6. Diskussion

Die vorliegende Arbeit erfolgte im Rahmen eines Projektes zur Aufklärung von Mechanismen der Neuroprotektion und Ischämietoleranz bei cerebraler Ischämie.

An einem in-vivo Ischämie-Modell der Maus waren zwei SAGE-cDNA-Expressionsbibliotheken (Kontrolle, 14 h nach Reperfusion nach 2 h fokaler cerebraler Ischämie) erstellt worden, die die Grundlage einer Rangliste differentiell exprimerter SAGE-*tags* (Gene) nach fokaler cerebraler Ischämie bildeten.

Das Ziel dieser Arbeit war die Bestätigung potentiell neuer bzw. bisher wenig charakterisierter differentiell exprimierter Gene der Ischämie-SAGE-cDNA-Expressionsbibliothek, sowie die weitere molekularbiologische Charakterisierung verifizierter Transkripte.

6.1. Aspekte der SAGE-*tag*-Gen-Zuordnung

Die Verifizierung von SAGE-Genexpressionsdaten durch Methoden wie Northern Blot, RT-PCR oder in situ Hybridisierung erfordert die vorherige korrekte Zuordnung des jeweiligen SAGE-*tags* zu dem korrespondierenden Transkript bzw. Gen. Die korrekte Identifikation der jeweiligen Transkripte ist von entscheidender Relevanz für spätere funktionelle Studien von detektierten differentiell exprimierten Genen.

Die Zuordnung der hier untersuchten Kandidaten SAGE-*tags* erfolgte initial über eine Referenzdatenbank mit *tags* aus Maus Unigene-Clustern (file, *SAGEmap tag ug-rel-Nla3-Mm.txt,* hergeleitet aus *Mus musculus Unigene Build #96, Lal et al 1999*) sowie manuell durch Sequenz-Homologierecherche in den verfügbaren EST- und Gendatenbanken mit Hilfe des BLAST-Algorithmus. Obwohl jedes spezifische SAGE-*tag* theoretisch genügend Informationen enthält, um das zugrunde liegende Transkript mit Hilfe der zur Verfügung stehenden Referenzsequenz-Datenbanken eindeutig zu identifizieren, ergaben sich dabei Schwierigkeiten. Diese erklären sich einerseits aus Mängeln der Referenzsequenz-Datenbanken und andererseits aus dem SAGE-*tag* selbst:

1. *geringe Spezifität der tags*

   Ein Teil der Kandidaten-SAGE-*tag*s ließ sich jeweils mehreren Unigene-Clustern, die unterschiedliche Gene repräsentieren, zuordnen (*multiple match tags*). Nach Literaturangaben trifft das für etwa 1/3 aller zuordenbaren *tag*s einer SAGE-Bibliothek zu. Besonders hoch (2/3) ist der Anteil in der Gruppe der *high copy tags* (*Lee et al 2002*).

Eine Untersuchung von Lee et al. (*2002*) ergab, dass sich in der Literatur 45 % aller vermeintlich eindeutig zugeordneten SAGE-*tags* bei Reevaluation doch multiplen Genclustern zuordnen lassen.

Vielfach wurde versucht, die Spezifität der *tags* durch Verlängerung der SAGE-*tag*-Sequenz zu erhöhen (*Ryo et al 1998, Inoue et al 1999, Ryo et al 2000, Saha et al 2002, Matsumura et al 2003, 2005*). Andererseits konnte jedoch gezeigt werden, dass eine Verlängerung von 10 bp auf 35 bp die Spezifität nur um etwa 10 % erhöht (*Lee et al 2002*).

2. *unzureichende Qualität der Referenzsequenz-Datenbanken*

Die Spezifität der SAGE-*tags* hängt wesentlich von der Qualität der zur Verfügung stehenden Referenzsequenz-Datenbanken ab (*Lee et al 2002*). Diese enthalten jedoch nur einen Teil an Transkripten aus gut charakterisierten mRNA oder cDNA Sequenzen mit einer vollständigen 3`UTR, so dass vorwiegend EST-Sequenzen (*expressed sequence tag sequences*) als Referenz dienen müssen. Davon besitzen nur wenige ein eindeutiges PolyA-Signal mit einem eindeutigen PolyA-Sequenzanteil. Etwa 10-15 % der verfügbaren Datenbanksequenzen sollen - bedingt durch interne Anlagerung des Oligo(dT) Primers bei der Generierung der cDNA Bibliotheken - unvollständige bzw. artefizielle 3` Enden aufweisen (*Bonaldo et al 1996*).

Auch die angegebene Orientierung der ESTs stimmt zum Teil nicht, oder sie besitzen keine Orientierung. Besonders bei ESTs mit wenigen Sequenzeinträgen muss weiterhin mit Sequenzierungsfehlern gerechnet werden (*Bouck et al 1999*), in humanen ESTs z.B. etwa 1 % (*Hillier et al 1996*). Es gibt ferner ESTs, die durch fehlerhaftes Alignment einem falschen Unigene Cluster zugeordnet sind (*Bouck et al 1999*). Zusätzlich sind viele Unigene Cluster „überclustert“ (*Bowtell 1999*).

Diese Mängel führen dazu, dass bei der automatischen *tag*-zu-Gen Zuordnung (SAGEmap *„reliable“ tag-to-gene mapping*; http: //www.ncbi.nlm.nih.gov/sage) mit einer Fehlerrate von etwa 10 % gerechnet werden muss (*Lash et al 2000*). Gunnersen fand nach Überprüfung mittels manueller Zuordnung sogar eine Fehlerrate von 33 % (*Gunnersen et al 2002*).

In einer Untersuchung von Lee wurde demonstriert, dass die Spezifität der Transkriptzuordnung bei ausschließlicher Verwendung der zur Zeit noch limitierten gut charakterisierten „nur Gen-Sequenzen“ als Referenz wesentlich höher ist als bei Einbeziehung von EST-Sequenzen (44 % vs. 11 %) (*Lee et al 2002*).

3. *echte und artifizielle interne tags*

Besondere Schwierigkeiten in der Zuordnung und Transkriptidentifikation können interne *tags* bereiten. Echte interne *tags* können durch alternative Polyadenylierung, alternatives Splicen, Polymorphismus an der NlaIII-Schnittstelle (*Boon et al 2002*) oder bei internen mitochondrialen *tags* durch geringere Genauigkeit der Transkription (*Anisimov et al 2002a*) erklärt werden. Artifizielle interne, richtig orientierte SAGE-*tags* durch falsche Anlagerung des Oligo(dT)-Primers an internen A-reichen mRNA-Regionen bei der cDNA-Synthese (*de Waard et al 1999, Pauws et al 2000*) oder durch inkompletten NlaIII-Verdau (*Boon et al 2002*) wurden bereits in mehreren SAGE-Studien beschrieben. Daneben konnte bei der Analyse des *tags* NM03 ein weiterer Fehlermechanismus der SAGE-Prozedur aufgedeckt werden. Das *tag* NM03 liegt intern, revers orientiert vor einer T-reichen Region (der Maus Calmodulin-mRNA). Über die mögliche Entstehung fanden sich in der Literatur keine Angaben, sie könnte jedoch wie folgt erklärt werden (Abb. 23):

- Bei der Synthese des ersten cDNA-Stranges wird eine interne T-reiche mRNA-Sequenz in die komplementäre A-reiche Sequenz umgeschrieben.
- An diese lagert sich während der Synthese des zweiten cDNA-Stranges überschüssiger Oligo(dT)-Primer an.
- Nach dem NlaIII-Verdau der Doppelstrang-cDNA entsteht ein artifizielles 3`terminales cDNA Fragment, welches zur Generierung eines revers orientierten falschen SAGE-*tags* führt.

