA
ms	Milli Sekunde
MSC	Multiple Cloning Site
Myc	myelocytomatosis viral oncogene homolog
µg	Mikrogramm (10 <sup>-6</sup> g)
µl	Mikroliter (10 <sup>-6</sup> l)
µM	mikromolar (10 <sup>-6</sup> M)
NaCl	Natriumchlorid
NaOH	Natronlauge
ng	Nanogramm (10 <sup>-9</sup> g)
nm	Nanometer (10 <sup>-9</sup> m)
NPPOC	3'-Nitrophenylpropyloxycarbonyl)
NTA	Nitrilotriessigsäure
OD	Optische Dichte
PAGE	Polyacrylamid-Gelelektrophorese
PBS	Phosphate buffered saline

---

PCR	Polymerase chain reaction
Pen	Penicillin
Pfu-Polymerase	Pyrococcus furiosus
pmol	Picomol ( $10^{-12}$ mol)
QCMD	Quartz Crystal Microbalance with Dissipation
Rb	rabbit
RFP	Red fluorescent protein
RNA	Riboxynukleinsäure
rpm	revolutions per minute (Umdrehungen pro Minute)
RU	Response Units
RT	Raumtemperatur ( etwa 22°C)
s	sense
SAPE	Streptavidin R-Phycoerythrin
SDS	sodium dodecyl sulfate
sec	Sekunde
SELEX	systematic evolution of ligands by exponential enrichment
SPR	Surface plasmon resonance
T	Thymin
SNP	single nucleotide polymorphism
TAE-Puffer	Tris-Acetat-EDTA-Puffer
Taq-Polymerase	Thermus aquaticus
TBS	Tris Buffered Saline
TEMED	N,N,N',N'-Tetramethyl-ethylendiamin
TF	Transkriptionsfaktor
Tris	Tris(hydroxymethyl)aminomethan
U	Unit (Menge an Enzym, die 1µg DNS schneidet)
ÜNK	Übernachtkultur
UV/Vis	Ultra-Violettes Licht / für den Menschen sichtbares Licht
X-Gal	5-Brom-4-Chlor-3-indolyl- $\beta$ -D-galactosid

## 8 Abbildungsverzeichnis

1.1	Die 20 häufigsten Krebstodesursachen in Deutschland im Jahr 2006. . . . .	1
1.2	Trendgraphiken zu den fünf häufigsten Todesursachengruppen in Deutschland für Frauen (links) und Männer (rechts). (Quelle: www.dkfz.de/krebsatlas) . .	2
1.3	Partner-Institute des “Regulatory Genomics” EU Projektes . . . . .	3
2.1	Die vier Basen der DNA. . . . .	11
2.2	Zuckerphosphatrückgrat der DNA mit beliebigen Nukleobasen (B). . . . .	12
2.3	Basenpaarung der Basen und DNA-Helix. . . . .	12
2.4	Uracil (links) und Ribose (rechts). . . . .	13
2.5	Schematische Darstellung der räumlichen Proteinstrukturen. . . . .	15
2.6	Bindung von Helix-Turn-Helix-Proteinen (links) und homöotischen Proteine (rechts) an eine DNA-Bindungsstelle in der großen Furche der DNA. . . . .	16
2.7	Bindung eines Zinkfingers (links) und eines Zinkfingerproteins mit mehreren Zinkfingern (rechts) an ihre Bindungssequenz(en) in der DNA. . . . .	17
2.8	Bindung eines Leucin-Zipper-Proteins und eines Helix-Loop-Helix-Proteins an ihre Bindungssequenzen in der DNA. . . . .	18
2.9	Aufreinigung der Transkriptionsfaktoren mit Hilfe eines Maltose-bindenden-Proteins. . . . .	20
2.10	Struktur des Nickel-Chelat-Komplex. R = Protein. . . . .	21
2.11	Fluoreszenzspektrum von R-Phycoerythrin. . . . .	22
2.12	Schema der der Ionisation und Desorption bei der MALDI-TOF Analyse . . .	23
2.13	Typischer DNA-Microarray (links). Lichtaktivierte DNA-Synthese mit Hilfe eines Mikrospiegelsystems (rechts). . . . .	24
2.14	DNA-Synthese I. . . . .	25
2.15	DNA-Synthese II . . . . .	25
2.16	DNA-Synthese III . . . . .	26
2.17	Sensorgramm einer Biacore-Messung . . . . .	27
3.1	Coomassie Brilliant Blue G-250 . . . . .	51
3.2	Struktur eines Kupfer-Bicinchoninsäure- Komplexes . . . . .	52
3.3	UV-Absorption von Tryptophan und Thyrosin. . . . .	53
4.1	DNA-Microarray . . . . .	57
4.2	Optimierung der Proteinexpression. . . . .	61
4.3	Aufreinigung von Gli2 als Fusionsprotein mit MBP im Vektor pMAL c2X. . .	62
4.4	Faktor Xa Verdau . . . . .	62

4.5	MALDI Spektrum von MBP-3xMyc-Gli2 . . . . .	63
4.6	Aufreinigung von Gli2 als Fusionsprotein mit 8x His-Tag im Vektor pQE Tri- system Strep 1 . . . . .	65
4.7	Proteinaufreinigung von Gli2 mit His-Tag unter verschiedenen Waschkonditionen.	66
4.8	Proteinaufreinigung von Gli2 mit His-Tag über FPLC. . . . .	67
4.9	Typisches Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA- Sequenzen auf einem mit dem Geniom 1X synthetisierten Chip. . . . .	68
4.10	Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA-Sequenzen.	69
4.11	Ergebnis der Bindung von Gli2 mit His-Tag an verschiedene DNA-Sequenzen	69
4.12	PCR auf Chips . . . . .	70
4.13	Ergebnis der Bindung von Gli2 mit MBP-Tag an verschiedene DNA-Sequenzen einer Haarnadelstruktur und an eine mit PCR hergestellte Nukleotidsequenz. Die sichtbaren Spots zeigen die Haarnadelstruktur. An den mit PCR hergestellten Sequenzen konnte keine Proteinbindung detektiert werden. . . . .	71
4.14	Abhängigkeit der Signalintensität von der aufgetragenen Proteinmenge im Fall von Gli2-His. . . . .	72
4.15	Abhängigkeit der Signalintensität von der aufgetragenen Proteinmenge im Fall von Gli2-MBP. . . . .	72
4.16	Silanisierung von Silizium- oder Glasoberflächen mit 3-Aminopropyltriethoxysilan (APTES) . . . . .	73
4.17	Die Aktivierung der silanisierten Oberflächen mit 6-Maleinimidohexansäure- N-succinimidylester (EMCS, blau) erleubt die Kopplung der DNA über die Thiolgruppe (grün). . . . .	74
4.18	Microarray auf Glas-Objektträger zur Optimierung der DNA-Konzentration. .	75
4.19	Kontrolle zur Effizienz der Ausbildung der Haarnadelstruktur. . . . .	75
4.20	Arraydesign des Geniom Chips für das Fusionsprotein Gli2. . . . .	76
4.21	Vergleich zwischen His-Tag und MBP-Tag. . . . .	77
4.22	Vergleich der Bindungsstärke von Gli2 an alle möglichen 1-Basenpaar-Permutationen der Konsensus-Sequenz auf DNA-Chips mit verschiedenen DNA-Konzentrationen.	78
4.23	Vergleich der Bindungsstärke von Gli2 His- und MBP-Tag . . . . .	79
4.24	Vergleich <i>flanking region</i> . . . . .	80
4.25	Vergleich der Bindungsstärke von Gli2 an alle möglichen 1-Basenpaar-Permutationen der Konsensus-Sequenz mit verschiedenen endständigen Nukleotiden. In Groß- buchstaben ist die permutierte Base angegeben. . . . .	80
4.26	Vergleich Spacer . . . . .	81
4.27	Vergleich Kompetitor . . . . .	82
4.28	Vergleich Farbstoffe . . . . .	83
4.29	Gewichtete Positionsmatrix von MBP-Gli2 . . . . .	83
4.30	Gewichtete Positionsmatrix und Sequenzlogo von MBP-Gli2 . . . . .	84
4.31	Sequenzmusteranalyse von MBP-Gli2 . . . . .	84
4.32	Vergleich der verschiedenen Experimente mit Gli2 als “ <i>color coded list</i> ” . . . .	85
4.33	<i>Scatter Plot</i> der Intensitäten von Gli2 . . . . .	86

4.34	Ausschnitte von Bildern der dynamischen Detektion. . . . .	87
4.35	Ausschnitte von Bildern der bleaching Kontrolle (links) sowie Waschkontrolle (rechts). . . . .	87
4.36	Ergebnis der dynamischen Detektion von Gli2. . . . .	88
4.37	Schema der Signalverstärkung durch die Verwendung eines biotinylierten Anti-Streptavidin-Antikörpers. . . . .	89
4.38	Signalverstärkung . . . . .	89
4.39	Intensitäten aller untersuchten 9 mere (links) inklusive der 10 stärksten Signale (rechts). . . . .	90
4.40	Idealisierte Darstellung der Bindungskinetik einer Protein-DNA -Wechselwirkung in einem BIAcore-T100-System . . . . .	92
4.41	Sensorgramme von Gli2 aufgenommen mit dem Biacore T100 . . . . .	93
4.42	QCMD Messergebnisse für die einzelnen Kristalle . . . . .	96
4.43	Werte für die Viskosität und das Schermodul . . . . .	99
4.44	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-PBX1 (links). Ausschnitt eines Arrays mit gebundenen PBX1 Transkriptionsfaktor (rechts). . . . .	103
4.45	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	103
4.46	Oben: Sequenzlogo von PBX1 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	104
4.47	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-RBPJK (links). Ausschnitt eines Arrays mit gebundenen RBPJK Transkriptionsfaktor (rechts). . . . .	105
4.48	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	105
4.49	Oben: Sequenzlogo von RBPJK mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	106
4.50	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-ATF1 (links). Ausschnitt eines Arrays mit gebundenen ATF1 Transkriptionsfaktor (rechts). . . . .	107
4.51	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	107
4.52	Oben: Sequenzlogo von ATF1 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	108

