

4 Diskussion

Ziel der Arbeit war erstens die Herstellung von rekombinanten *Arabidopsis thaliana*-Proteinen und zweitens die Etablierung von funktionellen Studien im Hochdurchsatz. Aus diesem Grunde wurden diese Themen in zwei Abschnitten zusammengefasst und diskutiert. Zur Gewinnung rekombinanter Proteine im ersten Teil der Arbeit wurden cDNA-Bibliotheken aus der Pistill und der Infloreszenz von *Arabidopsis thaliana* erstellt und charakterisiert. Ausgehend von der Infloreszenz cDNA-Bibliothek wurde ein Uniklonset von Expressionsklonen erstellt. Zu diesem Zweck wurden potentielle Expressionsklone dieser Bibliothek unter Verwendung von Hochdichte-Proteinfiltern selektiert und in einer Proteinexpressions-Unterbibliothek neu angeordnet. Basierend auf BLAST- und Cluster-Analysen dieser Klone, erfolgte anschließend die Erstellung eines Uniklonsets aus 1.502 Klonen, das um 192 vollständige cDNA-Klone erweitert wurde. Alle 1.694 Klone dieses Erweiterten Uniklonsets wurden im Hochdurchsatz exprimiert und die entsprechenden rekombinanten Proteine aufgereinigt.

Im zweiten Abschnitt wurden mit den aufgereinigten Proteinen *Arabidopsis*-Protein-Microarrays erstellt, welche anschließend für Phosphorylierungsstudien *in vitro* verwendet wurden. Diese Studien wurden unter Verwendung verschiedener Kinasen durchgeführt: einerseits mit der gut beschriebenen Maus Proteinkinase A (PKA) und andererseits mit zwei *Arabidopsis* MAP-Kinasen (MPK3 und 6), die eine wichtige Rolle in der Signaltransduktion spielen. Ziel war es, nach Weiterentwicklung des Testsystems, unter Verwendung der PKA, potentielle Phosphorylierungs-Targets der MAP-Kinasen zu identifizieren.

4.1 Gewinnung rekombinanter *Arabidopsis* Proteine

4.1.1 Herstellung und Charakterisierung von zwei cDNA-Bibliotheken

Die cDNA-Bibliotheken repräsentieren im Idealfall die Gesamtheit der exprimierenden Gene eines Gewebes, eines Zelltyps oder eines Organismus' zu einem bestimmten Zeitpunkt. Basierend auf der Analyse dieser Bibliotheken, kann man wichtige Einblicke in den Ablauf biologischer Prozesse bekommen. So können z.B. neue Gene und deren endgültige Spleißprodukte identifiziert werden, aber auch die Expressionsintensität der Gene des entsprechenden Stadiums des jeweiligen Gewebes, Zelltyps bzw. Organismus' analysiert werden (Clark *et al.*, 1999).

cDNA-Bibliotheken können auch als Quelle zur Expression rekombinanter Proteine des entsprechenden Gewebes und Stadiums genutzt werden. Dazu werden die cDNAs entweder

direkt in einen geeigneten Expressionsvektor (Büssow *et al.*, 1998; Gutjahr, 2003; Kramer, 2003) oder zunächst in einen Entryvektor des GATEWAY-Systems kloniert. Ausgehend von diesem Entryvektor kann dann die cDNA über Rekombination je nach Bedarf in unterschiedliche Expressionsvektoren transferiert werden (Heyman *et al.*, 1999; Walhout *et al.*, 2000). Im Rahmen dieser Arbeit wurden beide Wege zur Herstellung von cDNA-Bibliotheken verwendet. Die Infloreszenz cDNA-Bibliothek wurde direkt im *E. coli* Expressionsvektor pQE30NASTattB hergestellt. Dieses bakterielle Expressionssystem wurde wegen seiner guten Eignung für den Hochdurchsatz ausgewählt (vgl. 1.3.2). Eine weitere cDNA-Bibliothek aus dem Pistillgewebe von *Arabidopsis* wurde in den Entryvektor pENTR1a des GATEWAY-Systems kloniert und steht für zukünftige funktionelle Studien in verschiedensten Expressionssystemen zur Verfügung.

Die Charakterisierung von jeweils 96 Klonen beider cDNA-Bibliotheken bezüglich der durchschnittlichen Insertlänge lieferte Werte von 1,01 Kbp (Pistill) bzw. 1,06 Kbp (Infloreszenz). Diese Werte liegen damit relativ nahe an der durchschnittlichen Genlänge von 1,3 Kbp des *Arabidopsis*-Genoms (*Arabidopsis*-Genome-Initiative, 2000), was für eine gute Qualität der hergestellten cDNA-Bibliotheken spricht. Ähnliche durchschnittliche Genlängen wurden auch in den Genomen von *Oryza sativa* L. *ssp. indica* (1,1 Kbp; (Yu *et al.*, 2002)) und *Oryza sativa ssp. japonica* (1,3 Kbp; (Goff *et al.*, 2002)) gefunden. Auch die beiden von (Kramer, 2003) erstellten Gerste cDNA-Bibliotheken ergaben eine durchschnittliche Insertlänge von ungefähr 1 Kbp (Perikarp: 996 Bp und Embryosack: 932 Bp).

Der Anteil an Religanden konnte für 96 Klone der Infloreszenz cDNA-Bibliothek mit Hilfe einer PCR mittels genspezifischer Primer für *phyB* (Insert im Ausgangsvektor) ermittelt werden und lag bei 2,1%. Diese Religationen können sowohl auf eine unvollständige Restriktion des Vektors pQE30NASTattB als auch auf eine Kontamination des geschnittenen Vektors mit dem ausgeschnittenen Insert (*phyB*) zurückzuführen sein.

Dagegen wurde der Anteil der Religanden der 95 Klone der Pistill cDNA-Bibliothek nur ausgehend von der theoretischen PCR-Produktlänge der Religanden ermittelt und lag bei 6,3%. Dabei ist es nicht auszuschließen, dass sich unter den 6,3 % einige Klone befinden, die keine Religanden sind, aber zufällig ein PCR-Produkt in derselben Größe lieferten. Aufgrund der unterschiedlichen Ermittlung der Religationenanteile sind die beiden Werte nicht unmittelbar miteinander vergleichbar.

Im Falle des Vektors pENTR1a wäre eine Negativselektion der Religanden aufgrund der Suizid-Funktion des *ccdB*-Genes zu erwarten gewesen; der hier erhaltene hohe Anteil von Religanden (6,3%) könnte jedoch auf unvollständige Aktivität des *ccdB*-Gens zurückzuführen sein.

Aufgrund der Testsequenzierung von 48 Pistill-Klonen wurde ersichtlich, dass basierend auf einer BLAST-Analyse allen sequenzierten Klonen einem AGI *gene code* zugeordnet werden konnten. Dabei zeigte sich die hohe Diversität der Klone. Nur ein Klon kam zweimal vor, die restlichen Klone (95,8%) waren je einmal vertreten (*Singeltons*). Der Anteil an Klonen in voller Länge lag bei 23,4%. Darüber hinaus zeigte sich, dass sich 31,3% der Klone im richtigen Leserahmen (d.h. der Leserahmen des RGS His₆-Tags) befanden.

In die Testsequenzierung der Infloreszenz cDNA-Bibliothek wurde eine größere Anzahl von Klonen (576) einbezogen, da diese cDNA-Bibliothek der Ursprung für ein Proteinexpressions-Uniklonset war. 418 Sequenzen mit ausreichender Qualität konnte ein AGI *gene code* zugeordnet werden. Bei 7 Sequenzen handelte es sich um *phyB*, welche den Religandenanteil darstellen und nicht in die weitere Auswertung eingingen. Die Analyse der Diversität der Klone mittels Cluster-Analyse, auf der Grundlage der AGI-*gene codes* ergab, dass 72,3% der Klone einmal vorkamen (*Singeltons*), während die restlichen 27,7% sich in 49 Cluster zusammenfassen ließen. 192 der 411 Klone (46,7%) lagen in voller Länge und 128 Klone (31,1%) im richtigen Leserahmen vor.

Ein Vergleich der Diversität der Infloreszenz cDNA-Bibliothek (72,3% *Singeltons*) mit der Pistill cDNA-Bibliothek (95,8% *Singeltons*) ist nicht sinnvoll, da die Anzahl der analysierten Klone der Pistill cDNA-Bibliothek zu gering dazu war. Dies ist auch eine mögliche Ursache für die Differenz der Klone, die in voller Länge vorliegen (Pistill: 23,4%; Infloreszenz: 46,7%). Weitere Gründe für einen geringeren Anteil an kompletten cDNA-Klonen können auch auf die mRNA Qualität (mögliche Degradation) und die cDNA-Synthese zurückzuführen sein. Bei der Synthese der Infloreszenz cDNA könnte die Reverse Transkriptase effektiver gewesen sein, welches die Häufigkeit erhöhte, ein Gen in voller Länge zu transkribieren.