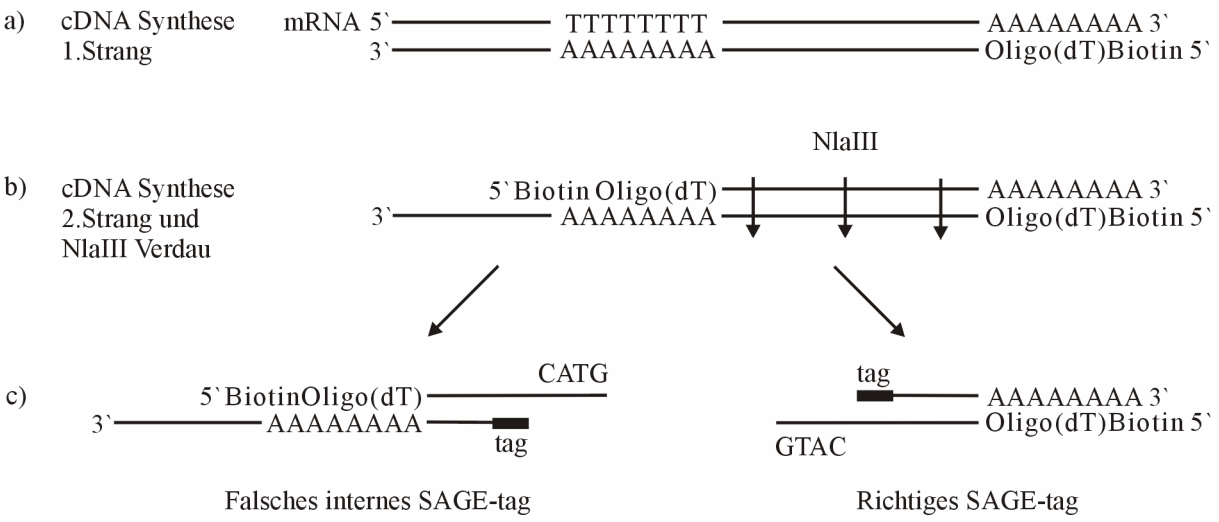

Abb. 23: Hypothese zur Entstehung interner revers orientierter SAGE-*tags* während der SAGE-Prozedur. Überschüssiger Oligo(dT)-Primer lagert sich während der Synthese des 2. cDNA-Stranges an eine interne A-reiche Sequenz an (a, b). Nach dem NlaIII-Verdau kann das artifizielle biotinylierte Fragment gebunden werden (c).

4. no match tags

SAGE-*tags*, die sich weder einem bekannten Gen, noch einem EST aus den Referenzsequenz-Datenbanken zuordnen lassen, werden als *no match tags* bezeichnet. Besonders hoch ist deren Anteil innerhalb der Fraktion niedrigexprimierter Gene und unter den nur einmal gezählten *tags* (*single copy tags*) (*Chen et al 2002a*).

No match tags werden in der Literatur sehr unterschiedlich interpretiert. Zum Teil wird angenommen, dass es sich hierbei vorwiegend um Artefakte durch Sequenzierfehler (bis zu 10 %) handelt (*Zhang et al 1997*). Besonders bei nur einmal gezählten *tags* (*single copy tags*), die sich nur in einer Base von bereits bekannten *tags* unterscheiden, wird dies vermutet. Chen und Mitarbeiter (*Chen et al 2002a*) fanden in ihren SAGE-Bibliotheken allerdings einen geringeren Sequenzierfehler (ca. 1,67 %). Sie konnten aus der Analyse von 101 SAGE-Bibliotheken weiterhin zeigen, dass die Chance, dass selbst ein einzeln gezähltes *tag* durch einen Sequenzierfehler entstanden ist, sehr gering ist. Und sogar bei einem Sequenzierfehler von 10 % beträgt die Wahrscheinlichkeit, dass ein durch einen Einzel-Basen-Sequenzierfehler (*single base error*) entstandenes *tag* ein zweites Mal entsteht, nur 1:329. Nach Untersuchungen von Chen repräsentiert ein Großteil der *no match tags* (knapp 70 %) neue Transkripte (*Chen et al 2002a*). Es wird davon ausgegangen, dass trotz der in den letzten Jahren drastisch gesunkenen Zahl neu identifizierter Gene durch das EST-Projekt (http://www.ncbi.nlm.nih.gov/ncigap/lib_report.html) weiterhin eine beachtliche Zahl von Genen existiert, die bisher nicht identifiziert wurden (*Hogenesch et al 2001, Saha et al*

2002). Dafür spricht auch, dass in vielen neueren humanen oder Maus-SAGE-Projekten immer noch eine große Anzahl von *no match* SAGE-*tags* detektiert wird (*Lee et al 2001, Zhou et al 2001*).

Im Gegensatz dazu berichten andere Autoren, dass sich viele der *no match tags* manuell durch Interspezies-Vergleich oder Verbindung von benachbarten Sequenzstrecken den 3` untranslatierten Regionen von bereits bekannten Genen zuordnen lassen (*Gunnersen et al 2002, Pauws et al 2000*).

No match tags können weiterhin von bekannten Genen stammen, welche aber vom Antisense-Strang der genomischen DNA aus transkribiert wurden (*Chen et al 2002a*), von alternativ geschnittenen Transkripten (*Neilson 2000*), von nicht kodierenden Transkripten oder Transkripten mit alternativen Polyadenylierungsschnittstellen (*Pauws et al 2001*).

6.2. SARA-PCR

Angesichts der oben genannten Schwierigkeiten bei der Transkriptidentifikation mit Hilfe von Referenzsequenz-Datenbanken wurde deshalb in dieser Arbeit eine direkte, Sequenzdatenbank-unabhängige Methode zur *tag*-spezifischen Identifikation korrespondierender Transkripte etabliert (SARA-PCR).

Diese beruht auf einer Amplifikation der SAGE-*tag*-spezifischen 3`Transkripte. In der Literatur werden mehrere Amplifikationsmethoden zur Identifizierung der korrespondierenden Transkripte beschrieben, welche im wesentlichen darauf basieren, dass ein *tag*-spezifischer 13-15 bp Vorwärtsprimer (CATG(NlaIII)- und SAGE-*tag*-Sequenz) zusammen mit einem 3`terminalen Minusprimer an einer z.T. M13-modifizierten, meist unverdauten Oligo(dT)cDNA eingesetzt wird.

Die SARA-PCR unterscheidet sich von diesen publizierten PCR-Methoden hinsichtlich cDNA-Modifikation und Primerkonstruktion.

6.2.1. cDNA-Synthese und cDNA-Modifikation

Anstelle eines Standard Oligo(dT)-Primers (*Peters et al 1999*) oder eines unmarkierten M13-Oligo(dT)-Primers (*Van den Berg et al 1999*), erfolgte die Synthese der SARA-PCR-cDNA mit einem Biotin-markierten M13-Oligo(dT)-Primer. Das ermöglichte nach dem Restriktionsverdau die Isolation der 3`terminalen cDNA-Fragmente und Entfernung der 5`terminalen ungebundenen Fragmente. Durch diese Reduktion der Komplexität der cDNA wurde eine entscheidende Erhöhung der Amplifikationsspezifität erreicht und eine Anlagerung des *tag*-spezifischen Plusprimers an falschen, nicht 3`terminal gelegenen

homologen Transkriptbereichen verhindert. Damit profitiert das PCR-Protokoll von der zusätzlichen Information die in jedem SAGE-*tag* enthalten ist: die Lage an der letzten 3`terminalen NlaIII-Erkennungssequenz der cDNA. Im Gegensatz dazu erfolgt bei anderen existierenden SAGE-*tag*-spezifischen PCR-Methoden (*Van den Berg et al 1999, Peters et al 1999, Chen et al 2000*) kein cDNA-Restriktionsverdau, so dass eine ausschließliche Amplifikation von 3`-terminalen Transkripten nicht garantiert werden kann.