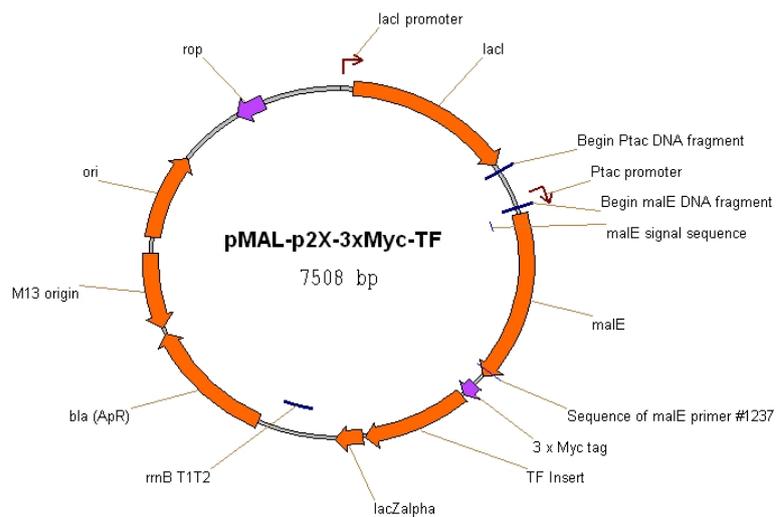
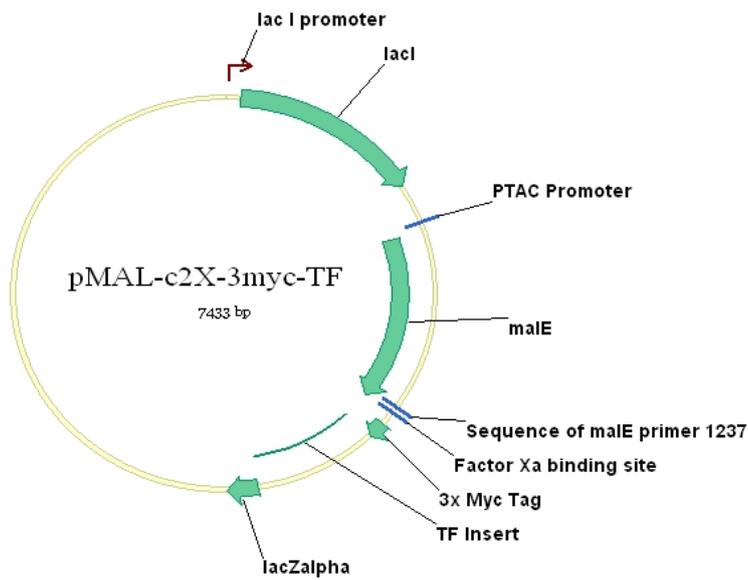
---

4.53	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-TCF4 (links). Ausschnitt eines Arrays mit gebundenen TCF4 Transkriptionsfaktor (rechts). . . . .	109
4.54	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	109
4.55	Oben: Sequenzlogo von TCF4 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	110
4.56	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-WT1 (links). Ausschnitt eines Arrays mit gebundenen WT1 Transkriptionsfaktor (rechts). . . . .	111
4.57	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	111
4.58	Sequenzlogo von WT1 mit 5% der stärksten Signale (links) und das daraus resultierende Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). . . . .	112
4.59	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteins MBP-3xMyc-CBFB (links). Ausschnitt eines Arrays mit gebundenen CBFB Transkriptionsfaktor (rechts). . . . .	113
4.60	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	113
4.61	Oben: Sequenzlogo von CBFB mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	114
4.62	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-SMAD4 (links). Ausschnitt eines Arrays mit gebundenen SMAD4 Transkriptionsfaktor (rechts). . . . .	115
4.63	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	115
4.64	Oben: Sequenzlogo von SMAD4 mit 5% der stärksten Signale (links) und das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	116
4.65	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-TTE3 (links). Ausschnitt eines Arrays mit gebundenen TFE3 Transkriptionsfaktor (rechts). . . . .	117
4.66	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	117

4.67	Oben: Sequenzlogo von TFE3-Ebox Motiv (links) mit 5% der stärksten Signale sowie das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). Unten: Positionsgewichtete Scorematrix (links) und das dazugehörige Sequenzlogo (rechts). . . . .	118
4.68	SDS-Gel, Proteinmarker sowie Western Blot des Fusionsproteines MBP-3xMyc-p53 (links). Ausschnitt eines Arrays mit gebundenen p53 Transkriptionsfaktor (rechts). . . . .	119
4.69	Intensitäten aller untersuchten Permutationen (links) inklusive der 10 stärksten Signale (rechts). . . . .	120
4.70	Sequenzlogo von p53-Ebox Motiv (links) mit 5% der stärksten Signale sowie das dazugehörige Sequenzlogo aus der Literatur (rechts; Quelle: Transfac). . .	120

## 9 Anhang

### Vektorkarten



## MALDI Spektren

### ATF1

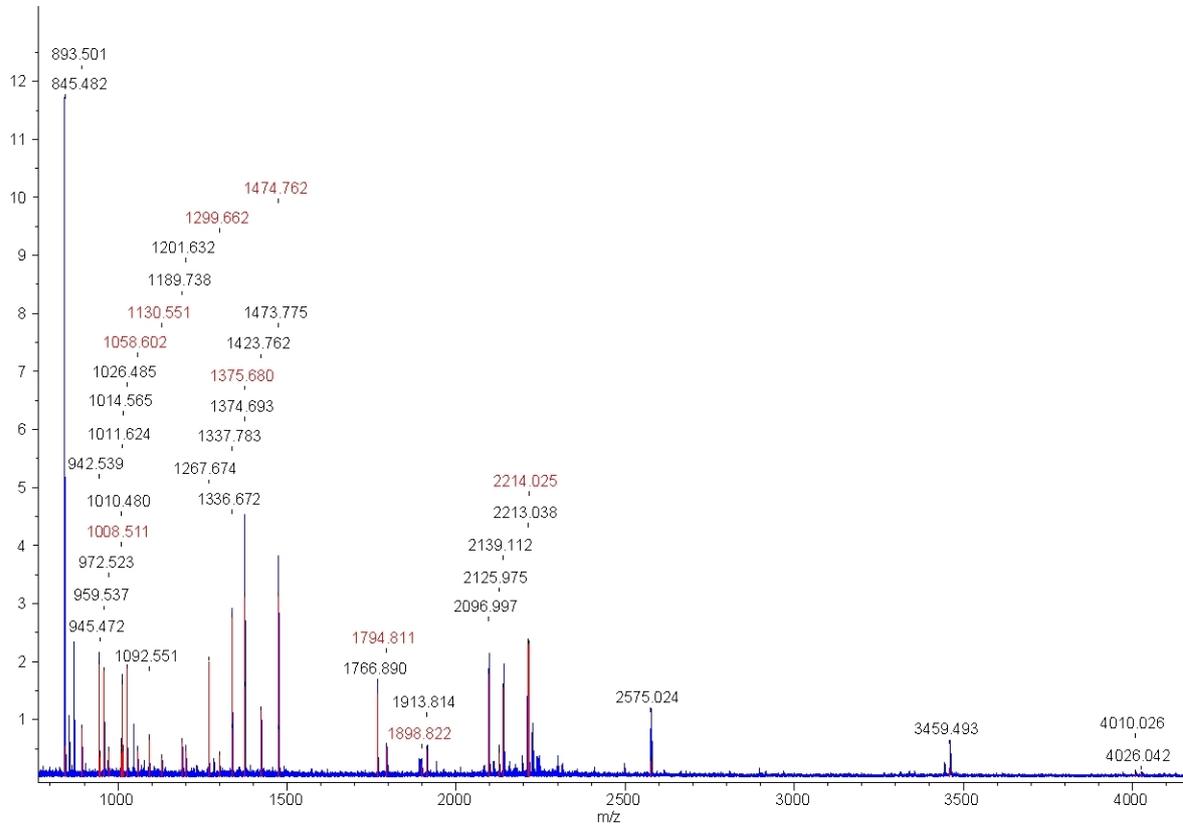
#### Sequenzierung des Vektorinserts

GAGCTCGAACAAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTCAGAATTCGGCAACATG  
GGCCGCTCTGAGCAAAAGCTCATTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCA  
ACGGTGAGCAGAAGTTAATATCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGCCGCACT  
GGTACCCCGTGAAGTCTGGAAGTTCTCTTCCAGGGGCCGAAGGAGACGCCACCATGGCTGCTACTTCTTGCCA  
CAAAGTGTGGTGATGACATCTCCTGTGACTCTCACCTCTCAGACAACCTAAGACAGATGACCCCAATTGAAAAGAG  
AAATAAGTTAATGAAAAACAGAGAAGCTGCTCGAGAATGTCGCAGAAAGAAGAAAGAATATGTGAAATGCCTGGA  
AAACCGAGTTGCAGTCCTGGAAAATCAAAATAAACTCTAATAGAAGAGTTAAAACTTTGAAGGATCTTTATTCC  
AATAAAAGTGTGGCGGCCGCACTAGAATTGTGAGCGCTCACAATTCTGAGAATTTGTATTTTCAGGGTGATCTCC  
GTGGATCTATTACAACCTCGAGTCTAGGGATCTCTAGA