Bezogen auf den Anteil der Klone, die sich im richtigen Leserahmen befanden, ist für beide in dieser Arbeit erstellten Bibliotheken (Pistill: 31,3% und Infloreszenz: 32,3%) kein direkter Vergleich mit anderen cDNA-Bibliotheken möglich, da für keine der bisher beschriebenen cDNA-Bibliotheken vor Neuordnung der Expressionsklone eine aussagekräftige Leserahmennaussage auf Sequenzbasis beschrieben wurde. Kramer 2003 testeten die Expression von 96 Klonen aus einer Gerste-Embryosack cDNA-Bibliothek und

zeigten, dass 18% dieser Klone eine im SDS-Polyacrylamidgel sichtbare Überexpressionsbande lieferten. Das lässt vermuten, dass in dieser Bibliothek mindestens 18% der Klone im richtigen Leserahmen vorliegen. Jedoch kann davon ausgegangen werden, dass der Anteil höher ist, da bei dieser Methode die schwach translatierten Proteine nicht erfasst werden. Im Gegensatz dazu kann es bei der Ermittlung des richtigen Leserahmens auf Sequenzbasis dazu kommen, dass Klone miterfasst werden, die zwar im richtigen Leserahmen sind aber aufgrund von Stopcodonen in der 5'-untranslatierte Region nicht exprimiert werden.

Die Pistill cDNA-Bibliothek setzte sich mit 660.800 Klonen aus deutlich mehr Klonen zusammen, als die Infloreszenz cDNA-Bibliothek, die aus 40.000 Klonen besteht. Als eine Hauptursache hierfür kann die Verwendung unterschiedlicher kompetenter Zellen angesehen werden. Die Transformationseffizienz der käuflich erworbenen DH5 α -Zellen war um den Faktor 10 höher als die der selbsthergestellten kompetenten SCS1/pSE111-Zellen. Weitere Faktoren können sowohl in den unterschiedlichen Vektoren, z.B. die variierende Vektorgröße (pENTR1a: 2.717 Bp und pQE30NASTattB: 3.575 Bp) als auch in der Ligationseffizienz zu finden sein. Ein weiterer, jedoch nicht so gewichtiger Faktor, um die unterschiedlichen Bibliotheksgrößen zu erklären, ist eine mögliche negative Selektion einiger Proteinexpressionsklone im pQE30NASTattB Vektor durch toxische oder das Wachstum beeinträchtigende Proteine, bei einer nicht vollständig unterdrückten Expression (*leaky expression*).

Da jedoch nur die Infloreszenz cDNA-Bibliothek in einen Expressionsvektor (pQE30NASTattB) kloniert wurde und der zeitliche und finanzielle Aufwand zu hoch gewesen wäre, alle 660.800 Pistill-Klone in einen Expressionsvektor zu überführen, wurden im Folgenden ausschließlich die Klone der Infloreszenz cDNA-Bibliothek verwendet.

Die Pistill cDNA-Bibliothek wurden in einen Entryvektor (pENTR1a) kloniert und vereinigt. Diese Klone stehen somit z. B. als Ressource für ein späteres Überführen in unterschiedliche Expressionsvektoren zur Verfügung.

Die positiven Ergebnisse, die aus der Testsequenzierung der Infloreszenz Klone gewonnen wurden, belegen die gute Qualität dieser cDNA-Bibliothek. Ausgehend von dieser cDNA-Bibliothek wurden im weiteren Verlauf potentielle Expressionsklone zur Erstellung einer Proteinexpressions-Unterbibliothek detektiert.

4.1.2 Identifizierung potentieller Expressionsklone und Herstellung der Proteinexpressions-Unterbibliothek

Die Klonierung der Infloreszenz cDNA-Bibliothek erfolgte in den *E. coli* Expressionsvektor pQE30NASTattB. *E. coli* wurde als Expressionswirt verwendet, da er eine stabile Expression ermöglicht, auf genetischer und biochemischer Ebene gut charakterisiert ist und eine hohe Reproduktivität aufweist. Aufgrund dieser Eigenschaften ist *E. coli* sehr gut für den hohen Durchsatz geeignet. Außerdem ist die im Haus entwickelte Pick- und Spotttroutine mit diesem Expressionssystem bereits etabliert.

Der verwendete Expressionsvektor ermöglicht die rekombinante Proteinexpression über Induktion des T5-Promoters mit IPTG. Darüber hinaus enthält der Vektor ein N-terminales RGS His₆-Tag, womit die Proteine nach der Expression fusioniert sind. Damit können die Fusionsproteine über *Immobilised Metal Affinity Chromatography* (IMAC), sowohl unter nativen, als auch unter denaturierenden Bedingungen aufgereinigt werden. Aufgrund der geringen Größe des Tags geht man davon aus, dass die Struktur und Funktion des fusionierten Proteins nicht beeinträchtigt wird und das Tag demzufolge nach der Aufreinigung nicht entfernt werden muss. Des Weiteren kann das Tag auch noch zur Detektion der rekombinanten Proteine mit Hilfe eines anti-RGS His₆-Antikörpers verwendet werden.

Die Notwendigkeit der Detektion der potentiellen Expressionsklone aus einer cDNA-Bibliothek ist darauf zurückzuführen, dass beim Klonieren der cDNA kein Einfluss darauf genommen werden kann, in welchem der drei Leserahmen die cDNA transferiert wird. Somit ist zu erwarten, dass etwa nur ein Drittel der klonierten Gene im richtigen Leserahmen vorliegen und daher als potentielle Expressionsklone identifiziert werden können. Dies konnte anhand der Testsequenzierung für die im Rahmen dieser Arbeit erstellten cDNA-Bibliotheken bestätigt werden (Pistill: 31,3% und Infloreszenz: 32,3% lagen im richtigen Leserahmen vor).

Die Klone der Infloreszenz cDNA-Bibliothek wurden mittels Robotertechnologie auf PVDF-Membranen gespottet. Anschließend wurde *in situ* die Expression der rekombinanten Fusionsproteine induziert, welche anschließend mit Hilfe des spezifischen anti-RGS His₆-Antikörpers identifiziert werden konnten.

Basierend auf dem Antikörper-*screening* der Hochdichte-Proteinfiler konnten 7.000 Klone als potentielle Expressionsklone identifiziert werden. Das entspricht einem Anteil von 17,5% der 40.000 Klone aus der cDNA-Bibliothek. Auch in den bereits oben erwähnten

Bibliotheken von (Büssow *et al.*, 1998; Gutjahr, 2003; Kramer, 2003) lag der Anteil an potentiellen Expressionsklonen bei ungefähr 20%. Mögliche Ursache dafür, dass die Anzahl an potentiellen Expressionsklonen niedriger ist als die zu erwartende Anzahl (ein Drittel) kann u.a. sein, dass gering exprimierende Proteine gar nicht oder in zu geringer Menge synthetisiert werden, um sie als Expressionsklone zu identifizieren. Berücksichtigt werden muss auch, dass ein Teil der exprimierten Proteine toxisch auf *E. coli* wirken kann bzw. die Proteine durch die Zellen degradiert werden können. Des Weiteren kann bei Klonen, die in voller Länge vorliegen, die 5'-untranslatierte Region zusätzliche Stopcodone enthalten, so dass der Klon das enthaltene Gen nicht exprimiert.

Neben dem *screening* der Hochdichte-Proteinfilter mit dem anti-RGS His₆-Detektionsantikörper, wurde ein weiterer Filtersatz mit einem anti-PhyB-Antikörper inkubiert. Somit konnten die Religanden (Ausgangsvektor mit Insert *phyB*) identifiziert und bei der Erstellung der entsprechenden Unterbibliotheken ausgeschlossen werden. 2.000 der 7.000 potentiellen Expressionsklone wurden als Religanden identifiziert. Dieser hohe Anteil an Religanden kann auf einen nicht vollständig geschnittenen Ausgangsvektor (*SalI* und *NotI*) zurückzuführen sein.

Aufgrund der systematischen Anordnung auf den klonalen Proteinfiltern war es möglich, die 5.000 potentiellen Expressionsklone (abzüglich der Religanden) zu identifizieren und diese in einer Proteinexpressions-Unterbibliothek neu anzuordnen. Um eine definierte Aussage über die Diversität der Bibliothek machen zu können und um die Erstellung eines Uniklonsets zu ermöglichen, wurde die gesamte Proteinexpressions-Unterbibliothek ausgehend vom 5'-Ende ansequenziert und analysiert.