Gegen Ende dieser Arbeit publizierte Chen (*Chen et al 2002*) eine Modifikation seiner früheren SAGE-*tag*-spezifischen PCR-Methode (*Chen et al 2000*), bei der ebenfalls eine Isolation der 3`terminalen NlaIII-cDNA-Restriktionsprodukte erfolgt. Allerdings wird bei dieser PCR nicht die unten diskutierte Konzeption des *tag*-spezifischen Plusprimers genutzt.

Wang (*Wang S M et al 1998*, *Wang et al 2000*) verwendete zur cDNA-Synthese einen M13-Oligo(dT)-Anker-Primer, der die 3`terminalen Basen dA, dG, dC trägt. Damit sollen sich cDNA-Stränge mit gleich langen PolyA-Teilen synthetisieren lassen. Bei der Sequenzierung der SARA-PCR-Produkte zeigte sich jedoch, dass auch mit dem hier verwendeten M13-Oligo(dT)-Primer ohne Anker-Nukleotid die Längendifferenz innerhalb eines spezifischen SARA-PCR-Produktes aufgrund unterschiedlich langer PolyA-Teile durchschnittlich nur 1-3 bp betrug.

Das Prinzip der Linkerligation an das 5` Doppelstrang-cDNA Ende als Grundlage für den Plusprimer wird in modifizierter Form auch bei Peters (*Peters et al 1999*) und Cheval (*Cheval et al 2002*) beschrieben. Bei beiden Methoden besteht der Plusprimer im Vergleich zur SARA-PCR jedoch ausschließlich aus der Linkersequenz, während als Minusprimer ein 3` transkriptspezifischer Primer dient. Somit können diese Methoden nicht zur *tag*-zu-Gen-Zuordnung, sondern nur zur Bestätigung eines zugeordneten Transkriptes eingesetzt werden.

6.2.2. Primerkonstruktion

Im Gegensatz zu anderen Protokollen profitiert die SARA-PCR weiterhin von der Länge des Plusprimers. Die Konstruktion basiert auf dem Prinzip, nach dem die cDNA modifiziert ist und setzt sich aus dem spezifischen SAGE-*tag* (11 bp), der NlaIII-Restriktionssequenz CATG (4 bp) und der linkerspezifischen Sequenz zusammen. In der Literatur werden mehrere PCR-Methoden beschrieben, bei denen nur die CATG-SAGE-*tag*-Sequenz als Plusprimer dient (*Polyak et al 1997, Matsumura et al 1999, Peters et al 1999, Van den Berg et al 1999, Chen et al 2000, Chen et al 2002*). Aufgrund der geringen Primerlänge (13-15 bp) ergeben sich dabei je nach A/T und G/C Gehalt der einzelnen SAGE-*tags* sehr unterschiedliche Amplifikationsbedingungen, die oft einer aufwendigen Optimierung bedürfen. Im Gegensatz

dazu, wurde in der SARA-PCR durch die Verlängerung des Primers um die Linkersequenz eine höhere Unabhängigkeit der Amplifikations- und Annealingbedingungen von der jeweiligen SAGE-*tag*-Sequenz erreicht. Zur Erhöhung der Spezifität des Primers für verschiedene Gen-Cluster mit gleichem 10 bp-*tag* aber unterschiedlicher 11. Base, wurde auch die 11. Base der SAGE-*tags* in die Primersequenz einbezogen (*Angelastro et al 2000a*). Mit den so konstruierten Plusprimern ließen sich für fast alle *tags* stringente Amplifikationsbedingungen durchführen und aufwendige PCR-Optimierungen entfielen. Dadurch erwies sich die SARA-PCR als geeignete Methode für Anwendungen mit hohem Durchsatz (*high throughput*).

Wang konnte zeigen, dass ähnliche Primer - bestehend aus einer universellen Linkersequenz und einer nur 3-4 bp langen genspezifischen Nucleotidsequenz - eine hohe Spezifität auch für Transkripte aufweisen, die sich nur in einer Base unterscheiden (*Wang et al 1999*).

Der Minusprimer der SARA-PCR besteht aus der M13-spezifischen Sequenz. Er wurde in gleicher Form auch in anderen Protokollen für SAGE-*tag*-spezifische PCR-Methoden beschrieben (*Polyak et al 1997, Van den Berg et al 1999, Chen et al 2002*). Im Gegensatz zu einem einfachen Anker-Oligo(dT)-Minusprimer gleicher Länge (*Peters et al 1999, Chen et al 2000*) trug er durch eine höhere Annealingtemperatur zu einer robusteren und spezifischeren Amplifikation der *tag*-spezifischen Transkripte bei.

6.3. Optimierung der SARA-PCR Bedingungen

„Lineare Preamplifikation“

Bei *tag*-spezifischen Plusprimern mit geringerer Amplifikationseffizienz ließ sich diese durch Erhöhung der Ausgangsmenge des spezifischen cDNA-Fragmentes mittels einer linearen Preamplifikation mit dem *tag*-spezifischen Plusprimer erhöhen (vgl. 5.5.2.4).

Amplifikationsschritte zur Erhöhung des zur Verfügung stehenden Materials sind auch als Modifikationen der SAGE Prozedur mehrfach in der Literatur beschrieben worden. Ebenfalls mit einer linearen Amplifikation arbeitet Vilain, die nach dem NlaIII-Verdau der cDNA einen T7-RNA Polymerase spezifischen Adapter an die cDNA-Fragmente ligiert. Danach erfolgt eine lineare Transkription der cDNA in mRNA (*Vilain et al 2003*). Bei Peters erfolgt eine Modifizierung der 5`cDNA-Enden des ersten Stranges (*Peters et al 1999*). Nachfolgend wird diese cDNA in einer PCR amplifiziert. Dabei könnte aber die Gefahr bestehen, dass insbesondere längere Transkripte weniger effizient amplifiziert werden als kürzere und dadurch im Ansatz unterrepräsentiert werden. Auch in anderen Publikationen wird durch eine cDNA-Amplifikation vor der SAGE-Prozedur eine geringere mRNA-Ausgangsmenge

notwendig (*Neilson 2000*). Bei Datson wird dagegen, um eine ausreichende Menge *ditags* zu generieren, nach der PAGE Gelextraktion der *ditags* eine Re-PCR der *ditags* beschrieben (*Datson et al 1999*). Diese Modifikationen beruhen jedoch nicht auf der Amplifikation eines spezifischen Transkriptes wie im Fall der modifizierten SARA-PCR, sondern auf der Amplifikation der gesamten Transkriptpopulation.

6.4. Anwendung der SARA-PCR

6.4.1. Sensitivität der SARA-PCR

Mit der SARA-PCR konnte für 89 % der getesteten SAGE-*tag*-spezifischen Primer ein positives PCR-Ergebnis erzielt werden. Diese Rate ist mit den Ergebnissen anderer publizierter SAGE-*tag*-spezifischer PCR-Methoden vergleichbar (*Chen et al 2002*).

Die von Chen (*Chen et al 2002*) beobachtete geringe Amplifikationseffizienz mit vermehrtem Auftreten unspezifischer Banden in der PCR, besonders bei *tag*s die in niedriger Kopienzahl in der jeweiligen Bibliothek gefunden wurden, ließ sich in der SARA-PCR bei einer vergleichbaren Menge eingesetzter cDNA (5 ng) nicht feststellen.