#### Proteinsequenz der Bindeseite des Fusionsproteins

S S N N N N N N N N N L G I E G R I S E F G N M G R S E Q K L I S E E D L  
N G E Q K L I S E E D L N G E Q K L I S E E D L N G S A G I R S G T A A A L  
V P R G S L E V L F Q G P E G D A T M A A T S L P Q T V V M T S P V T L T S  
Q T T K T D D P Q L K R E I R L M K N R E A A R E C R R K K K E Y V K C L E  
N R V A V L E N Q N K T L I E E L K T L K D L Y S N K S V G G R T R I V S A  
H N S E N L Y F Q G D L R G S I T T R V

## MALDI-Spektrum



ATF1 MALDI

M/z	GEFUNDENE FRAGMENTE
715.37	NREAAR
731.40	SVGGRTR
732.41	GSITTRV
738.35	DLYSNK
788.46	EIRLMK
804.45	EIRLMK
815.40	TDDPQLK
844.47	TLIEELK
890.40	EAARECR
941.53	SGTAAALVPR
971.51	TDDPQLKR
1009.47	ISEFGNMGR
1013.56	VAVLENQNK
1025.48	ISEFGNMGR
1080.58	TLKDLYSNK
1186.72	TLIEELKTLK
1194.60	DLYSNKSVGGR

1209.58	EYVKCLENR
1373.69	LISEEDLNGEQK
1373.69	LISEEDLNGEQK
1472.77	LISEEDLNGSAGIR
1481.69	ISEFGNMGRSEQK
1497.69	ISEFGNMGRSEQK
1685.85	CLENRVAVLENQNK
1840.03	VAVLENQNKTLIEELK
1845.90	SEQKLISEEDLNGEQK
1957.86	SSNNNNNNNNNNNLGIEGR
1961.96	IVSAHNSENLYFQGDLR
2219.11	TRIVSAHNSENLYFQGDLR
2396.27	LISEEDLNGSAGIRSGTAAALVPR
2676.36	IVSAHNSENLYFQGDLRGSITTRV
2729.32	LISEEDLNGEQKLISEEDLNGEQK
2828.40	LISEEDLNGEQKLISEEDLNGSAGIR
2949.31	SSNNNNNNNNNNNLGIEGRISEFGNMGR
2965.31	SSNNNNNNNNNNNLGIEGRISEFGNMGR
3993.00	GSLEVLFGQPEGDATMAATSLPQTVVM TSPVTLTSQTTK
4009.02	GSLEVLFGQPEGDATMAATSLPQTVVM TSPVTLTSQTTK
4009.02	GSLEVLFGQPEGDATMAATSLPQTVVM TSPVTLTSQTTK
4025.03	GSLEVLFGQPEGDATMAATSLPQTVVM TSPVTLTSQTTK

**Tabelle 9.1:** Im Massenspektrometer gefundene Teilfragmente von ATF1

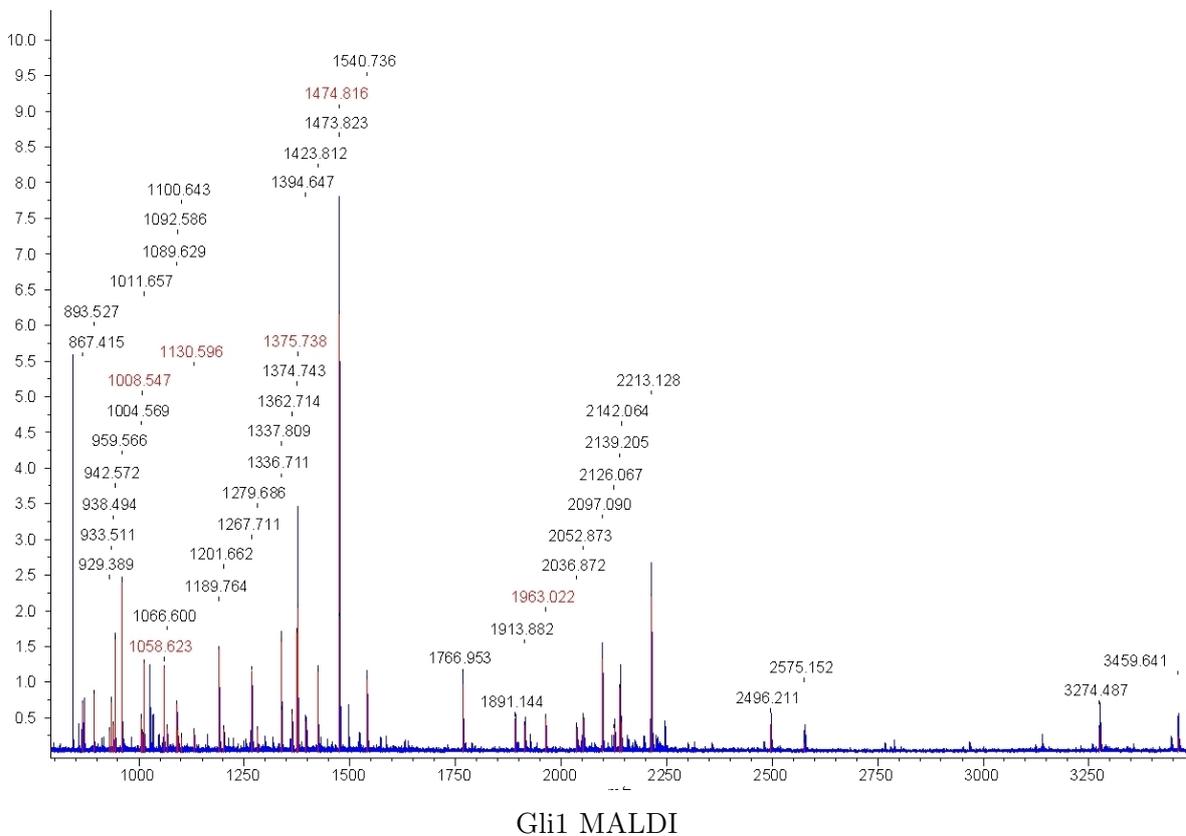
## Gli1

### Sequenzierung des Vektorinserts

GCCGCCAGCGGTCGTCAGACTGTTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACA  
ATAACAATAACAACAACCTCGGGATCGAGGGAAGATTTTCAGAATTCGGCAACATGGGCCGCTCTGAGCAAAAAGCT  
CATTTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCAACGGTGAGCAGAAAGTTAATA  
TCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGGCCGCACTGGTACCCCGTGGAAAGTCTGG  
AAGTTCTCTCCAGGGGCCGAAGGAGACGCCACCATGGCTTTGGGGATGCTGGATGGGCGGGAGGACCTCGAGAG  
AGAGGAGAAGCGTGAGCCTGAATCTGTGTATGAAACTGACTGCCGTTGGGATGGCTGCAGCCAGGAATTTGACTCC  
CAAGAGCAGCTGGTGCACCACATCAACAGCGAGCACATCCACGGGAGCGGAAGGAGTTCGTGTGCCACTGGGGG  
GCTGCTCCAGGAGCTGAGGCCCTTCAAAGCCCAGTACATGCTGGTGGTTCACATGCGCAGACACACTGGCGAGAA  
GCCACACAAGTGCACGTTTGAAGGGTCCGGAAGTCACTACTCACGCCTCGAAAACCTGAAGACGCACCTGCGGTCA  
CACACGGGTGAGAAGCCATACATGTGTGAGCACGAGGGCTGCAGTAAAGCCTTCAGCAATGCCAGTGACCGAGCCA  
AGCACCAGAAATCGGACCCATTCCAATGAGAAGCCGTATGTATGTAAGCTCCCTGGCTGCACCAAACGCTATACAGA  
TCCTAGCTCGCTGCGAAAACATGTCAAGACAGTGCATGGTCCTGACGCCCATGTGACCAAACGGCACCCTGGGGAT  
GGCCCCCTGCCTCGGGCACCAGGCGGCCGCACTAGAAATTGTGAGCGCTCACAATTCTGAGAAATTTGATTTTCAGG  
GTGATCTCCGTGGATCTATTACAACCTCGAGTCTAGGGATCCTCTAGAGTCGACCTGCAGGCAAGCTT

## Proteinsequenz der Bindeseite des Fusionsproteins

A A S G R Q T V D E A L K D A Q T N S S S N N N N N N N N N N L G I E G R I S  
 E F G N M G R S E Q K L I S E E D L N G E Q K L I S E E D L N G E Q K L I S E  
 E D L N G S A G I R S G T A A A L V P R G S L E V L F Q G P E G D A T M A L G  
 M L D G R E D L E R E E K R E P E S V Y E T D C R W D G C S Q E F D S Q E Q L  
 V H H I N S E H I H G E R K E F V C H W G G C S R E L R P F K A Q Y M L V V H  
 M R R H T G E K P H K C T F E G C R K S Y S R L E N L K T H L R S H T G E K P  
 Y M C E H E G C S K A F S N A S D R A K H Q N R T H S N E K P Y V C K L P G C  
 T K R Y T D P S S L R K H V K T V H G P D A H V T K R H R G D G P L P R A P G  
 G R T R I V S A H N S E N L Y F Q G D L R G S I T T R V