4.1.3 Charakterisierung der Proteinexpressions-Unterbibliothek

Ausgehend von der Sequenzinformation wurde zunächst eine Qualitätskontrolle durchgeführt. 4.398 Sequenzen wurden hinsichtlich ihrer Qualität als ausreichend befunden. Von diesen Sequenzen wurde eine BLAST-Analyse gegen die MIPS-*Arabidopsis*-Protein-Datenbank durchgeführt. 99,6% der Sequenzen konnte ein AGI *gene code* zugeordnet werden. Bei den restlichen Sequenzen, denen kein AGI *gene code* zugeordnet werden konnte, könnte es sich um bisher noch nicht beschriebene Gene handeln.

Basierend auf der BLAST-Analyse konnte gezeigt werden, dass sich 61,8% der sequenzierten Ausgangsklone in Bezug auf den RGS His₆-Tag im richtigen Leserahmen befanden und dass 39,4% der Proteinexpressions-Unterbibliotheks-Klone in voller Länge vorlagen. Dabei zeigte sich, dass sich im Vergleich zur Testsequenzierung (vor dem Rearray) der Anteil vollständiger cDNA-Klone etwas verringert hat, von 46,7%

(vor dem Rearray) auf 39,4% (nach dem Rearray). Eine mögliche Ursache hierfür kann sein, dass bei einigen vollständigen cDNA-Klonen die 5'-untranslatierte Region transkribiert wird, welche Stopcodonen aufweisen kann. Diese Klone liefern kein Fusionsprotein mit dem RGS His₆-Tag und sind damit nicht detektierbar und wurden deshalb nicht in die Proteinexpressions-Unterbibliothek aufgenommen. Der Anteil der Klone im richtigen Leserahmen hat sich dagegen fast verdoppelt, von 32,3% (vor dem Rearray) auf 61,8% (nach dem Rearray). Diese Anreicherung zeigt, dass sowohl die Identifizierung der potentiellen Expressionsklone, als auch deren erneute Anordnung in eine Proteinexpressions-Unterbibliothek außerordentlich wichtig ist. Vergleichbare Studien, in denen auch eine große Anzahl von Expressionsklonen auf Sequenzbasis analysiert wurde, finden sich bei Kramer 2003 und Lüking *et al.* 2003b.

Kramer 2003 sequenzierten 4.100 putative Expressionsklone der Gerste-Embryosack cDNA-Bibliothek. Von 2.416 Sequenzen mit ausreichender Qualität lagen 43% im richtigen Leserahmen vor.

In der Studie von Lüking *et al.* 2003b wurden 2.413 Proteine der humanen fötalen Hirngewebe Bibliothek (Büssow *et al.*, 1998) auf einen Microarray gespottet. Dafür wurden die verwendeten Klone sequenziert und 2.303 der 2.413 Klone konnten einer Sequenz zugeordnet werden. Basierend auf einer BLAST-Analyse (gegen die nichtredundanten Proteindatenbank NRPROT) konnte für 1.603 Klone der Leserahmen ermittelt werden. Von diesen 1.603 Klonen waren 55% im richtigen Leserahmen. Somit liegt der in der vorliegenden Arbeit erzielte Anteil von 61,8% Klonen mit richtigem Leserahmen deutlich über den Werten aus den Studien von Kramer 2003 und Lüking *et al.* 2003b, womit die gute Qualität des in dieser Arbeit durchgeführten Rearrays unterstrichen wird.

In der Studie von Lüking *et al.* 2003b lagen von den 882 Klonen, die sich im richtigen Leserahmen befanden, 382 (43%) in voller Länge vor. Im Vergleich dazu befanden sich in der vorliegenden Arbeit 926 Klone (34,1%) in voller Länge, wenn man von den 2.715 Sequenzen ausgeht, die im richtigen Leserahmen vorlagen. Somit war im Vergleich zu Lüking *et al.* 2003b zwar der Anteil der Klone, die sich im richtigen Leserahmen befanden höher; allerdings war der Anteil der Klone, die in voller Länge vorlagen, geringer. Eine mögliche Begründung für den geringeren Anteil an Klonen, die in voller Länge vorliegen, könnte sein, dass die Reverse Transkriptase nicht so effektiv war bzw. bei der Größenfraktionierung der cDNA auch große cDNAs verloren gingen.

Eine mögliche Ursache dafür, dass sich nach der Neuordnung der Expressionsklone nicht alle Klone im richtigen Leserahmen befinden, liegt darin, dass einige der kleinen Proteine,

die nicht im richtigen Leserahmen vorliegen, exprimiert werden. Wenn die Länge dieser Proteine ausreicht, um nicht von *E. coli* abgebaut zu werden, können diese mit dem anti-RGS His₆-Antikörper detektiert werden und gelangen somit in das potentielle Expressionssubset (Büssow *et al.*, 1998; Lüking *et al.*, 2003b).

Zur weiteren Analyse wurden die 4.398 *Arabidopsis*-Sequenzen mit allen kodierenden Sequenzen von *Arabidopsis* zusammen geclustert. Diese Analyse zeigte, dass 31,7% der Expressionsklone einmal vorkamen, während die restlichen Klone mehrfach repräsentiert waren und sich in 637 Cluster zusammenfassen lassen.

Basierend auf diesen Analysen wurde aus der Proteinexpressions-Unterbibliothek das Uniklonset erstellt.

4.1.4 Erstellung des Uniklonsets

Als erstes Selektionskriterium für die Erstellung des Uniklonsets wurde die Position des ersten Stopcodons einbezogen. Dazu wurden die DNA-Sequenzen translatiert.

Basierend auf der Position des ersten Stopcodons wurden alle Sequenzen eliminiert, die ein Stopcodon bei ≤ 70 Aminosäuren aufwiesen (1.442 Klone, entspricht 32,8% der Expressionsklone). Somit wurde gewährleistet, dass nur solche Klone einbezogen wurden, die Proteine mit einer Größe von mindestens 7,7 kDa exprimieren. Diese Größe wurde gewählt, da Lüking *et al.* 2003b zeigten, dass bei der Identifizierung von Expressionsklonen auf klonalen Proteinfiltren auch Peptide (welche im Leserahmen des RGS His₆-Tags vorlagen) mit einer Größe von 5 – 8 kDa zu detektieren waren.

Die verbliebenen 2.956 Sequenzen wurden in die folgenden zwei Sequenzgruppen unterteilt:

1. Sequenzen, die bei der Cluster- und BLAST-Analyse gleiche AGI *gene codes* als Clusterpartner bzw. besten BLAST-Hit aufwiesen (2.754 Sequenzen)
2. Sequenzen, die basierend auf der Cluster-Analyse keinem AGI *gene code* zugeordnet werden konnten bzw. Sequenzen, die bei der Cluster- bzw. BLAST-Analyse unterschiedliche AGI *gene codes* als Clusterpartner bzw. besten BLAST-Hit hatten (202 Sequenzen).

Diese Sequenzgruppen sind die Grundlage zur Erstellung des Uniklonsets. Die Selektionskriterien, mit denen die Sequenzen für das Uniklonset aus diesen beiden Sequenzgruppen ausgewählt wurden, sind im folgenden kurz dargestellt, um die Daten im Anschluss daran besser diskutieren zu können.

Die Selektionskriterien waren für die Sequenzen der beiden Sequenzgruppen unterschiedlich. Da bei den Sequenzen der ersten Sequenzgruppe BLAST- und Cluster-AGI *gene code* identisch waren, erfolgte hier die Selektion basierend auf der BLAST- und der

Cluster- BLAST-Analyse. Dagegen war in der zweiten Sequenzgruppe (BLAST- und Cluster-AGI *gene code* nicht identisch) nicht sicher, dass die Auskunft der BLAST-Analyse korrekt war und somit wurden die Parameter dieser Analyse nicht mit einbezogen. Hier gingen als Selektionskriterien das Stopcodon und die Clustergröße ein.

Bei den Sequenzen der ersten Gruppe wurden alle Klone, die nicht im Leserahmen des RGS His₆-Tags (11,6%) vorlagen, sowie alle Klone, welche einen E-Wert (Irrtumswahrscheinlichkeit) über e^{-20} aufwiesen und somit als nicht sicher genug galten, nicht berücksichtigt.