Bei einem kleinen Teil der untersuchten *tags* (NM01, NM04, EST01down, EST09down) blieb die *tag*-spezifische 3`Amplifikation trotz Optimierung erfolglos. Dies könnte damit erklärt werden, dass die SAGE-*tag* abhängige Sequenz des Plusprimers nicht immer ausreichend gute Bedingungen für eine optimale Primeranlagerung und effiziente Amplifikation erlaubt. Weiterhin könnte es sich bei nicht amplifizierbaren *tags* auch um Artefakte durch Sequenzierfehler (*Zhang et al 1997*) oder unvollständigen NlaIII-Verdau der Original-SAGE-cDNA (*Neilson 2000*) handeln.

6.4.2. Spezifität der SARA-PCR - Multiple match tags und no match tags

Anhand der Sequenzanalyse von Amplifikaten für SAGE-*tags* bekannter Gene (ß-Actin, Gapdh, Cathepsin, Hsp70, Metallothionein II, Spi 2, Moesin), konnte die Spezifität der SARA-PCR gezeigt werden.

Keines der getesteten SARA-PCR-Produkte wies eine interne CATG-Sequenz auf, was die korrekte 3`terminale Position der amplifizierten Fragmente zeigt und für einen effizienten NlaIII-Restriktionsverdau der cDNA spricht.

Es konnte gezeigt werden, dass die SARA-PCR für *multiple match tags* eine eindeutige Zuordnung zu einem Unigene Cluster ermöglicht (EST08down, EST16down).

Die Analyse der SARA-PCR-Amplifikate der *no match tags* identifizierte mitochondriale Sequenzen (EST02, Nu6M) und bereits bekannte Gene (NM03, EST01). Diese

Beobachtungen deckten sich nicht mit den Ergebnissen von Chen (*Chen et al 2002a*) die zeigten, dass sich unter den *no match tags* etwa 70 % neue, uncharakterisierte Sequenzen befinden. Sie korrelierten eher mit den Ergebnissen von Gunnersen (*Gunnersen et al 2002*) und Pauws *(Pauws et al 2000*), nach denen ein Großteil der *no match tags* keine neuen Sequenzen darstellt. Mögliche Gründe für die Diskrepanzen der verschiedenen publizierten Studien könnten Unterschiede hinsichtlich verwendetem Gewebe, oder Spezies, aber auch technische Aspekte wie Güte der Sequenzierungstechnik, und „Sequenzierungstiefe" der jeweiligen SAGE-Bibliothek (Menge der sequenzierten *tags*) sein.

Vereinzelt kam es zur Amplifikation von Transkripten (EST05, EST06, NM07, EST02down), deren Datenbanksequenz sich an der Position des SAGE-*tags* von der der zugrunde liegenden initial identifizierten SAGE-*tags* unterschied. Die Ursache lag möglicherweise in einer unspezifischen Anlagerung des *tag*-spezifischen Plusprimers an Transkripten deren SAGE-*tags* sich nur in der ersten oder zweiten Base vom Original-*tag* unterschieden. Eine höhere Kopienzahl und ähnliche oder geringere Länge des unspezifischen Transkriptes könnten dabei einen Amplifikationsvorteil begünstigt haben.
Eine alternative Erklärung für die Diskrepanz zwischen Amplifikat und homologer Datenbanksequenz wäre in Sequenzierfehlern innerhalb der Datenbankeinträge (*Bouck 1999*) oder nicht erfassten *single nucleotide* Polymorphismen (*Wang D G et al 1998*) in den 9-10 Basen nach dem letzten CATG (*Michiels et al 1999*) zu suchen. Dies könnte am ehesten für das Amplifikat EST05 zutreffen, da sich hierfür mehrere sequenzidentische, homologe ESTs fanden, die sich jeweils an unterschiedlicher SAGE-*tag*-Position in einer Base vom Original-*tag* EST05 unterschieden.

Im Vergleich zur PCR von Chen (*Chen et al 2002*) wurde mit der SARA-PCR mit den getesteten Plusprimern für *single* und *multiple match tags* ein höherer Anteil singulärer PCR-Produkte generiert (90-100 %). Bei den *no match tags* lag die Amplifikationseffizienz mit 50 % unter der von Chen erreichten, dies hängt möglicherweise mit der geringen Zahl getesteter *no match tags* zusammen.

6.5. SARA-PCR Quantifizierung: Northern blot und Real Time RT-PCR mit der Light Cycler-Technik

Nach der Etablierung des SARA-PCR Protokolls wurde getestet, ob sich die Methode auch zur quantitativen Bestimmung des Genexpressionsniveaus, insbesondere zur Validierung von SAGE-Expressionsdaten, eignet. Dafür wurden aus den möglichen Bestätigungsverfahren der Northern Blot und die RT-PCR bzw. Real Time RT-PCR ausgewählt.

Im Northern Blot kann jeweils die Expression nur eines Gens unter verschiedenen Bedingungen analysiert werden, er gilt vielfach als der Goldstandard in der Quantifizierung von RNA-Expressionsniveaus. Limitierend ist, dass eine relativ große Menge an RNA benötigt wird und er insbesondere bei der Analyse vieler Gene sehr zeitaufwendig ist.

In der hier vorliegenden Arbeit zeigte sich, dass der Einsatz von SARA-PCR-Amplifikaten als Sonde im Northern Blot nur bei einem kleinen Teil der PCR-Produkte sinnvoll war. Über 70 % (25/35) der Amplifikate besaßen eine transkriptspezifische Länge <200 bp (ohne M13$(dT)_{24}$ und Linkersequenz) und führten so vermehrt zu einem unzureichenden Hybridisierungsergebnis. Insgesamt erwies sich der Northern Blot aufgrund des hohen Arbeits- und Zeitaufwandes nicht als optimale Methode zur Validierung der SAGE-Expressionsdaten. Einen weiteren Nachteil stellte die Detektionsgrenze für gering exprimierte Gene dar. Andererseits ließen sich aber wichtige Informationen über die Länge der untersuchten Transkripte gewinnen.

Alternativ konnte demonstriert werden, dass mit dem SARA-PCR Protokoll trotz mehrerer Modifikationsschritte der cDNA eine zuverlässige semiquantitative Quantifizierung als eine Art modifizierter Real Time RT-PCR mit der Light Cycler-Technik möglich ist.

In der Literatur wurden vielfach SAGE-Expressionsdaten mit herkömmlichen RT-PCR Methoden validiert (*Hashimoto et al 1999a,b*). Die Anwendung einer SAGE-*tag*-spezifischen PCR zur semiquantitativen Bestimmung des Genexpressionsniveaus ist bisher nicht beschrieben. Auch die von Chen (*Chen et al 2002*) publizierte, eine der SARA-PCR ähnliche *tag*-spezifische PCR, liefert keine quantitativen Aussagen.

Die Anwendung der Real Time RT-PCR eignete sich besonders für kurze SARA-PCR-Amplifikate. Die Spezifität der Amplifikation wurde mittels Schmelzkurven der jeweiligen PCR-Produkte überprüft, eine gelelektrophoretische Auswertung ist deshalb entbehrlich. Im Hinblick auf die quantitative Untersuchung vieler SAGE-Kandidatengene (*Screening*) könnte somit eine signifikante Zeit- und Arbeitsersparnis erzielt werden.