M/z	GEFUNDENE FRAGMENTE
892.51994	DLLPNPPK
958.558373	TAVINAASGR
1010.649626	DKPLGAVALK
1091.57832	IAATMENAQK
1099.635609	TDRPSQQLR
1188.756361	AGLTFLVDLIK
1200.655054	TWEEIPALDK
1266.703316	LYPFTWDAVR
1335.703798	SYEEELAKDPR
1336.801384	VNYGVTVLPTFK

1422.804535	VTVEHPDKLEEK
1765.945321	FGGYAQSGLLAEITPDK
1890.137152	LIAYPIAVEALS LIYNK
1912.874471	HMNADTDYSIAEAAFNK
2096.082969	EFLENYLLTDEGLEAVNK
2125.059697	GEIMPNIPQMSAFWYAVR
2138.197964	GQPSKPFVGVLSAGINAASPNK
2141.056386	GEIMPNIPQMSAFWYAVR
2212.120586	FPQVAATGDGPDIIFWAHDR
2574.144344	DAQTNSSSNNNNNNNNNNLGIEGR
3458.633565	QTVDEALKDAQTNSSSNNNNNNNNNNLGIEGR

**Tabelle 9.2:** Im Massenspektrometer gefundene Teilfragmente von GLI1

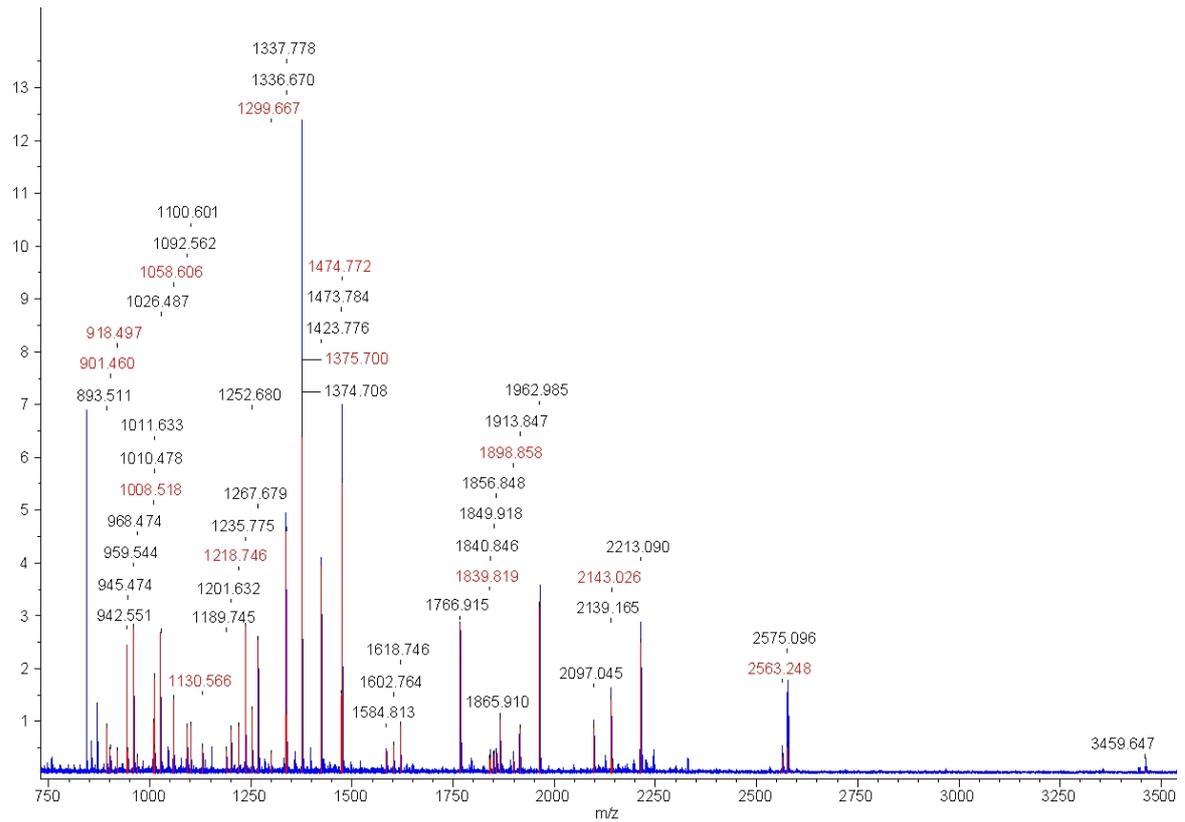
## RBP

### Sequenzierung des Vektorinserts

AACAACAACCTCGGGATCGAGGGAAGATTTTCAGAATTCGGCAACATGGGCCGCTCTGAGCAAAGCTCATTCTG  
AAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCAACGGTGAGCAGAAGTTAATATCCGAGGA  
GGATCTTAATGGCTCTGCAGGAATTAGGTCTGGGTACCGCGGCCGCACTGGTACCCCGTGGAAAGTCTGGAAGTTCTC  
TTCCAGGGGCCGAAGGAGACGCCACCATGGCTAAGGCAGAGTATACATTTTATGAGGGAATGGGCCCTGTCCTTG  
CCCCAGTCACTCCTGTGCCTGTGGTAGAGAGCCTTCAGTTGAATGGCGGTGGGGACGTAGCAATGCTTGAACCTTAC  
AGGACAGAATTTCACTCCAAATTTACGAGTGTGGTTTGGGGATGTAGAAGCTGAAACTATGTACAGGTGTGGAGAG  
AGTATGCTCTGTGTCGTCAGACATTTCTGCATTCCGAGAAGTTGGAGATGGGTCCGGCAACCAGTCCAGGTTCC  
CAGTAACTTTGGTCCGAAATGATGGAATCATTTATTCCACCAGCCTTACCTTTACCTACACACCAGAACCAGGGCC  
GCGGCCACATTGCAGTGCAGCAGGAGCAATCCTTCGAGCCAATTCAGGCGGCCGCACTAGAAATTGTGAGCGCTCAC  
AATTCTGAGAATTTGTATTTTCAGGGTGATCTCCGTGGATCTATTACAACCTCGAGTCTAGGGATCCTCTAGAGTCG  
ACCTGCAGGCAAGCTT

### Proteinsequenz der Bindeseite des Fusionsproteins

S S N N N N N N N N N L G I E G R I S E F G N M G R S E Q K L I S E E D L  
N G E Q K L I S E E D L N G E Q K L I S E E D L N G S A G I R S G T A A A L  
V P R G S L E V L F Q G P E G D A T M A K A E Y T F Y E G M G P V L A P V T  
P V P V V E S L Q L N G G D V A M L E L T G Q N F T P N L R V W F G D V E  
A E T M Y R C G E S M L C V V P D I S A F R E G W R W V R Q P V Q V P V T L  
V R N D G I I Y S T S L T F T Y T P E P G P R P H C S A A G A I L R A N S G  
G R T R I V S A H N S E N L Y F Q G D L R G S I T T R V



RBP MALDI

M/z	GEFUNDENE FRAGMENTE
732.413003	GSITTRV
817.415462	ANSGRTR
941.543488	SGTAAALVPR
987.503883	EGWRWVR
1009.47037	ISEFGNMGR
1025.478755	ISEFGNMGR
1234.767015	QPVQVPVTLVR
1373.70039	LISEEDLNGEQK
1373.70039	LISEEDLNGEQK
1472.776176	LISEEDLNGSAGIR
1481.693277	ISEFGNMGRSEQK
1497.688191	ISEFGNMGRSEQK
1601.756374	VWFGDVEAETMYR
1617.737923	VWFGDVEAETMYR
1675.988594	WVRQPVPVTLVR
1839.837871	CGESMLCVVPDISAFR
1845.895601	SEQKLISEEDLNGEQK
1848.910295	GSLEVLFGPEGDATMAK
1855.840295	CGESMLCVVPDISAFR
1864.901933	GSLEVLFGPEGDATMAK