Ausgehend von der Cluster-Analyse wurden alle Klone die je einmal vorkamen (37,5%; *Singeltons*), direkt, ohne weitere Selektion, in das Uniklonset transferiert. Die restlichen Sequenzen ließen sich in 384 Cluster zusammenfassen. Da in dem zu erstellenden Uniklonset jedes Gen möglichst nur einmal vertreten sein sollte, waren hier weitere Selektionskriterien erforderlich. Diese waren der Startpunkt des Subjekts (Treffer-Sequenz), der Startpunkt des Querys (Abfrage-Sequenz) und der E-Wert aus der BLAST-Analyse. Die Kriterien waren in der eben beschriebenen Reihenfolge festgelegt und sollten alle so klein wie möglich sein. Im Falle des Startpunktes des Subjekts heißt das, je kleiner der Wert, desto länger ist der Klon bzw. im Idealfall, wenn der Wert gleich eins ist, liegt das klonierte Gen in voller Länge vor. Beim Startpunkt des Querys bedeutet der kleinste Wert, dass der Klon mit der kürzesten 5'-untranslatierten Region genommen wurde. Sollte aufgrund der Aussage bezüglich des Startpunktes des Subjekts und dem Startpunkt des Querys keine Sequenzauswahl innerhalb eines Clusters möglich sein, wurde als weiteres Kriterium der E-Wert nochmals einbezogen. Es wurde der Klon mit dem kleinsten E-Wert aus dem Cluster genommen. Damit wurde der Grad der Wahrscheinlichkeit, dass es sich bei dem ausgewählten Klon um das bei der BLAST-Analyse gefundene Gen handelt, nochmals erhöht (vgl. 3.4.2.2).

Ausgehend von den 384 Clustern wurden jedoch mehr als 384 Klone in das Uniklonset transferiert. Insgesamt wurden 419 Klone für das Uniklonset ausgewählt. Das ist darauf zurückzuführen, dass im Falle von 21 Clustern den Sequenzen innerhalb eines Clusters mehr als ein BLAST-AGI *gene code* zugeordnet werden konnte und dass aus diesen Clustern mehrere Klone (ein Klon pro BLAST-Hit) nach den oben genannten Kriterien selektiert wurden. Mit dieser Vorgehensweise sollten die wenigen Fälle berücksichtigt werden, bei denen unterschiedliche Gene mittels Cluster-Analyse aufgrund hoher Sequenzähnlichkeiten in eine Cluster eingeordnet wurden. Diese verschiedenen Gene sollten im Uniklonset erfasst werden. Jedoch wird hierbei auch in Kauf genommen, dass einige wenige Gene mehrfach in

dem Uniklonset repräsentiert sind. Diese Mehrfachrepräsentierung kommt dadurch zustande, dass im Laufe der Evolution einige der Gene auf den fünf Chromsomen mehrfach verliegen und somit unterschiedlichen *Accession*-Nummern zugeordnet werden können (Mayer, 2001).

Von den Sequenzen der ersten Sequenzgruppe gingen somit 48,1% der Klone in das Uniklonset ein.

Wie bereits beschrieben, ging die Selektion der Sequenzen der zweiten Sequenzgruppe von weniger Kriterien aus. Die Kriterien waren hier: die Position des Stopcodons und die Clustergröße. Die zweite Sequenzgruppe ließ sich in zwei weitere Untergruppen unterteilen. In der einen Untergruppe konnten den Sequenzen mittels Cluster-Analyse kein AGI *gene code* zugeordnet werden. Für die Sequenzen der anderen Untergruppe konnte zwar auf Grundlage der Cluster-Analyse ein AGI *gene code* ermittelt werden, dieser war jedoch nicht identisch mit dem der BLAST-Analyse.

Von beiden Untergruppen wurden auch wieder alle *Singeltons* direkt in das Uniklonset transferiert (82 der 202 Klone). Die restlichen Klone lagen in Clustern vor. Von den Clustern, die sich aus weniger als sechs Sequenzen zusammensetzten, gingen alle Klone in das Uniklonset ein. Dagegen gingen bei Clustern mit mehr als sechs Klonen nur die Klone ein, die kein Stopcodon besaßen. Zusätzlich zu den 82 *Singeltons* gingen somit noch 92 weitere Klone in das Uniklonset ein. Das bedeutet, dass von Clustern der Sequenzgruppe 2 teilweise mehr als ein Klon in das Uniklonset eingehen konnte. Diese Vorgehensweise sollte ermöglichen, dass sich unter den ausgewählten Klone auch Klone im richtigen Leserahmen befanden, da hier keine Selektion bezüglich der Leserahmen möglich war.

Ausgehend von 2.956 potentiellen Expressionsklonen der Proteinexpressions-Unterbibliothek, deren Sequenzen qualitätsgerecht waren und kein Stopcodon bei ≤ 70 Aminosäuren besaßen, gingen 1.502 Klone (50,7%) in das Uniklonset ein. Zusätzlich zu den 1.502 Uniklonset-Klonen kamen noch 192 Klone ((Kersten *et al.*, 2003) und Klone von Transkriptionsfaktoren) hinzu, die in voller Länge vorlagen. Diese nunmehr insgesamt 1.694 Klone bildeten zusammen das Erweiterte Uniklonset.

4.1.5 Herstellung von Protein-Microarrays des Erweiterten Uniklonsets

Alle Klone des Erweiterten Uniklonsets wurden im 96 *well* Format exprimiert und die Proteine unter denaturierenden Bedingungen aufgereinigt. Ausgehend von drei zufällig ausgewählten 96 *well* Mikrotiterplatten wurde mit Hilfe des Bradford-Tests eine Proteinausbeute von durchschnittlich 8,5 μg bestimmt. Diese lag jedoch unter der erzielten durchschnittlichen Ausbeute von (Kersten *et al.*, 2003) mit 24,5 μg bei 96 *Arabidopsis*-

Klonen. Diese Diskrepanz könnte auf das vergleichsweise kleinere Elutionsvolumen zurückzuführen sein, welches in der vorliegenden Arbeit während der manuellen Reinigung eingesetzt wurde (80 µl Elutionspuffer). Die Proteine von Kersten *et al.* 2003 dagegen wurden unter Verwendung des BioRoboter 8000 (Qiagen) in 350 µl eluiert. Trotz der etwas niedrigeren Gesamtausbeute erfolgte in der vorliegenden Arbeit eine bewusste Entscheidung zugunsten der manuellen Aufreinigung, da hiermit höhere Proteinkonzentrationen erzielt werden können (Uniklonset: 106 µg/ml, Kersten *et al.* 2003: 70 µg/ml). Somit können Microarrays mit einem höheren Proteingehalt generiert werden.

Die 1.694 Proteine des Erweiterten Uniklonsets wurden in einem 11 x 11-Pattern gespottet. Nach dem Spotten der Proteine konnten durchschnittlich 95% der His-getagten Proteine auf den Microarrays mit Hilfe eines anti-RGS His₆-Antikörpers nachgewiesen werden. Dieses Ergebnis belegt, neben den ermittelten Proteinkonzentrationen unter Verwendung des Bradford-Test (Kap. 4.2), dass die Hochdurchsatz-Proteinaufreinigung effizient genug war, um die gewonnen Proteine auf Microarrays nachzuweisen.

4.2 Phosphorylierungsstudien als Funktionelle Analyse

Ausgehend von den aufgereinigten Proteinen des Erweiterten Uniklonsets erfolgten funktionelle Analysen der Proteine unter Verwendung der Microarray-Technologie. Dazu wurden die Proteine des Erweiterten Uniklonsets mittels eines Microarrayers auf Nitrozellulose ähnliche FASTTM-Slides gespottet. Anschließend sollte die Phosphorylierung der gespotteten Proteine mit ausgewählten Kinasen untersucht werden.

4.2.1 Vergleich von Phosphorylierungsstudien

Beim Vergleich der bisher entwickelten Protein-Microarray-basierten Phosphorylierungs-Assays mit dem im Rahmen dieser Arbeit entwickelten Assay, konnten einige Unterschiede festgestellt werden. Diese Unterschiede beziehen sich auf die Oberflächen der Microarrays, die Anzahl der gespotteten Proteine und die Detektion der phosphorylierten Targets.

(Zhu *et al.*, 2000) verwendeten für die Phosphorylierungs-Assays PDMS-beschichtete *nanowell*-Microarrays. Bei diesem Assay wurden 119 Hefe Proteinkinasen und 17 verschiedene Substrate eingesetzt. Die in den *nanowells* kovalent gebundenen Substrate wurden unter Verwendung der Hefe Kinasen und γ -ATP (³³P) phosphoryliert. Sie konnten, unter Verwendung eines Phosphorimager zeigen, dass 32 der 119 Kinasen ein bis zwei der 17 Substrate phosphorylieren (26,9%).