Weitere hier nicht angewendete Methoden zur Bestätigung von Genexpressionsdaten sind der Dot Blot (Reverser Northern Blot) und DNA-Microarrays. Beim Dot Blot wird jedes zu

analysierende Gen auf eine Nylon-Filter Membran geblottet und mit einer markierten RNA Probe hybridisiert. Die Expressionsniveaus werden über die Intensität des Radioaktivitäts- oder Fluoreszenzsignals der gebundenen Probe quantifiziert.
Bei DNA-Microarrays werden verschiedene DNA-Oligonucleotide korrespondierend zu tausenden Genen auf eine Membran geblottet und mit markierten ss-cDNA-Proben hybridisiert. Die Quantifizierung erfolgt durch Messung der Fluoreszenz- bzw. Radioaktivitätsintensität der gebundenen Probe auf jedem Punkt. Im Gegensatz zum Northern Blot erlauben jedoch diese beiden Hybridisierungsverfahren keinen Rückschluss auf die jeweilige Transkriptlänge und entsprechend geringer fällt die Spezifität des detektierten Signals aus.

6.6. Unterschiede im detektierten Expressionsunterschied zwischen SAGE-Daten, Northern Blot, und Real Time RT-PCR

Bei der Evaluation des Potentials der SARA-PCR als quantitative Methode wurde für SAGE-*tags* mit hoher Kopienzahl (ß-Actin, Gapdh) bzw. hoch signifikantem p-Wert (<0,001) (EST01, Mt2) eine gute Korrelation der detektierten Expressionsunterschiede zwischen Northern Blot, Real Time RT-PCR und SAGE-Expressionsdaten gefunden.
Im Gegensatz dazu war die Korrelation bei differentiell exprimierten *tags* mit geringer Kopienzahl und höherem p-Wert (0,01<p<0,05) (EST10, EST12down) deutlich geringer. Hierbei wurde in den Kontrollverfahren (Northern Blot, Real Time RT-PCR) eine Tendenz zur Detektion eines geringeren Expressionsunterschiedes als in den SAGE-Daten festgestellt. Diese Beobachtung deckte sich mit Ergebnissen von Lal *et al.,* die das sogar für statistisch signifikantere *tags* mit einem p<0,01 fanden (*Lal et al 2001*). Als Erklärung dafür können folgende Ursachen genannt werden:

1. Angesichts der hohen Komplexität neuronalen Gewebes (*Velculescu et al 1999, Sandberg 2000, Datson et al 2001*) waren die beiden hier verwendeten SAGE-Bibliotheken mit je ca. 32000 *tags* (Kontrolle, Ischämie) relativ klein. Die untersuchten *tags* lagen vorwiegend im niedrig exprimierten Bereich (*Trendelenburg et al 2002*), was insgesamt die Repräsentativität und statistische Signifikanz der quantitativen Aussagen begrenzt (*Dinel et al 2005*). Der Vergleich mit Literaturangaben stützte diese Erfahrung (SAGE vs. RT-PCR) besonders für *tags* mit niedriger Kopienzahl (*Anisimov et al 2002*). In weniger komplexen, homogeneren Geweben konnte dagegen auch mit kleineren SAGE-

Bibliotheken eine gute Korrelation zwischen SAGE- und Northern Blot-Ergebnissen erzielt werden (*Gunnersen et al 2000*).

2. Da die RNA-Präparationen von Northern Blot, RT-PCR und SAGE-Bibliothek von unterschiedlichen Tieren stammten, müssen interindividuelle Unterschiede im Genexpressionsniveau sowohl vor als auch nach der Ischämie in Betracht gezogen werden, wobei versucht wurde, diesen Fehler durch Poolen der RNA mehrerer Tiere zu verringern.
3. Eine zusätzliche Fehlerquelle kann in der Auswertung der Real Time RT-PCR Ergebnisse liegen. Hier wurde für das Kandidatengen und das Standardgen (ß-Actin, Gapdh) vereinfachend eine optimale Amplifikationseffizienz von 2 postuliert, die in der Praxis jedoch nicht unbedingt erreicht wird. Dennoch wurde diese vereinfachte Auswertung im Hinblick auf die Zeit-, Kosten-, und Arbeitsersparnis bei der Untersuchung vieler Transkripte als hinreichend genau akzeptiert.

6.7. Gesamtbilanz

Im ersten Teils dieser Arbeit sollten potentiell neue bzw. bisher wenig charakterisierte differentiell exprimierte *tags* einer Ischämie-SAGE-cDNA-Expressionsbibliothek des Maushirnes ihrem korrespondierenden Transkript zugeordnet und in ihrer differentiellen Expression bestätigt werden. Nach Etablierung der SARA-PCR wurden insgesamt 33 SAGE-Kandidatengene ($p<0,05$; mind. 3-facher Expressionsunterschied in den SAGE-Daten) mit der *tag*-Zuordnung EST / *no match* 14 h nach 2 h fokaler cerebraler Ischämie im Maushirn untersucht und davon 26 sicher zugeordnet. Durch die ständig zunehmende Zahl der verfügbaren Sequenzen in den Datenbanken, ließ sich ein Teil der initial als *no match* oder EST eingestuften SAGE-*tags* im Verlauf dieser Arbeit zu näher charakterisierten Transkripten, ESTs oder Sequenzen mit hoher Homologie zu einem humanen Gen zuordnen. Andererseits zeigte sich, dass ein Teil der SAGE-*tags*, welche sich ursprünglich nur einem Unigene-Cluster zuordnen ließen, inzwischen bei erneuter Datenbank-Zuordnung als *multiple match tag* klassifiziert werden müssen (z.B. EST01, EST02, EST13, EST05down, EST15down). Daraus wird deutlich, dass trotz der ständig zunehmenden Zahl zur Verfügung stehender Referenzsequenzen die Spezifität einer ausschließlich Sequenzdatenbank-basierten *tag*-Transkript-Zuordnung nicht zwangsläufig mit ansteigt. Das heißt, dass die SARA-PCR auch in Zukunft eine wichtige Methode zur SAGE-*tag*-spezifischen Identifikation korrespondierender Transkripte darstellen wird.

Von 26 zugeordneten SAGE-*tags* bestätigten sich 12 % (EST01, EST10, EST12down) als differentiell exprimiert. Die ursprüngliche Annahme, dass zumindest ein Teil der nicht bestätigten Kandidaten SAGE-*tags* auf eine initial falsche Sequenzdatenbank-basierte *tag*-Gen-Zuordnung zurückzuführen war, konnte nach der *tag*-spezifischen Transkriptidentifikation mit der SARA-PCR widerlegt werden. Deshalb müssen andere Gründe für den hohen Anteil nicht bestätigter Kandidatengene diskutiert werden:

1. *Auswahl der SAGE-tags*

   Ein wesentlicher Faktor hierbei ist die Auswahl der untersuchten SAGE-*tags*. Schon in der Literatur wurde gezeigt, dass der Anteil falsch Positiver (Rate mit der unterschiedliche *tag*-Anzahlen fälschlicherweise als signifikante Differenzen fehlgedeutet werden) eng mit dem Signifikanzniveau (p-Wert) der ausgewählten SAGE-*tags* korreliert. Es konnte gezeigt werden, dass sich bei sehr stringenten Auswahlkriterien ($p<0{,}001$) über 90 % der differentiell exprimierten *tags* bestätigen lassen (*Gunnersen et al 2002*), während sich bei einem $p<0{,}01$ nur noch 75 % *(Lal et al 2001*) und bei $p<0{,}02$-$0{,}05$ nur etwa 50 % der *tags* (*Ryu et al 2002, Gunnersen et al 2002*) bestätigen ließen. In relativ kleinen SAGE-Bibliotheken mit je etwa 5000 *tags* ließ sich nur die Hälfte der *tags* als differentiell exprimiert bestätigen, wobei alle falsch positiven im niedrig abundanten Bereich <10 *tags* lagen (*de Waard et al 1999, Parle-McDermott et al 2000*).