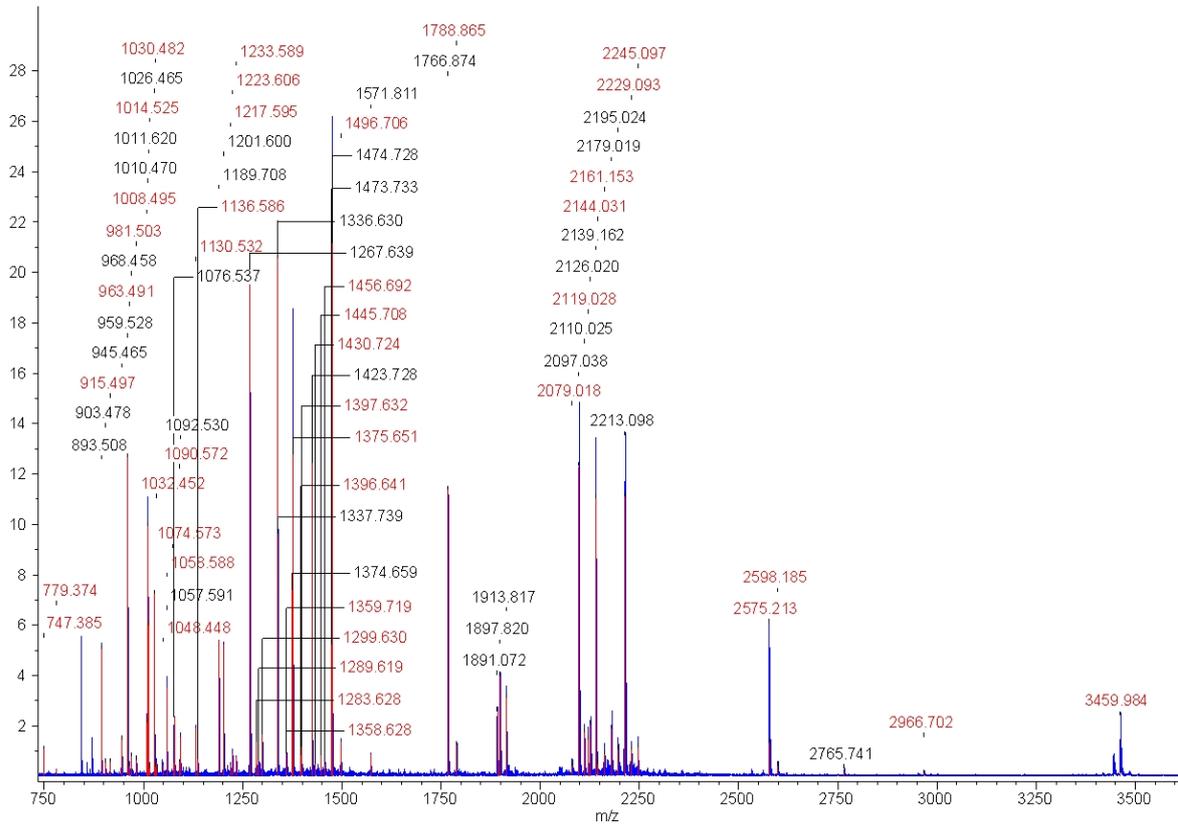
1957.858656	SSNNNNNNNNNNLGIIEGR
1961.977084	IVSAHNSENLYFQGDLR
2219.108328	TRIVSAHNSENLYFQGDLR
2368.076242	CGESMLCVVPDISAFREGWR
2384.071156	CGESMLCVVPDISAFREGWR
2396.265951	LISEEDLNLSAGIRSGTAAALVPR
2676.361977	IVSAHNSENLYFQGDLRGSITTRV
2729.324313	LISEEDLNLSAGIRSGTAAALVPR
2772.411629	SGTAAALVPRGSLEVLFGPEGDATMAK
2788.406544	SGTAAALVPRGSLEVLFGPEGDATMAK
2828.403961	LISEEDLNLSAGIRSGTAAALVPR
2949.313205	SSNNNNNNNNNNLGIIEGRRISEFGNMGR
2965.30812	SSNNNNNNNNNNLGIIEGRRISEFGNMGR
3423.539625	VWFGDVEAETMYRCGESMLCVVPDISAFR
3439.534539	VWFGDVEAETMYRCGESMLCVVPDISAFR
3439.534539	VWFGDVEAETMYRCGESMLCVVPDISAFR
3455.529454	VWFGDVEAETMYRCGESMLCVVPDISAFR
3461.703788	NDGHIYSTLTFTYTPEPGPRPHCSAAGAILR

**Tabelle 9.3:** Im Massenspektrometer gefundene Teilfragmente von RBP

## WT1

### Sequenzierung des Vektorinserts

GAGCTCGAACAAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTCAGAATTCGGCAACATG  
GGCCGCTCTGAGCAAAGCTCATTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCA  
ACGGTGAGCAGAAGTTAATATCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGCCGCACT  
GGTACCCCGTGAAGTCTGGAAGTTCTTCCAGGGGCCGAAGGAGACGCCACCATGGCTGTGCGACGTGTGCCT  
GGAGTAGCCCGACTCTTGACGGTCCGCATCTGAGACCAGTGAGAAACGCCCTTCATGTGTGCTTACCCAGGCT  
GCAATAAGAGATATTTTAAGCTGTCCCACTTACAGATGCACAGCAGGAAGCACACTGGTGAGAAACCATAACAGTG  
TGACTTCAAGGACTGTGAACGAAGTTTTCTCGTTCAGACCAGCTCAAAGACACCAAAGGAGACATACAGGTGTG  
AAACCATTCCAGTGTAACAACTTGTGACGAAAGTTCTCCCGTCCGACCACCTGAAGACCCACACCAGGACTCATA  
CAGGTGAAAAGCCCTTCAGCTGTCCGTGGCCAAGTTGTGCAAAAAGTTTTGCCGGTCAGATGAATTAGTCCGCCA  
TCACAACATGCATCAGAGAAACATGACCAAACTCCAGCTGGCGCTTGGCGGCCGCACTAGAATTGTGAGCGCTCAC  
AATTCTGAGAATTTGTATTTTCAGGGTATCTCCGTGGATCTATTACAACCTCGAGTCTAGGGATCCTCTAGA



WT1 MALDI

**Proteinsequenz der Bindeseite des Fusionsproteins**

S S N N N N N N N N N L G I E G R I S E F G N M G R S E Q K L I S E E D L N  
 G E Q K L I S E E D L N G E Q K L I S E E D L N G S A G I R S G T A A A L V P  
 R G S L E V L F Q G P E G D A T M A V R R V P G V A P T L V R S A S E T S E K  
 R P F M C A Y P G C N K R Y F K L S H L Q M H S R K H T G E K P Y Q C D F K D  
 C E R R F S R S D Q L K R H Q R R H T G V K P F Q C K T C Q R K F S R S D H L  
 K T H T R T H T G E K P F S C R W P S C Q K K F A R S D E L V R H H N M H Q R  
 N M T K L Q L A L G G R T R I V S A H N S E N L Y F Q G D L R G S I T T R V

M/z	GEFUNDENE FRAGMENTE
702.39276	KIEEGK
892.501011	DLLPNPPK
902.471044	QTVDEALK
944.458086	DVGVDNAGAK
958.521152	TAVINAASGR
967.450553	SYEEELAK
1010.61284	DKPLGAVALK
1056.584188	LVIWINGDK
1075.529581	IAATMENAQK
1091.522783	IAATMENAQK
1188.700558	AGLTFLVDLIK

1200.592852	TWEEIPALDK
1266.631224	LYPFTWDAVR
1335.622723	SYEEELAKDPR
1336.731944	VNYGVTVLPTFK
1422.720484	VTVEHPDKLEEK
1473.72023	QTVDEALKDAQTR
1570.803765	TWEEIPALDKELK
1765.867217	FGGYAQSGLLAEITPDK
1890.064625	LIAYPIAVEALS LIYNK
1896.812586	HMNADTDYSIAEAAFNK
1912.809577	HMNADTDYSIAEAAFNK
2096.030403	EFLENYLLTDEGLEAVNK
2109.017747	GEIMPNIPQMSAFWYAVR
2125.012253	GEIMPNIPQMSAFWYAVR
2138.15438	GQPSKPFVGVLSAGINAASPNK
2178.011616	GETAMTINGPWAWSNIDTSK
2194.016898	GETAMTINGPWAWSNIDTSK
2212.090341	FPQVAATGDGPDIIFWAHDR
2764.733457	LIAYPIAVEALS LIYNKDLLPNPPK