MacBeath und Schreiber 2000 führten ihre Phosphorylierungsstudien auf BSA-N-hydroxysuccinimid (BSA-NHS) beschichteten Microarrays durch. Dabei wurden drei unterschiedliche Kinase-Substrat-Interaktionspaare untersucht, welche jeweils in vier identischen Spots pro Microarray vorlagen. Anschließend wurde jeder dieser Microarrays mit einer der drei Kinasen und γ -ATP (³³P) phosphoryliert. Dabei konnten sie zeigen, dass jeweils nur die spezifischen Substrate der entsprechenden Kinasen phosphoryliert wurden. Zur Detektion der phosphorylierten Substrate verwendeten MacBeath und Schreiber 2000 eine Photoemulsion, in welche die Microarrays gegeben wurden. Anschließend wurden die Microarrays wie ein Foto entwickelt. Diese Microarrays wurden dann mit einem Lichtmikroskop in Teilbereichen aufgenommen, und diese später zusammengefügt. MacBeath und Schreiber 2000 diskutieren diese recht aufwendige und nicht für eine Untersuchung im Hochdurchsatz geeignete Detektion, damit dass zu dem damaligen Zeitpunkt eine ausreichende Auflösung weder mittels Röntgenfilm noch mittels Phosphorimager erzielt werden konnte.

Bei den Phosphorylierungsstudien von Kramer *et al.* 2004 wurden 768 rekombinante Gerste-Proteine auf FASTTM-Slides gespottet und anschließend mit Gerste Casine Kinase2 α

phosphoryliert. Dabei wurden 384 Proteine jeweils vierfach in einem 10 x 10-Pattern auf einen Microarray transferiert. Bei der Detektion der beiden Microarrays mittels Röntgenfilm konnten 21 potentielle Targets identifiziert werden (2,7%).

Im Rahmen der vorliegenden Arbeit wurden 1.694 *Arabidopsis*-Proteine FAST™-Slides gespottet. Die Proteine wurden in einem 11 x 11-Pattern gespottet. Dabei wurden auf jeden Microarray zwei identische Felder gespottet, wobei sich in jedem Feld ein Spot pro Protein befand. Die phosphorylierten Proteine wurden anschließend unter Verwendung eines Phosphorimagers (Bioimager FLA8000) detektiert. Diese Detektion wurde gewählt, da die Detektion mittels Phosphorimager 10 – 100 x empfindlicher als der Röntgenfilm ist, und daher eine wesentlich kürzere Expositionszeit benötigt wird bzw. auch schwächere Signale detektiert werden können. Zudem ist bei dieser Methode das System weniger schnell gesättigt, so dass sowohl sehr starke als auch sehr schwache Signale mit einer Exposition erfasst werden können. Außerdem geht der lineare Signalbereich über 5 Größenordnungen (100.000 : 1). Somit ist der Phosphorimager für Quantifizierungen besser geeignet als die Verwendung von Röntgenfilmen. Hiermit konnte eine der Aufgaben dieser Arbeit dahingehend gelöst werden, die Signalerfassung der phosphorylierten *Arabidopsis*-Proteine mit einem Phosphorimager durchzuführen. Diese Signalerfassung stellt die Grundlage des ebenfalls in dieser Arbeit entwickelten Quantifizierungssystems dar.

Die bisher beschriebene qualitative und leichter visuell zu detektierende Variante, die Duplikate eines Proteins nebeneinander zu spotten (Kramer *et al.*, 2004; MacBeath und Schreiber, 2000), konnte in der vorliegenden Arbeit aufgrund der Optimierung der Reproduzierbarkeit und Quantifizierung, nicht beibehalten werden. In der vorliegenden Arbeit wurden die *Arabidopsis*-Proteine zunächst in einem dichten Pattern (14 x 14 und 15 x 15) auch jeweils vierfach in ein Feld gespottet, dabei zeigte sich bei intensiven radioaktiven Signalen ein Überstrahlungseffekt. Dieser hatte zur Folge, dass Spots mit intensiven Signalen nicht quantifiziert werden konnten. Um diesen Effekt zu vermeiden, wurden die *Arabidopsis*-Proteine in den weiteren Versuchen zwar in Duplikaten, jedoch räumlich voneinander getrennt auf die Microarrays gespottet. Des Weiteren wurde als Spotdichte ein 11 x 11-Pattern festgelegt. Mit diesem verwendeten Spotpattern ist eine parallele Analyse von 1.728 Proteinen (3.456 Spots, ohne *guide dots* und Kontrollen) pro Microarray möglich. Im Gegensatz zu Kramer *et al.* 2004, hier wurden 384 Proteine auf einen Microarray gespottet und phosphoryliert, ist somit eine effektivere Identifikation von potentiellen Phosphorylierungs-Targets möglich.

Alle oben beschriebenen Phosphorylierungs-Detektionen erfolgten unter Verwendung von radioaktiv markiertem ATP. Da diese Methode immer noch die einfachste und sensitivste Detektions-Methode ist, jedoch gesundheitliche Risiken mit sich bringt, werden Detektions-Alternativen gesucht. So verwendeten Houseman *et al.* 2002 zur Detektion von phosphorylierten Substraten auf den Microarrays einen anti-Phosphotyrosin-Antikörper, welcher anschließend mit einem zweiten Fluoreszenz markierten Antikörper inkubiert und detektiert wurde. Auch Lesacherre *et al.* 2002 setzen Fluoreszenz markierte anti-Phosphoserin- und anti-Phosphotyrosin-Antikörper zur Detektion von phosphorylierten Peptiden ein.

Vorteile der Detektion phosphorylierter Substrate mittels Fluoreszenz sind die Einschränkung der gesundheitlichen Risiken und die Möglichkeit, herkömmliche Microarrayscanner zur Auswertung von Phosphorylierungsstudien einzusetzen. Jedoch ist diese Detektions-Methode noch in der Entwicklung und noch nicht mit radioaktiv markierter Detektion zu vergleichen. Des Weiteren muss bedacht werden, dass bei der Detektion mittels Antikörper nur die Targets ermittelt werden können, die an einem bestimmten Epitop phosphoryliert werden. Somit ist diese Antikörpermethode zwar zur Identifizierung potentieller Phosphorylierungs-Targets mit bekannten Epitopen geeignet, aber nicht als *screening tool* zur Erfassung möglichst vieler Phosphorylierungs-Targets.

Um die Spezifität von Kinasen zu ermitteln, können auch Peptid-Microarrays verwendet werden (Falsey *et al.*, 2001; Houseman *et al.*, 2002; Lesacherre *et al.*, 2002; Lizcano *et al.*, 2002).

Lizcano *et al.* 2002 untersuchten die Substratspezifität der humane Serin/Threonin Proteinkinase Nek6 unter Verwendung von Peptid-Microarrays. Bekannt ist, dass diese Kinase *in vivo* die hydrophoben Motive der 3 N-terminalen Leuzin-Reste der S6K (p70 *ribosomal S6 kinase*) und SGK (*serum- and glucocorticoid-induced protein kinase*) phosphoryliert. Auf den Peptid-Arrays mit 710 Peptiden, mit bekannten Phosphorylierungsstellen humaner Proteine, konnten die Phosphorylierung von 7 Peptiden durch die Nek6 gezeigt werden.

Houseman *et al.* 2002 charakterisierten die Spezifität der c-Scr (*nonreceptor tyrosin kinase*), indem sie drei unterschiedliche Peptide auf Microarrays transferierten, wobei ein Peptid ein bekanntes Substrat der c-Scr darstellt. Unter Verwendung eines spezifischen Phosphotyrosin-Antikörpers konnte die spezifische Phosphorylierung des Substrat-Peptids gezeigt werden.

Ein Vorteil der von uns verwendeten Protein-Arrays im Vergleich zu der Peptid-Arrays ist, dass keine Kenntnisse über die Sequenzinformation der Targetsequenzen vorhanden sein müssen. Darüber hinaus sind die Arrays mit rekombinaten Proteinen nicht so teuer wie die Arrays mit Peptiden. Des weiteren sind Protein-Arrays universeller einsetzbar, da diese auch für andere funktionelle Studien, wie zur Untersuchung der Wechselwirkung der Proteine mit anderen Molekülen wie Proteine, DNA und kleine Moleküle, verwendet werden können.

4.2.2 Phosphorylierungsstudien mit den Proteinen des Erweiterten Uniklonsets

Für die in dieser Arbeit durchgeführten Phosphorylierungsstudien wurden die Proteinkinasen PKA, MPK3 und MPK6 gewählt.

Für die Phosphorylierungsstudien wurde zunächst der von Kramer *et al.* 2004 etablierte Microarray-basierte Phosphorylierungs-Assay weiterentwickelt, wobei unter Verwendung der Proteinkinase A ein Quantifizierungssystem etabliert wurde. Anschließend erfolgte die Identifizierung potentieller Phosphorylierungs-Targets der *Arabidopsis* Mitogen-aktivierten Proteinkinasen (MPK3 und MPK6) ausgehend von allen gespotteten Proteinen des Erweiterten Uniklonsets. Zu jedem Phosphorylierungs-Versuch wurde generell ein Kontroll-Microarray mitgeführt. Dieser Phosphorylierungsansatz enthielt γ -ATP (^{33}P), aber keine Kinase. Da in keinem dieser Kontroll-Microarrays Signale detektiert wurden, kann davon ausgegangen werden, dass unter den Versuchsbedingungen keine Bindung des γ -ATP (^{33}P) an die gespotteten Proteine bzw. keine Autophosphorylierung erfolgte.