   Von den hier untersuchten *tags* (ESTs, no match) wies nur eines (EST01) einen stringenten p-Wert $<0{,}001$ auf (*Trendelenburg et al 2002*), während über die Hälfte weniger signifikante p-Werte ($0{,}02<p<0{,}05$) besaßen. Somit musste mit einem deutlichen Anteil falsch positiver *tags* gerechnet werden. Möglicherweise wäre es in diesem Zusammenhang zukünftig sinnvoll, auch *tags* mit geringeren Expressionsunterschieden aber signifikanteren p-Werten in die Auswertung einzubeziehen. Alternativ hätte auch eine Erhöhung der *tag*-Gesamtzahl beider SAGE-Bibliotheken (Ischämie, Kontrolle) die Signifikanz der p-Werte aller *tags* gesteigert (*Audic et al 1997*) und damit eine stringentere Auswahl ermöglicht. Andererseits stellen die hohen Kosten zur Generierung größerer SAGE-Bibliotheken einen limitierenden Faktor für ein Standardlabor dar.

   Wie schon in der Literatur beobachtet (*Robert-Nicoud et al 2001*), bekräftigt diese Arbeit jedoch auch, dass bei der Auswertung von SAGE-Expressionsdaten SAGE-*tags* mit weniger signifikanten p-Werten nicht unbeachtet bleiben dürfen, da sich auch unter ihnen differentiell exprimierte Gene finden lassen.

2. *Auswertung der Expressionsdaten*

   Im Weiteren wird der Anteil falsch Positiver bei der Verifizierung von SAGE-Expressionsdaten auch dadurch bestimmt, ab wann ein detektierter Expressionsunterschied als biologisch relevant gewertet wird. In dieser Arbeit wurden nur mindestens 2-fache Expressionsunterschiede im Northern Blot und in der RT-PCR als signifikant angesehen, während in der Literatur vielfach auch geringere Werte berücksichtigt wurden (*Gunnersen et al 2002*). Da die Genexpression interindividuell stark schwanken kann (*Pritchard et al 2001*), ist die biologische Relevanz eines detektierten Genexpressionsunterschiedes nur durch anschließende funktionelle Studien zu klären. So können einerseits z.B. auch geringe Expressionsunterschiede eine funktionelle biologische Bedeutung haben während andererseits die detektierte mRNA Menge nicht immer mit der Proteinmenge korrelieren muss (*Gygi et al 1999*, *Read et al 2001*). Dessen muss man sich bei der Festlegung einer Signifikanzschwelle zur Unterscheidung zwischen differentiell exprimierten und falsch positiven Genen bewusst sein.

3. *während der SAGE-Prozedur generierte artifizielle Expressionsunterschiede*

   Spinella postuliert, dass durch unterschiedliche Amplifikationseffizienz der *ditags* bei der SAGE-Prozedur (*Spinella et al 1999*) trotz gleicher Ausgangsmenge eines *tags* in zwei SAGE-Bibliotheken artifizielle quantitative Unterschiede generiert werden können.

   Quantitative Artefakte sollen weiterhin bei zu hohen Temperaturen beim Präzipitieren der *ditags* nach dem NlaIII-Verdau durch Verlust A/T-reicher und Anstieg G/C reicher *ditags* entstehen können (*Margulies et al 2001*). Ob das in einer der beiden hier verwendeten SAGE-Bibliotheken der Fall gewesen sein könnte, ließe sich jedoch nur durch aufwendigere statistische Auswertung der Sequenzierungs-Rohdaten klären.

   Sequenzierfehler während des SAGE-Protokolls, die zu einer Verringerung oder Erhöhung der Zahl eines *tags* führen, haben im Hinblick auf die Generierung eines artifiziellen Expressionsunterschiedes wahrscheinlich weniger Bedeutung.

6.8. Weitere Charakterisierung verifizierter Kandidatengene

6.8.1. Charakterisierung von EST10

6.8.1.1. Konsensussequenz EST10

Das SARA-PCR-Produkt des SAGE-*tags* EST10 konnte dem Maus-Homologen des humanen *PTTG-Binding Factor* (PBF) zugeordnet werden.
Es wurde eine 2307 bp lange Konsensussequenz erstellt, die 5`-terminal 13 bp länger und 3`-terminal 51 bp kürzer als die TIGR Datenbank-Konsensussequenz war. Da das 3`-terminale Ende des EST10 Transkriptes mit der SARA-PCR eindeutig definiert werden konnte, stellt der 3´-terminal längere Sequenzbereich der TIGR-Sequenz wahrscheinlich eine Spleißvariante oder einen Artefakt dar.
Die Größendiskrepanz der experimentell bestimmten Konsensussequenz im Vergleich mit dem im Northern Blot detektierten ca. 2400 bp langen Transkript lässt sich durch die Länge des PolyA-Teils erklären. Obwohl es theoretisch möglich wäre, dass mit der 5`-RACE-PCR nicht das vollständige 5`-UTR Ende des Trankriptes isoliert wurde, spricht die Beobachtung, dass alle Klone in diesem Bereich endeten, für eine vollständige Identifizierung der 5`-terminalen Sequenz.

6.8.1.2. Putative Proteinstruktur und Lokalisation

Aus der experimentell bestimmten EST10-Konsensussequenz wurde basierend auf dem längsten ORF mit dem Startcodon AUG und einem der drei möglichen Stoppcodons eine 174 AS lange putative Proteinsequenz abgeleitet.
Sie zeigt beim Vergleich mit dem humanhomologen Protein PBF (*Yaspo et al 1998, Chien et al 2000*) eine Reihe struktureller Gemeinsamkeiten: das spaltbare N-terminale Signalpeptid, die PSI-Domäne, eine Transmembrandomäne, die Kern-Signalsequenz, N-Glycosylierungs- und -Myristoylierungsstellen sowie Phosphorylierungsstellen für cAMP u. cGMP abhängige Kinasen. Die Struktur weist charakteristische Merkmale für Membranproteine auf. Das Signalpeptid (10-15 hydrophobe AS, N-terminal einige positiv geladene AS und wenige polare AS in der Nähe der Spaltstelle) ist typisch für Proteine, welche im endoplasmatischen Retikulum weiter prozessiert werden, wie z.B. Proteine für Lysosomen, Sekretvesikel und Zytoplasmamembranen. N-Glycosylierungsstellen sind für Membranproteine und sezernierte Proteine typisch. Über Myristyl-Reste kann eine Amid-Bindung zwischen N-terminalen Aminogruppen und Fettsäuren erfolgen, dies dient der Fixation eines Protein in der Lipiddoppelschicht einer Membran. Die PSI-Domäne wurde nach einer extrazellulären, für

Plexine, Semaphorine und Integrine charakteristischen Domäne, benannt (*Bork et al 1999*). Sie besteht aus ca. 50 Aminosäure-Resten mit gewöhnlich 8 Cysteinresten, welche untereinander Disulfidbrücken bilden. Proteine die eine PSI-Domäne enthalten, zeigen fast alle eine rezeptorähnliche Struktur. Über die Funktion der PSI-Domäne ist bisher wenig bekannt. Man vermutet, dass sie z.B. bei der Bindung von Proteinen eine Rolle spielen könnte (*Ohta et al 1995*).