**Tabelle 9.4:** Im Massenspektrometer gefundene Teilfragmente von WT1

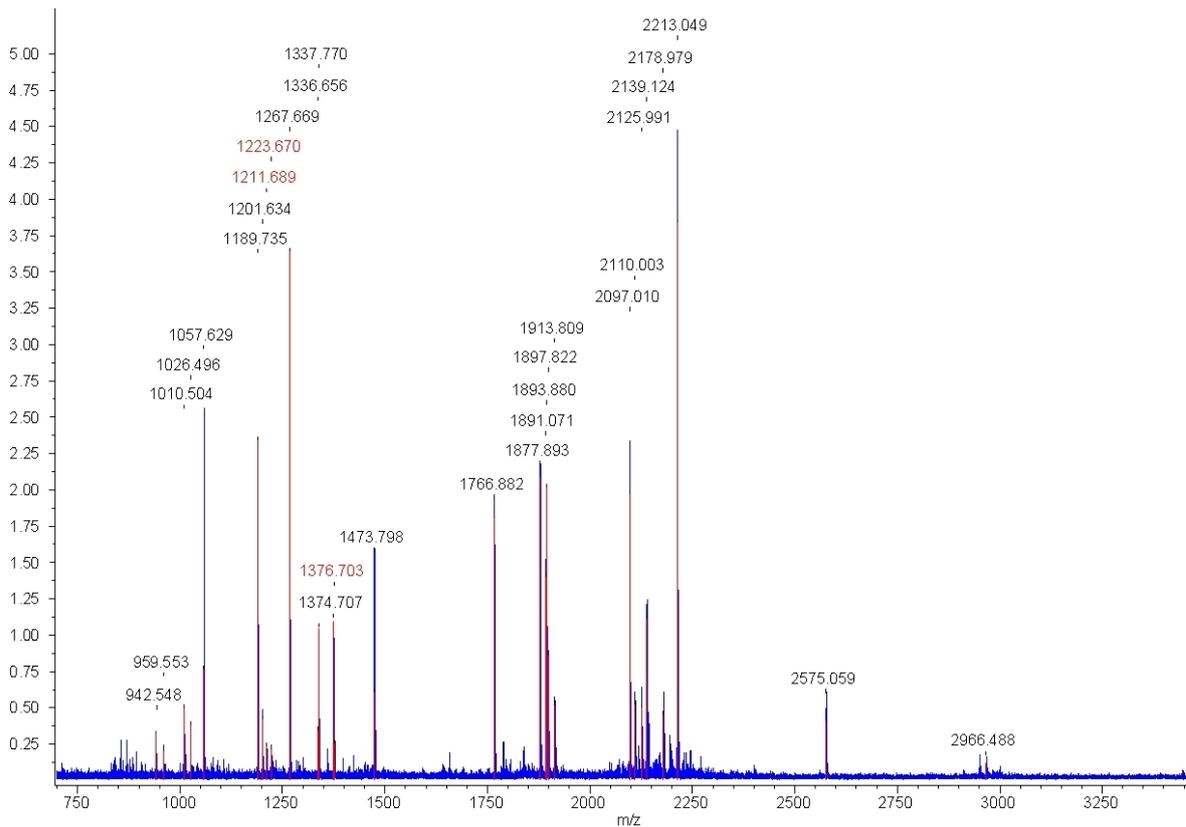
## TCF4

### Sequenzierung des Vektorinserts

GAGCTCGAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTCAGAATTCGGCAACAT  
GGGCCGCTCTGAGCAAAAGCTCATTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCT  
CAACGGTGAGCAGAAGTTAATATCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGCCGC  
ACTGGTACCCCGTGAAGTCTGGAAGTTCTTCCAGGGCCCCGAAGGAGACGCCACCATGGCTAGATCTAGCAA  
TAATGACGATGAGGACCTGACACCAGAGCAGAAGGCAGAGCGTGAGAAGGAGCGGAGGATGGCCAACAATGCCCG  
AGAGCGTCTGCGGGTCCGTGACATCAACGAGGCTTTCAAAGAGCTCGGCCGCATGGTGCAGCTCCACCTCAAGAG  
TGACAAGCCCCAGACCAAGCTCCTGATCCTCCACCAGGCGGTGGCCGTCATCCTCAGTCTGGAGCAGCAAGTCCG  
AGAAAGGAATCTGAATCCGAAAGCTGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGTCTCAGAGCCTCCCC  
TCTCTCCTTGCCGGCCACACCCTGGAATGGGAGACGCATCGAATGGCGCCGCACTAGAATTGTGAGCGCTCA  
CAATTCTGAGAATTTGTATTTTCAGGGTGATCTCCGTGGATCTATTACAACCTCGAGTCTAGGGATCCTCTAGAG

## Proteinsequenz der Bindeseite des Fusionsproteins

S S N N N N N N N N N L G I E G R I S E F G N M G R S E Q K L I S E E D L  
 N G E Q K L I S E E D L N G E Q K L I S E E D L N G S A G I R S G T A A A L  
 V P R G S L E V L F Q G P E G D A T M A R S S N N D D E D L T P E Q K A E R  
 E K E R R M A N N A R E R L R V R D I N E A F K E L G R M V Q L H L K S D K  
 P Q T K L L I L H Q A V A V I L S L E Q Q V R E R N L N P K A A C L K R R E  
 E E K V S S E P P P L S L A G P H P G M G D A S N G G R T R I V S A H N S E  
 N L Y F Q G D L R G S I T T R V



TCF4 MALDI

M/Z	GEFUNDENE FRAGMENTE
717.395578	AACLKR
802.418482	SDKPQTK
831.413354	RMANNAR
835.407583	DINEAFK
847.408268	RMANNAR
867.500044	MVQLHLK
869.471915	ERNLNPK
883.494958	MVQLHLK
941.540424	SGTAAALVPR
960.455947	MANNARER
976.450862	MANNARER

1009.495839	ISEFGNMGR
1025.488238	ISEFGNMGR
1090.577108	VRDINEAFK
1127.612113	NLNPKAACLK
1290.656815	DINEAFKELGR
1322.749275	ELGRMVQLHLK
1338.74419	ELGRMVQLHLK
1373.69894	LISEEDLNGEQK
1373.69894	LISEEDLNGEQK
1472.790583	LISEEDLNGSAGIR
1481.693277	ISEFGNMGRSEQK
1497.688191	ISEFGNMGRSEQK
1590.664539	SSNNDDEDLTPEQK
1651.907961	MVQLHLKSDKPQTK
1667.902875	MVQLHLKSDKPQTK
1845.895601	SEQKLISEEDLNGEQK
1876.885116	GSLEVLFFQGPEGDATMAR
1892.872647	GSLEVLFFQGPEGDATMAR
1946.845356	SSNNDDEDLTPEQKAER
2115.989518	VSSEPPPLSLAGPHPGMGDASN
2131.984433	VSSEPPPLSLAGPHPGMGDASN
2142.288859	LLILHQAVAVILSLEQQVR
2396.265951	LISEEDLNGSAGIRSGTAAALVPR
2427.432563	LLILHQAVAVILSLEQQVRER
2631.212261	EEEKVSSEPPPLSLAGPHPGMGDASN
2647.207175	EEEKVSSEPPPLSLAGPHPGMGDASN
2729.324313	LISEEDLNGEQKLISEEDLNGEQK
2800.417777	SGTAAALVPRGSLEVLFFQGPEGDATMAR
2816.412692	SGTAAALVPRGSLEVLFFQGPEGDATMAR
2828.403961	LISEEDLNGEQKLISEEDLNGSAGIR
2926.696776	SDKPQTKLLILHQAVAVILSLEQQVR
3449.552886	GSLEVLFFQGPEGDATMARSSNNDDEDLTPEQK
3465.547801	GSLEVLFFQGPEGDATMARSSNNDDEDLTPEQK

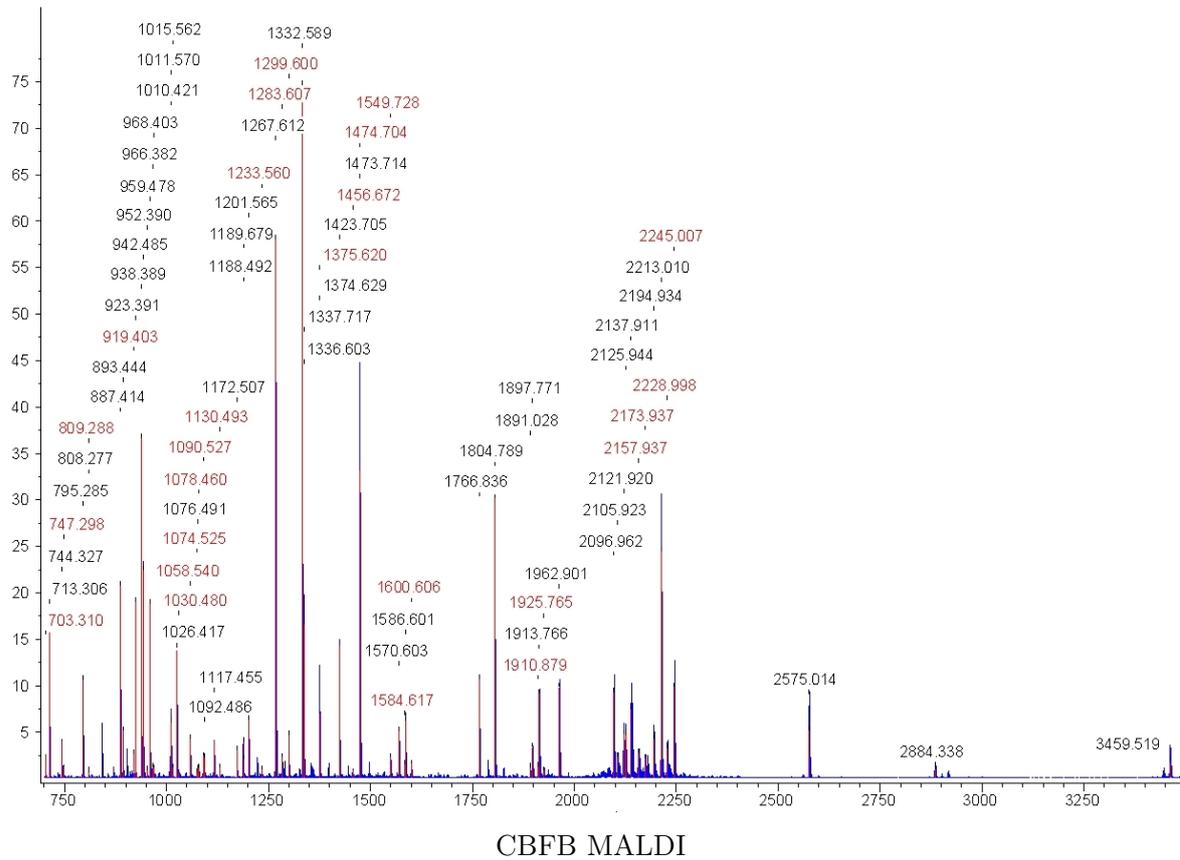
**Tabelle 9.5:** Im Massenspektrometer gefundene Teilfragmente von TCF4