Für alle drei Kinasen wurde je ein Set zusammengestellt (PKA-Set, MPK3-Set und MPK6-Set), welches aus den identifizierten potentiellen Phosphorylierungs-Targets der entsprechenden Kinasen bestand.

4.2.2.1 Quantifizierung mittels Proteinkinase A

In den Phosphorylierungsstudien unter Verwendung der Proteinkinase A, welche zur Etablierung des Quantifizierungssystems eingesetzt wurden, konnten 35 potentielle Phosphorylierungs-Targets ermittelt werden. Zur Bestätigung wurden diese erneut kultiviert, aufgereinigt, auf eine PVDF-Membran transferiert und dort erneut phosphoryliert. Von 35 potentiellen Phosphorylierungs-Targets konnten 28 bestätigt werden. Dabei können falschpositive Proteine ausgeschlossen werden, da alle im Coomassie-Brilliant-Blau gefärbten Gel identifizierten Protein phosphoryliert wurden, d.h. jedes im Coomassie-Brilliant-Blau gefärbten Gel visuell identifizierte Proteine wurde phosphoryliert. Für die Proteine, die weder im Coomassie-Brilliant-Blau gefärbten Gel noch im phosphorylierten Blot detektiert werden konnte, kann keine Aussage gemacht werden. Mögliche Ursachen

dafür, dass 7 Targets nicht visuell bestätigt werden konnten, kann einerseits sein, dass die radioaktiven Signale der phosphorylierten Proteine auf dem Blot zu gering waren, um detektiert zu werden. Andererseits kann es sein, dass die Proteine für diese erneute Aufreinigung entweder nicht ausreichend exprimiert wurden, die Aufreinigung nicht effizient war oder dass die aufgetragene Proteinmenge nicht ausreichend war. Ein weiterer Hinweis für die Richtigkeit der identifizierten Targets ist, dass alle im Coomassie-Brilliant-Blau gefärbten Gel sichtbaren Banden der Negativkontrollen keine Signale im Radiogramm zeigten. Dabei ist zu beachten, dass die Negativkontrollen aufgrund der für den Phosphorylierungs-Assay gewählten Quantifizierungsmethoden als negativ gewertet wurden.

Zusätzlich wurde die Qualität der identifizierten Targets durch den Vergleich der sich in dem PKA- und in dem Uniklonset befindlichen Sequenzen kontrolliert. Dabei wurden sowohl eine Motivsuche von vier bekannten Konsensussequenzen (Kallus, 2001) durchgeführt als auch eine Kontrolle der Serin- und Threoninegehalte. Bei der Motivsuche konnte für alle vier Konsensussequenzen eine Anreicherung der entsprechenden Motivgehalte im PKA-Set ermittelt werden. Diese Anreicherung zeigt, dass die Identifizierung der potentiellen Targets generell möglich ist. Jedoch zeigte sich auch, dass jede der vier Konsensussequenzen häufiger als 35 x (Anzahl der Targets im PKA-Set) im Uniklonset vorkommen, d.h. nicht jedes Protein, welches eines dieser Motive aufwies, wurde phosphoryliert. Das kann darauf zurückzuführen sein, dass lediglich das Vorliegen des Motivs nicht ausreichend für die jeweilige Phosphorylierung ist. Es ist aber auch nicht auszuschließen, dass das gewählte quantitative Kriterium zur Charakterisierung eines Proteins als Target (nur die Proteine, deren mittlere Signalintensität über der 10fachen Standardabweichung des mittleren Hintergrundsignals aller Unterfelder lagen, wurden als potentielle Phosphorylierung-Targets identifiziert) evtl. zu streng ist. Da es jedoch Ziel dieser Arbeit war, ein *screening tool* zu etablieren, welches zu einer quantifizierten Eingrenzung der Anzahl von potentiellen Phosphorylierungs-Targets dienen sollte, die anschließend unabhängig verifiziert werden sollten, wurde dieses strenge Kriterium beibehalten. Bei dieser Vorgehensweise kann man annehmen, dass auch der Anteil an falschpositiven Targets eingeschränkt wird. In weiteren vergleichenden Studien sollte man versuchen, für dieses Kriterium den exakten Grenzwert zu ermitteln, wodurch es zu einer Maximierung der Selektion aller sich im Set befindlichen potentiellen Targets kommen könnte.

Der Vergleich des Serin- und Threoningehaltes im PKA- und Uniklonset zeigte eine Anreicherung des Seringehaltes, was als Hinweis darauf angesehen werden könnte, dass die Proteinkinase A zwar zu Serin/Threonin-Proteinkinase gehört, jedoch die Serin-Reste vielleicht bevorzugt phosphoryliert.

4.2.2.2 Phosphorylierungsstudien unter Verwendung der *Arabidopsis* Kinasen MPK3 und MPK6

Wie bereits in der Einleitung erläutert, gehören die MAP-Kinasen zu den Serin/Threonin-Proteinkinasen. Beide MAP-Kinasen (MPK) MPK3 und MPK6 werden in eine der vier Gruppen (Gruppe A), in welche die MPKs unterteilt werden, eingeordnet (vgl. 1.5.1; (MAPKgroup, 2002). Die MPKs sind Bestandteile von Signaltransduktionskaskaden, den MAPK-Kaskaden. Diese Kaskaden bestehen aus MAPKK-Kinasen (MAPKKK), MAPK-Kinasen (MAPKK) und MAP-Kinasen (MPK) und sind an der Weiterleitung der durch Rezeptoren oder Sensoren erkannten externen Stimuli von der Zelloberfläche zum Zellkern beteiligt. Im Zellkern kann es dann zur Phosphorylierung von Transkriptionsfaktoren kommen. Die Weiterleitung des Stimulus erfolgte über sogenannte *second messenger*, durch die es u. a. zu Änderungen in der Ionenpermeabilität der Plasmamembran (Nünberger und Scheel, 2001) oder zu Akkumulationen von reaktiven Sauerstoffspezies (ROS, z.B. Superoxidanionen, Hydroxylradikale oder Wasserstoffperoxid) kommen kann (Neill *et al.*, 2002). Die ROS können direkt toxisch auf das Phytopathogen wirken oder in die Verstärkung der pflanzlichen Zellwand involviert sein (Lamb und Dixon, 1997).

Durch diese Weiterleitung kommt es zur Aktivierung von Abwehrmechanismen (Zwerger und Hirt, 2001). Solche Kaskaden ermöglichen es den Pflanzen, auf verschiedene biotische und abiotische Stressfaktoren zu reagieren. Eine Beteiligung von MAP-Kinasen an der pflanzlichen Signaltransduktion konnte für eine große Anzahl von Stimuli beschrieben werden. Beispielsweise für mechanischen und osmotischen Stress (Bogre *et al.*, 1996; Ichimura *et al.*, 2000; Mizoguchi *et al.*, 1996; Tena und Renaudin, 1998), Trockenheit und Kälte (Ichimura *et al.*, 2000; Jonak *et al.*, 1996; Mizoguchi *et al.*, 1996), sowie Elicitor- als auch Pathogenkontakt (Asai *et al.*, 2002; Nühse *et al.*, 2000; Romeis *et al.*, 1999; Zhang *et al.*, 1998). Des Weiteren sind sie auch bei der Weitergabe der Pflanzenhormone, wie Ethylen, Abszissinsäure und Auxin beteiligt (Jonak *et al.*, 2002; Lu *et al.*, 2002; Tena *et al.*, 2001).

Bei den in der vorliegenden Arbeit durchgeführten Phosphorylierungsstudien konnten für das MPK3-Set 48 potentielle Phosphorylierungs-Targets und für das MPK6-Set 39 potentielle Phosphorylierungs-Targets identifiziert werden. Die geringere Anzahl an

identifizierten Targets bei dem MPK6-Set kann einerseits darauf zurückzuführen sein, dass die Kinase weniger Targets phosphoryliert oder für die MPK6 weniger potentielle Phosphorylierungs-Targets im Erweiterten Unikonset vorlagen, als für die MPK3. Andererseits bleibt zu bedenken, dass die rekombinate MPK6 mit einer geringeren Aktivität vorlag (persönliche Mitteilung unserer Kooperationspartner Dr. Justin Lee und Prof. Dr. Dierk Scheel), daher zwar eine höhere Konzentration eingesetzt wurde, diese aber vielleicht nicht ausreichend war, um eine zur MPK3 vergleichbare Aktivität zu erreichen. Es kann aber auch sein, dass die in der vorliegenden Arbeit verwendeten Assay-Bedingungen (wie z. B. Kofaktoren) für die MPK6 weniger optimal waren, als für die MPK3, da die Bedingungen ausgehend vom Myelin-assoziiertes basisches Protein (MBP), einem artifiziellen Substrat der Kinasen, optimiert wurden.