Anhand der gefundenen Strukturmerkmale kann vermutet werden, dass EST10 in Form eines rezeptorähnlichen Proteins in der Zellmembran oder im Zytoplasma vorliegt. Nach Bindung eines Liganden könnte über die Kern-Signalsequenz eine Translokation in den Nucleus erfolgen. Beim humanhomologen PBF wurde vermutet, dass es als Zelloberflächenrezeptor nach Bindung eines Liganden über coated-pit vermittelte Endocytose aufgenommen wird (*Yaspo et al 1998*). Im Gegensatz dazu fanden Chien und Pei (*Chien et al 2000*) eine Lokalisation im Zytoplasma und Zellkern. Sie zeigten weiterhin eine Interaktion von PBF mit PTTG (*Pituitary Tumor-Transforming Gene 1 Protein*) über die 30 C-terminalen Aminosäuren des PBF mit der Kernsignalsequenz.

6.8.1.3. Expression und mögliche Funktion

Die Funktion von EST10 im Kontext der fokalen cerebralen Ischämie ist unklar. Im Rahmen dieser Arbeit wurde eine etwa 2-fache Induktion von EST10 14 h nach 2 h fokaler cerebraler Ischämie gezeigt. Über eine differentielle Expression des humanhomologen PBF nach (fokaler) cerebraler Ischämie finden sich in der Literatur keine Angaben. PBF wurde ubiquitär exprimiert gefunden, besonders hoch in der Plazenta (*Chien et al 2000*) und in Hypophysentumoren (*McCabe et al 2003*).

PBF soll, über die Interaktion mit PTTG, einem Protein welches als Schlüsselregulator der Mitose, bei der Zell-Proliferation, Zell-Transformation, DNA-Reparatur und Genregulation (*Zhang et al 1999, Zou et al 1999, Bernal et al 2002*) eine Rolle spielt, die Translokation von PTTG in den Nucleus fördern (*Chien et al 2000*). Es wurde weiterhin vermutet, dass die Coexpression von PBF und PTTG zur Aktivierung von FGF2 durch PTTG erforderlich ist (*Chien et al 2000*). FGF2 wurde nach fokaler cerebraler Ischämie differentiell exprimiert gefunden (*Lin et al 1997*). Außerdem wurde in mehreren Arbeiten eine neuroprotektive Funktion von FGF2 nach fokaler cerebraler Ischämie mit Reduktion des Infarktvolumens durch Reduktion der DNA-Fragmentation, Verhinderung der Herunterregulation von antiapoptotischen Bcl-2 und Schutz vor exzitotoxisch induziertem Zelltod gezeigt (*Kirschner*

et al 1995, Ay et al 2001, Watanabe et al 2004). Im weiteren konnte gezeigt werden, dass PTTG über FGF2 (bFGF) die Angiogenese induziert (*Ishikawa et al 2001*).

Inwieweit PBF tatsächlich – möglicherweise vermittelt über FGF2 – neuroprotektive Eigenschaften besitzt, bleibt zu klären (Abb. 24). Auch ob zwischen der differentiellen Expression von EST10 und FGF2 nach fokaler cerebraler Ischämie ein Zusammenhang besteht, müssen weitere Experimente klären. Es gibt z.B. Hinweise, dass die differentielle Expression von FGF2 während der fetalen Entwicklung des humanen Gehirns und auch beim Adulten nicht das Ergebnis einer veränderten PBF-Expression darstellt (*Boelaert et al 2003*).

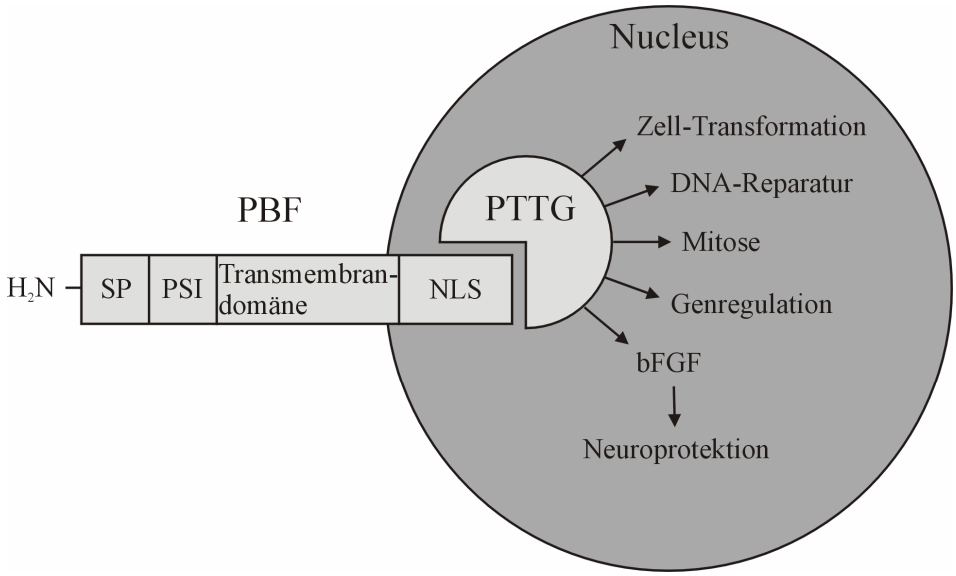

Abb. 24: PBF als neuroprotektives Protein? PBF soll durch die Interaktion mit PTTG über die 30 C-terminalen Aminosäuren der Kernsignalsequenz (NLS) die Translokation von PTTG in den Nucleus fördern (*Chien et al 2000*). PTTG spielt als Schlüsselregulator der Mitose, bei der Zell-Proliferation, Zell-Transformation, DNA-Reparatur und Genregulation (*Zhang* et al *1999, Zou et al 1999, Bernal et al 2002*) eine Rolle. Weiterhin soll die Coexpression von PBF und PTTG zur Aktivierung von bFGF (FGF2) durch PTTG erforderlich sein (*Chien et al 2000*). bFGF wurde als neuroprotektives Gen nach fokaler cerebraler Ischämie differentiell exprimiert gefunden (*Lin 1997*).

6.8.1.4. Ausblick

Zur Klärung der Funktion von EST10 im Rahmen der transienten fokalen cerebralen Ischämie wäre die Untersuchung der Proteinexpression dieses Gens, z.B. mit Western blot sinnvoll. Weiterhin wäre es interessant zu wissen, in welchen Hirnarealen, bzw. in welchen Zellen dieses Protein beim Schlaganfall exprimiert wird. Dazu könnte man Verfahren wie Immunohistochemie oder in situ Hybridisierung verwenden. Zur Aufklärung der molekularen Funktion dieses Proteins könnte man weiterhin untersuchen, in wieweit über die PSI-Domäne

möglicherweise eine Proteinbindung erfolgt, und welche funktionelle Relevanz dies haben könnte.

6.8.2. Charakterisierung von EST01

Die Sequenzanalyse des SARA-PCR-Produktes EST01 ergab eine Sequenzhomologie zu den VL30-Elementen der Maus.

6.8.2.1. Aufbau der VL30-Elemente

Erstmalig wurde die VL30 RNA von Duesberg und Scolnick (1977) in Typ-C murinen Leukämie Virus Partikeln zusätzlich zur 35-38S MLV-RNA gefunden, sie sedimentierte bei rund 30S (*Duesberg 1977*).