**CBFB****Sequenzierung des Vektorinserts**

GCCGCCAGCGGTCGTCAGACTGTTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACA  
ATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTTCAGAATTCGGCAACATGGGCCGCTCTGAGCAAAAGCT  
CATTTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCAACGGTGAGCAGAAAGTTAATA  
TCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGCCGCACTGGTACCCCGTGGAAAGTCTGG  
AAGTTCTCTTCCAGGGGCCGAAGGAGACGCCACCATGGCTATGCCGCGCTCGTGCCCGACCAGAGAAGCAAGTT  
CGAGAACGAGGAGTTTTTTAGGAAGCTGAGCCGCGAGTGTGAGATTAAGTACACGGGCTTCAGGGACCGGCCAC  
GAGGAACGCCAGGCACGCTTCCAGAACGCCTGCCGCGACGGCCGCTCGGAAATCGCTTTTGTGGCCACAGGAACCA  
ATCTGTCTCTCCAGTTTTTTCCGGCCAGCTGGCAGGGAGAACAGCGACAAACACCTAGCCGAGAGTATGTCGACTT  
AGAAAAGAGAAGCAGGCAAGGTATATTTGAAGGCTCCCATGATTCTGAATGGAGTCTGTGTTATCTGGAAAGGCTGG  
ATTGATCTCCAAAGACTGGATGGTATGGGCTGTCTGGAGTTTGTGAGGAGCGAGCCCAGCAGGAGGATGCATTAG  
CACAAACAGGCCTTTGAAGAGGCTCGGAGAAGGACCGCAATTTGAAGATAGAGACAGGTCTCATCGGGAGGAAAT  
GGAGGTGAGAGTTTTCACAGCTGCTGGCAGTAACTGGCAAGAACAACAAGACCCGGCGCCGCACTAGAATTGTG  
AGCGCTCACAAATCTGAGAATTTGTATTTTCAGGGTGATCTCCGTGGATCTATTACAACCTCGAGTCTAGGGATCCT  
CTAGAGTCGACCTGCAGGCAAGCTT

**Proteinsequenz der Bindeseite des Fusionsproteins**

A A S G R Q T V D E A L K D A Q T N S S S N N N N N N N N N N L G I E G R I S  
E F G N M G R S E Q K L I S E E D L N G E Q K L I S E E D L N G E Q K L I S E  
E D L N G S A G I R S G T A A A L V P R G S L E V L F Q G P E G D A T M A M P  
R V V P D Q R S K F E N E E F F R K L S R E C E I K Y T G F R D R P H E E R Q  
A R F Q N A C R D G R S E I A F V A T G T N L S L Q F F P A S W Q G E Q R Q T  
P S R E Y V D L E R E A G K V Y L K A P M I L N G V C V I W K G W I D L Q R L  
D G M G C L E F D E E R A Q Q E D A L A Q Q A F E E A R R R T R E F E D R D R  
S H R E E M E V R V S Q L L A V T G K K T T R P G G R T R I V S A H N S E N L  
Y F Q G D L R G S I T T R V



M/z	GEFUNDENE FRAGMENTE
712.298315	VVPDQR
732.413003	GSITTRV
743.319159	TTRPGGR
791.348354	EEMEVV
794.277417	FQNACR
807.269282	EEMEVV
871.498798	KTTRPGGR
886.405823	GWIDLQR
902.402177	QTVDEALK
906.517468	EAGKVYLK
922.383132	EYVDLER
927.513779	VVPDQRSK
937.381489	DRPHEER
941.477278	SGTAAALVPR
951.382319	TREFEDR
965.373792	EFEDRDR
1000.552624	TTRPGGRTR
1009.413383	ISEFGNMGR
1014.554586	VSQLLAVTGK
1025.409141	ISEFGNMGR

1033.522629	LSRECEIK
1116.447238	FENEFFFR
1122.498874	FQNACRDGR
1142.702308	VSQLLAVTGKK
1149.546159	QARFQNACR
1171.498691	SHREEMEV
1187.484613	SHREEMEV
1244.582587	FENEFFFRK
1292.633394	DRPHEERQAR
1301.607422	ECEIKYTGFR
1307.635745	EYVDLEREAGK
1331.581172	SKFENEFFFR
1344.699742	AASGRQTVDEALK
1373.621268	LISEEDLNGEQK
1373.621268	LISEEDLNGEQK
1472.706631	LISEEDLNGSAGIR
1481.693277	ISEFGNMGRSEQK
1491.731771	QTPSREYVDLER
1497.688191	ISEFGNMGRSEQK
1499.799262	APMILNGVCVIWK
1515.794177	APMILNGVCVIWK
1561.738587	YTGFRDRPHEER
1569.594739	LDGMGCLEFDEER
1585.592933	LDGMGCLEFDEER
1787.945134	EEMEVRSQLLAVTGK
1803.781658	AQQEDALAQQAFEEAR
1803.781658	EEMEVRSQLLAVTGK
1845.895601	SEQKLISEEDLNGEQK
1959.939866	AQQEDALAQQAFEEARR
1961.893401	IVSAHNSENLYFQGDLR
2003.110032	VYLKAPMILNGVCVIWK
2019.104946	VYLKAPMILNGVCVIWK
2104.915465	GSLEVLFFQGPEGDATMAMPR
2120.912141	GSLEVLFFQGPEGDATMAMPR
2120.912141	GSLEVLFFQGPEGDATMAMPR
2136.902907	GSLEVLFFQGPEGDATMAMPR
2219.108328	TRIVSAHNSENLYFQGDLR
2368.254798	APMILNGVCVIWKGWIDLQR
2384.249713	APMILNGVCVIWKGWIDLQR
2396.265951	LISEEDLNGSAGIRSGTAAALVPR
2438.099479	GWIDLQRDGMGCLEFDEER
2454.094394	GWIDLQRDGMGCLEFDEER

2574.006018	DAQTNSSSNNNNNNNNNNNLGI EGR
2676.361977	IVSAHNSENLYFQGD LRG SITTRV
2729.324313	LISEEDLN GEQK LISEEDLN GEQK
2799.368384	GSLEVL FQGPEG DATMAMPRVVPDQR
2815.363299	GSLEVL FQGPEG DATMAMPRVVPDQR
2815.363299	GSLEVL FQGPEG DATMAMPRVVPDQR
2828.403961	LISEEDLN GEQK LISEEDLN GSAGIR
2831.358213	GSLEVL FQGPEG DATMAMPRVVPDQR
2883.330121	SEIAFVATGTNLSLQFFPASWQGEQR
3028.511026	SGTAAALVPRGSLEVL FQGPEG DATMAMPR
3044.50594	SGTAAALVPRGSLEVL FQGPEG DATMAMPR
3044.50594	SGTAAALVPRGSLEVL FQGPEG DATMAMPR
3060.500855	SGTAAALVPRGSLEVL FQGPEG DATMAMPR
3211.568675	DGRSEIAFVATGTNLSLQFFPASWQGEQR
3355.472133	LDGMGCLEFDEERAQQEDALAQQAFEEAR
3371.467048	LDGMGCLEFDEERAQQEDALAQQAFEEAR
3452.711317	SEIAFVATGTNLSLQFFPASWQGEQRQTPSR
3458.511088	QTVDEALKDAQTNSSSNNNNNNNNNLGI EGR
3565.558474	DAQTNSSSNNNNNNNNNNNLGI EGRISEFGNMGR
3581.553389	DAQTNSSSNNNNNNNNNNNLGI EGRISEFGNMGR