Zur Bestätigung der identifizierten Targets in einem unabhängigen *in vitro* Testsystem wurden die 48 potentiellen Phosphorylierungs-Targets des MPK3-Sets erneut aufgereinigt, auf PVDF-Membranen transferiert und unter Verwendung der MPK3 phosphoryliert. Dies war leider nur für die MPK3 möglich, da zu diesem Zeitpunkt keine ausreichend aktive MPK6 mehr zugänglich war. Von 48 potentiellen Phosphorylierungs-Targets konnten 45 bestätigt werden. Mögliche Ursachen für die drei nicht identifizierten Targets sind wie bei dem Phosphorylierungs-Blot der Proteinkinase A zu diskutieren.

Bei den Literaturrecherchen zu den identifizierten potentiellen Phosphorylierungs-Targets zeigte sich, dass für die *Arabidopsis* MPK noch nicht bekannt ist, welche Targetproteine direkt phosphoryliert werden. Bisher gibt es nur Untersuchungen der durch Stress ausgelösten MAP-Kinasen-Signaltransduktionskaskade hinsichtlich bestimmter Gentranskripte, aus denen man erste Hinweis auf mögliche Targets gewinnen kann (z.B. Kimura *et al.*, 2003; Kreps *et al.*, 2002; Seki *et al.*, 2002).

Beispielsweise konnten sowohl die Aktivierung von MPK3 durch Trockenheit, Kälte, Salzstress und mechanische Beanspruchung (Mizoguchi *et al.*, 1996) als auch eine posttranslationale Aktivierung von MPK3 durch H₂O₂ beobachtet werden.

Des Weiteren konnte für die Transkripte der MEKK1, MPK3 und MPK1 eine Induktion durch osmotischen Stress und mechanische Beanspruchung gezeigt werden. Zusätzlich sind die Transkripte der MEKK1 und MPK3 durch Kälte induzierbar [Mizoguchi *et al.*, 1996]. Darüber hinaus konnten die Transkripte der MPK3 und MPK4 durch Bestrahlung mit UV-Licht induziert werden (Ulm *et al.*, 2002). Teige *et al.* 2004 konnten sowohl eine Aktivierung der MKK2 durch Kälte, Salzstress oder durch die stressinduzierte MEKK1 als auch die davon ausgehende Induktion der MAPK4 und MAPK6 zeigen.

Zudem konnte eine Aktivierung der MAP-Kinasen nicht nur bei abiotischen sondern auch bei den biotischen Stressfaktoren beobachtet werden. Asai *et al.* 2002 beschrieben bei *Arabidopsis* die vollständige MAP-Kinasen-Signaltransduktionskaskade, welche durch den bakteriellen Elicitor Flg22 ausgelöst wird. Diese MAPK-Kaskade besteht aus den Komponenten MEKK1, MKK4/5 und MPK3/6, welche die transkriptionelle Aktivierung der WRKY-Transkriptionsfaktoren WRY22/29 induzieren. Die Übertragung des durch den Elicitor Flg22 ausgelösten Signals erfolgt über FLS2 (*flagellin sensitive 2*), eine LRR (*leucin-rich repeat*) Rezeptor ähnliche Kinase (Gómez-Gómez und Boller, 2000).

Neben dem eben beschriebenen bakteriellen Elicitor bei *Arabidopsis* wurden für die MPK6 Orthologen in *Medicago sativa*, SIMK (*stress induced MAPK*), und in Tabak, SIPK (*salicylate induced MAPK*), weitere Elicitoren beschrieben, welche zur Aktivierung dieser Kinasen führen. Die SIMK-Aktivierung erfolgt durch unterschiedliche pilzliche Elicitoren (Cardinale *et al.*, 2000). Für SIPK konnte eine Aktivierung durch allgemeine Elicitoren (Droillard *et al.*, 2000; Lee *et al.*, 2001) und durch TMV (Tabak-Mosaik-Virus)-Infektion (Zhang *et al.*, 1998) beschrieben werden. Auch bei Petersilie konnte eine Aktivierung der PcMPK6 durch die Elicitoren PEP-13 und HrpZ beobachtet werden (Kroj *et al.*, 2003).

Wan *et al.* 2004 vermutet eine Überlappung von der von Asai *et al.* 2002 beschriebenen Signaltransduktionskaskaden mit der Signaltransduktion von Chitin Elicitoren, welche bei der Übertragung des Chitin-Stimulus von den Zellwänden pathogener Pilze und Insekten ausgelöst wird.

Einige in dieser Arbeit identifizierten potentiellen Phosphorylierungs-Targets stehen im Zusammenhang mit bereits beschriebenen Komponenten der MAP-Kaskaden. Aus diesem Grund ist ihre Phosphorylierung durch die MPKs nicht unwahrscheinlich.

Das als potentielles Phosphorylierungs-Target der MPK3 identifizierte Protein: *leucine-rich repeat family protein contains leucine rich repeat (LRR) domains* (At4g03260; exprimiert von Klon: 312_J15 des Uniklonsets) könnte, aufgrund seiner LRR-Domäne, an einer von der Elicitoren ausgelösten-Signaltransduktionskaskade beteiligt sein (Asai *et al.*, 2002; Gómez-Gómez und Boller, 2000; Wan *et al.*, 2004). Als ein weiteres interessantes Target konnte das durch Klon 313_G13 exprimierte Protein (At4g11280) bei Phosphorylierung mit der MPK6 detektiert werden. Hierbei handelt es sich um die ACC Synthase 6 (*1-aminocyclopropane-1-carboxylate synthase 6*). Ouaked *et al.* 2003 konnten zeigen, dass eine MAPK-Kaskade an der Ethylen-Signaltransduktion beteiligt ist, in dem bestimmte MPKs in *Arabidopsis* und *Medicago* durch die Vorläufersubstanz des Ethylens ACC aktiviert werden.

Dieses sind die aktivierten MPKs SIMK (*stress induced MAPK*) und MMK (*Medicago MAPK2*) in *Medicago* und in *Arabidopsis* die MPK6 sowie eine weitere MAP-Kinase.

Des weiteren wurden die Histonen H2A (im MPK3-Set) und H3 (in beiden Sets) als potentielle Phosphorylierungs-Targets identifiziert. Für Histone ist bekannt, dass diese als Reaktion auf extrazelluläre Signale phosphoryliert werden. Yamagata *et al.* 2001 zeigten eine Phosphorylierung von Histonen durch die *MAPK kinase-like protein kinase* in Zellkultur von Sojabohnen. Ob die Histone tatsächlich in der Signaltransduktion der MPKs involviert sind, bleibt noch zu zeigen.

Auch ein Protein des MPK3-Sets (At5g62070; *calmodulin-binding family protein*; exprimiert durch Klon 313_F05) kann ein Bestandteil der MAPK-Kaskaden sein. Perruc *et al.* 2004 zeigten, dass bei *Arabidopsis* ein Calmodulin bindendes Protein (At2g41010) als Negativregulator bei der Stresstoleranz gegenüber Salz- und osmotischem Stress beteiligt ist. Ein eventueller Zusammenhang zwischen einer möglichen Phosphorylierung von Calmodulin-bindenden Proteinen durch MPKs und deren Rolle in der Regulation der Stresstoleranz muss noch bestätigt werden. Calmodulin könnte über die Bindung an das Calmodulin-bindende Protein oder aber auch über die Aktivierung der Ca²⁺-calmodulinabhängigen Proteinkinase (CaM-Kinase) einen Einfluss auf die Stresstoleranz haben. CaM-Kinasen können z.B. Serin- und Threoninreste in ihren Zielproteinen phosphorylieren und damit eine Enzymaktivierung bewirken. Daran könnten ebenfalls auch die MPKs beteiligt sein.

Die Lokalisierung der aktivierten MPKs im Zellkern läßt Transkriptionsfaktoren als Targets vermuten. Diese Vermutung wird auch durch die in dieser Arbeit als Phosphorylierungs-Targets gefundenen Transkriptionsfaktoren bestätigt. So konnten folgende Transkriptionsfaktoren identifiziert werden: Der Klon des MPK3-Sets 315_G09 (At5g66940; *Dof-type zinc finger domain-containing protein*), sowie die Klone des MPK6-Sets 315_P07 (At1g06170; *basic helix-loop-helix (bHLH) family protein*), 314_N21 (At3g60390; *homeobox-leucine zipper protein HAT3*), 313_N06 (At5g54630; *zinc finger protein-related*) und 311_F11 (At5g58620; *zinc finger (CCCH-type)*), wobei letzterer zusätzlich auch im MPK3-Set vorkommt und mit seinem Zinkfingermotiv eine DNA-bindende Domäne aufweist.