**Tabelle 9.6:** Im Massenspektrometer gefundene Teilfragmente von CBF8

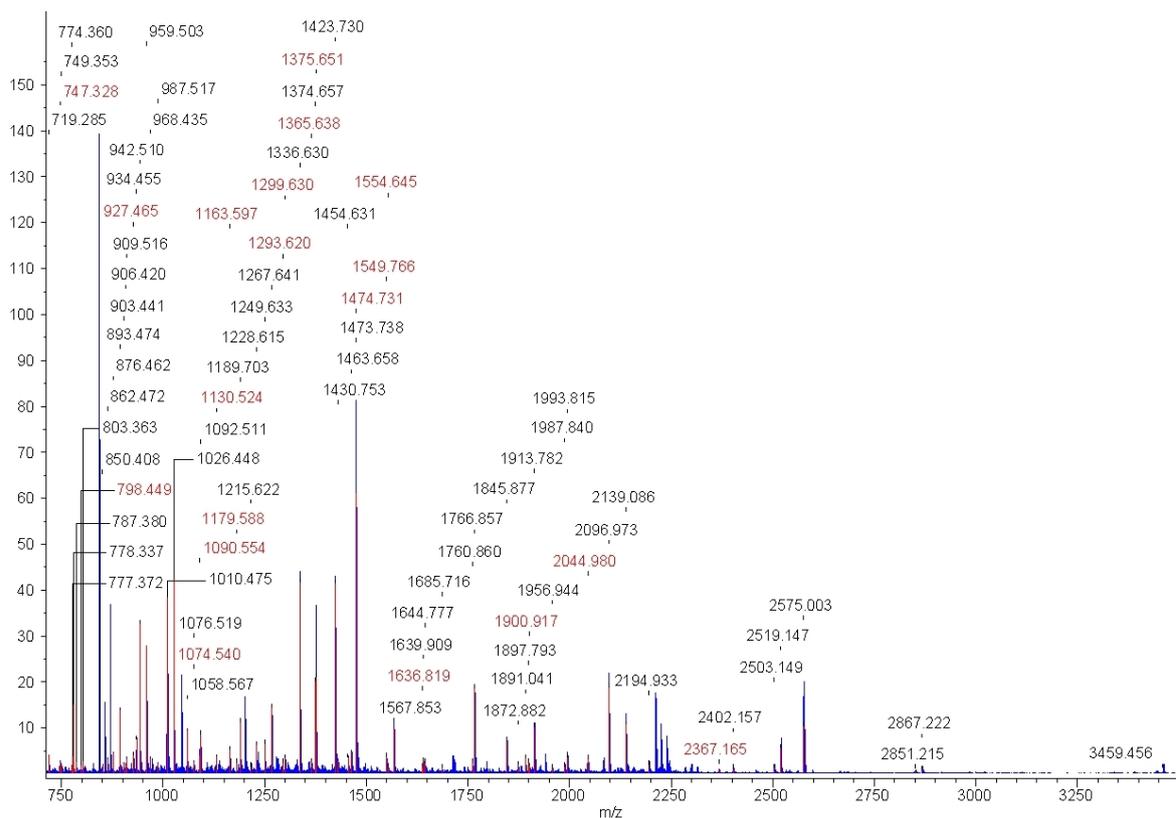
## TFE3

### Sequenzierung des Vektorinserts

GCCGCCAGCGGTCGTCAGACTGTGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACA  
ATAACAATAACAACAACCTCGGGATCGAGGAAGATTTGAGAATTCGGCAACATGGGCCGCTCTGAGCAAAAGCT  
CATTTCTGAAGAGGACTTGAATGGAGAACAGAAATTGATCAGTGAGGAAGACCTCAACGGTGAGCAGAAAGTTAATA  
TCCGAGGAGGATCTTAATGGCTCTGCAGGAATTAGGTCGGGTACCGCGCCGCACTGGTACCCCGTGGAAAGTCTGG  
AAGTTCTCTCCAGGGCCCGAAGGAGACGCCACCATGTCTGCTGAGCTGCCAACATCAAACGGGAGATCTCTGA  
GACCGAGGCAAAGGCCCTTTTGAAGGAACGGCAGAAAGACAATCACAACCTAATTGAGCGTCGCAGGCGATT  
AACATTAACGACAGGATCAAGGAACCTGGCACTCTCATCCCTAAGTCCAGTGACCCGAGATGCGCTGGAACAAGG  
GCACCATCCTGAAGGCCTCTGTGGATTATATCCGCAAGCTGCAGAAGGAGCAGCAGCGCTCCAAAGACCTGGAGAG  
CCGGCAGCGATCCCTGGAGCAGGCCAACCGCAGCTGCAGCTCCGAATTCAGGAAGTAACTGCAGGCCAGATC  
CATGGCCTGCCAGTGCCTCCACTCCAGGGCTGCTTTCCTTGGCGGCCGCACTAGAATTGTGAGCGCTCACAATT  
CTGAGAATTTGTATTTTTCAGGGTATCTCCGTGGATCTATTACAACCTCAGTCTAGGGATCCTCTAGAGTCGACCT  
GCAGGCAAGCTT

## Proteinsequenz der Bindeseite des Fusionsproteins

A A S G R Q T V D E A L K D A Q T N S S S N N N N N N N N N N L G I E G R I S  
 E F G N M G R S E Q K L I S E E D L N G E Q K L I S E E D L N G E Q K L I S E  
 E D L N G S A G I R S G T A A A L V P R G S L E V L F Q G P E G D A T M S A E  
 L P N I K R E I S E T E A K A L L K E R Q K K D N H N L I E R R R R F N I N D  
 R I K E L G T L I P K S S D P E M R W N K G T I L K A S V D Y I R K L Q K E Q  
 Q R S K D L E S R Q R S L E Q A N R S L Q L R I Q E L E L Q A Q I H G L P V P  
 P T P G L L S L G G R T R I V S A H N S E N L Y F Q G D L R G S I T T R V



TFE3 MALDI

M/z	GEFUNDENE FRAGMENTE
728.454474	ALLKER
732.413003	GSITTRV
774.398415	EQQRSK
777.329435	FNINDR
816.40898	SLEQANR
820.338517	SSDPEMR
822.423567	ASVDYIR
833.424296	SKDLESR
836.333432	SSDPEMR
869.522219	ELGTLIPK
902.433365	DLESRQR

---

902.433365	QTVDEALK
905.411787	EISETEAK
928.509028	LQKEQQR
933.447128	RFNINDR
941.501716	SGTAAALVPR
950.51853	ASVDYIRK
958.495592	WNKGTILK
1009.467489	ISEFGNMGR
1009.467489	DNHNLIER
1018.555979	FNINDRIK
1025.440505	ISEFGNMGR
1061.535303	REISETEAK
1100.568668	QRSLEQANR
1110.701246	IKELGTLIPK
1137.58907	KDNHNLIER
1165.595218	DNHNLIERR
1248.625205	SSDPEMRWNK
1264.550635	SSDPEMRWNK
1330.734396	EISETEAKALLK
1334.755801	GTILKASVDYIR
1344.699742	AASGRQTVDEALK
1373.649556	LISEEDLNGEQK
1373.649556	LISEEDLNGEQK
1413.768825	SLEQANRSLQLR
1472.730069	LISEEDLNGSAGIR
1481.693277	ISEFGNMGRSEQK
1497.688191	ISEFGNMGRSEQK
1671.850171	ELGTLIPKSSDPEMR
1687.845086	ELGTLIPKSSDPEMR
1845.895601	SEQKLISEEDLNGEQK
1961.959539	IVSAHNSENLYFQGDLR
2219.108328	TRIVSAHNSENLYFQGDLR
2396.265951	LISEEDLNGSAGIRSGTAAALVPR
2502.140789	GSLEVLFFQGPEGDATMSAELPNIK
2518.138714	GSLEVLFFQGPEGDATMSAELPNIK
2573.995078	DAQTNSSSNNNNNNNNNNLGIEGR
2658.332316	GSLEVLFFQGPEGDATMSAELPNIKR
2674.327231	GSLEVLFFQGPEGDATMSAELPNIKR
2676.361977	IVSAHNSENLYFQGDLRGSITTRV
2729.324313	LISEEDLNGEQKLISEEDLNGEQK
2828.403961	LISEEDLNGEQKLISEEDLNGSAGIR
2832.586163	IQELELQAQIHGLPVPPTPGLLSLGGR

---

3089.734953	IQELELQAQIHGLPVPPTPGLLSLGGRTR
3425.75007	SGTAAALVPRGSLEVLFGPEGDATMSAELPNIK
3429.946008	SLQLRIQELELQAQIHGLPVPPTPGLLSLGGR
3441.744984	SGTAAALVPRGSLEVLFGPEGDATMSAELPNIK
3458.448584	QTVDEALKDAQTNSSNNNNNNNNNNNLGIEGR
3565.558474	DAQTNSSNNNNNNNNNNNLGIEGRISEFGNMGR
3581.553389	DAQTNSSNNNNNNNNNNNLGIEGRISEFGNMGR

**Tabelle 9.7:** Im Massenspektrometer gefundene Teilfragmente von TFE3

## Danksagung

Mein Dank gilt als Erstes Dr. Jörg Hoheisel und Prof. Jürgen Wolfrum für die interessante Aufgabenstellung, für das Vertrauen und die mir gewährte Eigenständigkeit und Freiheit, mit der ich meine Experimente vorantreiben konnte.

Ein besonderer Dank gilt meiner Laborantin Marita Schrenk für die vielen Protein-Expression während der letzten Jahre und die tolle Zusammenarbeit.

Manuel Fugazza danke ich für die tolle Unterstützung bei allen biologischen Fragestellungen und für die tolle Zusammenarbeit während und nach seiner Diplomarbeit.

Anette Jacob danke ich für die vielen tollen und endlosen Diskussionen, für die Korrektur meiner Arbeit und für die vielen Stunden der Heiterkeit.

Svenni Danke ich für das verbessern meiner Tischtennis-Fähigkeiten sowie für die tolle Unterstützung im Labor. Und ja, du bekommst deine Grillade noch...

Dem Boys-Office (Fugi, Mic, Frank und Gast Sven) danke ich für die tolle Zeit, die vielen Diskussionen, die Feierabend-Biere, den Kaffee-Konsum, etc...

Das Kasperle dankt auch dem Kirschwässerle.

Bedanken möchte ich mich auch beim Herr Ober-Administrator Marc für die Ausbildung zum Unter-Administrator sowie für die Konvertierung zu Linux.

Allen Mitgliedern des Arbeitskreises Hoheisel danke ich für die angenehme Arbeitsatmosphäre und die vielen lustigen Stunden im Labor, die Abteilungspartys, die Ski-Seminar, ...

Ich möchte mich bei allen meinen Studienkollegen und Freunden bedanken die mich während des Studiums und der Promotion begleitet haben. Achim, Andrea, Aydin, Axel, Barzi, Bella, Christoph, Eva, Helmut, Jof, Jorge, Kai, Katsche, Kuli, Marc, Markus, Robert, Sigggi, Simon, Staudt, Theo und Zuhäl.

Nicht zuletzt danke ich meiner Familie, der Wunsch-Familie und meinen Freunden, ohne deren Rückhalt und Zuspruch diese Arbeit niemals Wirklichkeit geworden wäre.

Und natürlich Julia...

## Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbst verfasst habe und mich dabei keiner anderen, als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Ich habe an keiner anderen Stelle ein Prüfungsverfahren beantragt, beziehungsweise die Dissertation in dieser oder anderer Form bereits als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt.

Christian Krauß

Heidelberg, 12. Dezember 2008