Wie sich anhand der Histone und der Transkriptionsfaktoren zeigte, wurden auch gemeinsame potentielle Phosphorylierungs-Targets für beide MPKs identifiziert. Insgesamt waren das 26 potentielle Phosphorylierungs-Targets. Dieses wurde jedoch auch vermutet, da beide MPKs zu einer Untergruppe einer Proteinklasse gehören.

Als weiteres gemeinsames potentiell Phosphorylierungs-Target wurde z. B. die Casein Kinase (CK; 314_G10; At5g44100) ermittelt. Eine direkte Beteiligung der Casein Kinase in der Signaltransduktion der MAPK-Kaskaden ist für Pflanzen noch nicht beschrieben. In HeLa-Zellen wurde jedoch eine direkte Wechselwirkung der p38 MAP-Kinase mit der CK2 und deren stressinduzierten Aktivierung durch diese MAP-Kinase gezeigt (Sayed *et al.*, 2000).

Ein weiterer Klon, der in beiden Sets zu finden war, ist der Klon: 313_E14 (At5g48990; *kelch repeat-containing F-box family protein*). F-Box Proteine sind häufig an der Degradation von cytosolischen Proteinen beteiligt (Andrade *et al.*, 2001). F-Box Proteine erkennen und binden die abzubauenen Proteine, welche häufig phosphoryliert sind (Willems *et al.*, 1999). Das am F-Box-Protein gebundene Protein wird anschließend mehrfach ubiquitiniert und zum Proteasomen transportiert. Diese Polyubiquitinierung dient als Degradationssignal für das Proteasomen. Ob die Phosphorylierung der F-Box-Proteine *in vivo* eine Rolle spielt, bleibt noch zu klären.

Als eine weitere Kontrolle hinsichtlich der Spezifität der identifizierten Targets der jeweiligen Kinasen wurden neben den gemeinsamen Phosphorylierungs-Targets der beiden MPKs (26 insgesamt) auch die gemeinsamen Targets der PKA mit denen der MPK3 bzw. der MPK6 bestimmt. Für die MPK3 wurden drei und für die MPK6 vier gemeinsame Targets mit der PKA ermittelt. Wie zu erwarten war, lieferte der Vergleich der unterschiedlichen Serin/Threonin-Proteinkinasen einen niedrigeren Anteil an gemeinsamen Targets, als der Vergleich der beiden MAP-Kinasen, die zu einer Kinasegruppe gehören.

Da alle drei Kinasen zu den Serin/Threonin-Proteinkinasen gehören, liegt die Vermutung nahe, dass bei den ermittelten gemeinsamen potentiellen Phosphorylierungs-Targets beider Gruppen ausschließlich Serin- und Threonin-Reste phosphoryliert wurden, was aber in der vorliegenden Arbeit nicht weiter untersucht wurde. Die genaue Ermittlung der phosphorylierten Aminosäure ist prinzipiell entweder mit Massenspektroskopie (Glinski *et al.*, 2003) oder mit phospho-spezifischen Antikörpern (Houseman *et al.*, 2002; Lesaichere *et al.*, 2002) möglich.

Die Ergebnisse der Phosphorylierungsstudien zeigen, dass sowohl der etablierte Kinase-Assay als auch das dazu gehörende Quantifizierungssystem reproduzierbar und spezifisch sind.

Des Weiteren sind alle durchgeführten Kontrollen ein Hinweis darauf, dass alle Schritte, die zur Entstehung des Uniklonsets (Detektion der potentiellen Expressionsklone mittels Hochdichte Proteinfiltern, die Sequenzanalyse und davon ausgehend die Erstellung des

Uniklonsets) erfolgreich waren. Außerdem belegen sowohl die Kontrollen als auch die Ergebnisse dieser Phosphorylierungsstudien, dass die im Rahmen dieser Arbeit verwendete Aufreinigungstechnik für den Hochdurchsatz geeignet war.

An dieser Stelle soll nochmals betont werden, dass es sich bei dem hier weiterentwickelten Mikroarray-basierten Phosphorylierungssystem nur um ein *screening tool* handelt, welches zu einer *in vitro* Identifizierung der potentiellen Kinasetargets dient. Die mit diesem Testsystem identifizierten potentiellen Phosphorylierung-Targets müssen jedoch in anschließenden Folgeversuchen *in vivo* weiter verifiziert werden. Das ist aus folgenden Gründen erforderlich: In der vorliegenden Arbeit wurden die Targets unter denaturierenden Bedingungen aufgereinigt. Es kann also nicht ausgeschlossen werden, dass unter nativen Bedingungen die Phosphorylierungsstellen für die Kinasen nicht zugänglich sind. Des Weiteren besteht auch die Möglichkeit, dass *in vivo* die hier identifizierten Targets nicht mit den Kinasen interagieren können, da sie in unterschiedlichen Zellkompartimenten lokalisiert sein könnten. Dennoch zeigt die Identifikation von bekannten Phosphorylierung-Targets, dass diese Methode als eine Möglichkeit zur Ermittlung von unbekanntem potentiellen Phosphorylierungs-Targets angesehen werden kann.

In vivo Methoden, die angewandt werden können, um die identifizierten potentiellen Phosphorylierungs-Targets zu verifizieren, sind z.B. Methoden zur Untersuchung von Protein-Protein Wechselwirkungen, z.B. in Hefe (Yeast-Two-Hybrid) oder in Pflanzen direkt. Hierbei kommt es zur Verwendung von mikrospektroskopischen Ansätzen wie Fluoreszenz-Resonanz-Energie-Transfer (FRET) und Fluoreszenz-Korrelations-Spektroskopie (FCS) (Hink *et al.*, 2002). Mittels FRET-Methoden kann eine abstandsabhängige Wechselwirkung zwischen zwei angeregten Farbstoffmolekülen detektiert werden, bei der die Anregungsenergie von einem Donormolekül auf eine Akzeptormolekül übertragen wird, ohne daß Photonen abgestrahlt werden. Bei FCS wird meist nur eine Molekül farblich markiert. Bei dieser Methode kann man die Bewegung einzelner Moleküle beobachten, so bewegen sich z. B. gebundene Moleküle vergleichsweise langsamer.

Eine weitere *in vivo* Methode zur Bestätigung der potentiellen Phosphorylierungs-Targets stellt die Klonierung der potentiellen Targets in Epitop-getagte Pflanzenexpressionsvektoren dar. Dazu könnte das GATEWAY-System benutzt werden, um eine effiziente und schnelle Subklonierung zu ermöglichen. Anschließend werden pflanzliche Protoplasten mit diesen Plasmiden transformiert, und die Phosphorylierung dieser Targets kann durch Zugabe von radioaktiv markiertem Phosphat zum Medium untersucht werden. Bedingungen, die zu einer

Aktivierung der MAP-Kinase führen (Elicitor, abiotische Stressbehandlung oder Kotransfektion in eine aktive *upstream* MAP Kinase Kinase) können genutzt werden, um die Beteiligung von MAP-Kinase bei der Phosphorylierung der potentiellen Targets zu untersuchen.

Bei der Validierung von Phosphorylierungs-Targets mittels Protein-Protein-Interaktionsstudien bleibt aber zu beachten, dass die Wechselwirkung einer Kinase mit einem Protein nicht zur Phosphorylierung des Proteins führen muß.

Die in der Arbeit hergestellten Klone stellen die bisher umfangreichste Sammlung von rekombinaten *Arabidopsis thaliana* Expressionsklonen dar und stehen für weitere Studien zur Verfügung.

Diese Arbeit, ist die erste systematische Studie zur Identifizierung von Phosphorylierungs-Targets der MAP-Kinasen im Hochdurchsatz. Zur Identifizierung wurde ein Quantifizierungssystem entwickelt, bei dem nur phosphorylierten Signale, die sich deutlich vom Hintergrund unterschieden, verwendet wurden, um die dazugehörigen Proteine als Phosphorylierungs-Targets zu identifizieren. Eine Lockerung des in der vorliegenden Arbeit sehr streng verwendeten *cutoff* Wertes für die Targetselektion, könnte die Anzahl der identifizierten Phosphorylierungs-Targets in Zukunft evtl. erhöhen.

Desweiteren konnten in der vorliegenden Arbeit bisher noch nicht beschriebene Targets der MAP-Kinasen identifiziert werden. Diese *in vitro* identifizierten Phosphorylierungs-Targets liefern eine interessante Grundlage für weitere *in vivo* Untersuchungen. Der Phosphorylierungs-Assay ist für verschiedene Kinasen anwendbar und kann in Zukunft zur Identifizierung von Phosphorylierungs-Targets weiterer *Arabidopsis thaliana* Kinasen verwendet werden.