Die ca. 150-200 Mitglieder umfassende VL30-Familie ist genetisch nicht direkt mit infektiösen Retrovirus-Genfamilien verwandt, aber sie zeigt strukturelle Gemeinsamkeiten mit retroviralen Proviren. Einige der genomisch integrierten VL30-Elemente werden transkribiert, was in Abhängigkeit von der Verfügbarkeit von retroviralen Transportern zu potentiell infektiösen VL30-Elementen führt (*Song et al 2002*). VL30-Elemente werden vorwiegend als etwa 5-6 kb Transkripte (*Keshet et al 1980*), aber auch als größere (10 kb) (*Nilsson et al 1994*) sowie kleinere (1-4 kb) (*Norton et al 1988*) Transkripte exprimiert. Ein VL30-Element besteht aus einer internen ca. 5 kb langen Sequenz welche 3`- und 5`-terminal von ca. 0,5 kb langen *long terminal repeats* (LTR) flankiert wird. Die Repeats zeigen einen so genannten U3-R-U5 Aufbau. VL30-Elemente können anhand der U3-Region des LTRs in vier Subgruppen (U3 I-IV) eingeteilt werden, die bestimmte konservierte Sequenzmotive enthalten, welche in die Zelltyp-spezifische Transkriptionsregulation involviert sind. Zwischen den Subgruppen lassen sich zahlreiche Punktmutationen, Sequenzduplikationen, Deletionen und Insertionen finden. Die Sequenzhomologie im Bereich der U3-Region innerhalb einer U3 Subgruppe beträgt über 80 %, zwischen den verschiedenen U3 Subgruppen unter 65 %. Die R Regionen der LTRs zeigen hohe Homologie (87-99 %), ebenso die internen VL30 Sequenzen (96-98 %). An der Grenze der LTR zur zellulären DNA befinden sich kurze unvollständige invertierte Repeats. Das VL30-Element selbst wird von kurzen aus vier Nucleotiden bestehenden direkten Repeats flankiert, welche durch Sequenzduplikation an der Integrationsstelle entstanden sind. Downstream (am 3`-Ende) des 5`-LTR befindet sich eine tRNA Primer-Bindungsstelle, welche für die Initiation der reversen Transkription benötigt wird. Die VL30 verwenden so wie andere Retrotransposon-Familien die tRNA Spezies $tRNA^{gly}$, diese wird nicht von infektiösen murinen Retroviren verwendet.

Am inneren Sequenzbereich an der Grenze zum 3`-LTR befindet sich eine Polypurin-Sequenz, welche als Anlagerungsstelle für die reverse Transkription dient (Abb. 25). In geringer Zahl enthält das Maus-Genom auch singuläre LTR der VL30-Familie. Das humane Genom soll Retrotransposon-Sequenzen enthalten, die zu über 90 % identisch mit den Maus-VL30-Sequenzen sind (*Song et al 2004*).

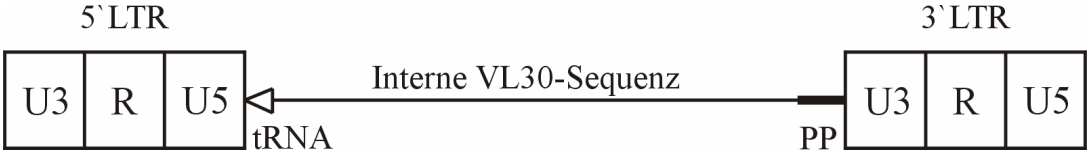

Abb. 25: Aufbau der Maus-VL30-Elemente. Die VL30-Elemente bestehen aus der internen ca. 5 kb langen Sequenz und werden 3` und 5` terminal von ca. 0,5 kb langen *long terminal repeats* (LTR) flankiert, die einen typischen U3-R-U5 Aufbau zeigen. Am 5`LTR befindet sich eine tRNA Primer-Bindungsstelle. An der Grenze zum 3`LTR befindet sich eine Polypurin-Sequenz (PP).

6.8.2.2. EST01 Homologierecherche und Expressionsuntersuchung

Eine spezifische Zuordnung von EST01 zu einem VL30-Element war zunächst nicht möglich. Da mehrere VL30-Elemente das SAGE-*tag* EST01 in der korrekten Position enthielten, musste vermutet werden, dass dieses *tag* möglicherweise die Summe der Expression verschiedener VL30-Elemente darstellt. Zur näheren Spezifizierung wurde deshalb eine 5`RACE-PCR durchgeführt.

Es konnte kein vollständiges VL30-Transkript in der erwarteten Länge von etwa 5-6 kb amplifiziert werden. Dafür wurden zwei kürzere PCR-Produkte erhalten, die sich als Maus VL30-LTR Sequenzen identifizieren ließen.

Durch die identischen repetetiven 3`und 5`terminalen Sequenzen (LTR) der VL30-Elemente kam es bei der 5`RACE-PCR sowohl zu einer 3`terminalen als auch zu einer 5`terminalen Anlagerung des EST01RACE-Primers am LTR. Die Amplifikation des kurzen 5`LTR ist im Gegensatz zur Amplifikation des 5-6 kb langen vollständigen VL30-Elementes (bei Primeranlagerung im 3`LTR) wesentlich effizienter, was die Entstehung der beiden kurzen PCR-Produkte erklären könnte. Innerhalb des LTR existieren aufgrund der Repeat-Sequenz mehrere potentielle Bindungsstellen für den EST01RACE-Primer, so dass zusätzlich unterschiedlich lange Amplifikate entstehen konnten.

Beide RACE-PCR-Produkte konnten keinem einzelnen VL30-Element zugeordnet werden, sondern enthielten Sequenzen verschiedener VL30. Dies ist mit Beobachtungen in der

Literatur vereinbar, die darauf hinweisen, dass sich größere Teile von LTR-Sequenzen verschiedener VL30-Mitglieder untereinander rekombinieren können (*Carter et al 1988*).
In der Literatur wird jedoch nicht nur die Expression vollständiger VL30-Elemente, sondern auch isolierter VL30-LTR (*Gavinski et al 1989*) beschrieben. Deshalb wurde die Länge des differentiell exprimierten Transkriptes mit den zwei EST01-5`RACE-PCR Produkten als Sonde im Northern Blot untersucht. Die Länge des detektierten Transkriptes von etwa 5,8 kb entspricht der eines vollständigen VL30-Elementes.
Allerdings können VL30-Elemente auch als Hybridsequenzen aus nicht VL30-spezifischen inneren Sequenzen mit flankierender VL30-LTR-spezifischer Sequenz an den Enden exprimiert werden (*Itin et al 1983*). Die verwendeten Northern Blot Sonden erlaubten diese Differenzierung jedoch nicht. In einer Real Time RT-PCR mit spezifischen Primern für die interne VL30-Sequenz konnte die differentielle Expression nach fokaler cerebraler Ischämie im Maushirn bestätigt werden. Da die internen VL30-Sequenzen mit 96-98 % zwischen den verschiedenen VL30-Elementen hoch homolog sind (*Nilsson et al 1994*), waren die eingesetzten Primer jedoch nicht für ein bestimmtes VL30-Element spezifisch.
Die Ergebnisse des Northern Blots und der Real Time RT-PCR sprachen also insgesamt dafür, dass das korrespondierende Transkript des SAGE-*tags* EST01 ein vollständiges VL30-Element darstellt. Die Frage, um welches differentiell exprimierte VL30-Element oder möglicherweise welche Gruppe(n) es sich dabei handelt, musste unklar bleiben. Anhand der Homologie der EST01-RACE-PCR-Produkte zu den verschiedenen LTRs ließ sich letztlich lediglich eine Eingrenzung zu den VL30-Elementen der U3-Subgruppen III (B10, Bvl1, Pb10, VL11, PA2, Vleco) oder IV (9λ5, RVL3, NVL1/2, VLov2, VLov1) (*Nilsson et al 1994*) vornehmen.
Dieses Zuordnungsproblem erinnert an das mehrfach in der Literatur diskutierte Zuordnungsproblem humaner SAGE-*tags* die von Alu-Repeat-Sequenzen abstammen (*Neilson 2000*). Eine weitere Spezifizierung ließe sich möglicherweise mit Hybridisierungstechniken mit sehr kurzen spezifischen Oligonukleotiden vornehmen. So konnten z.B. mit U3-spezifischen Oligonucleotiden als Sonde im Northern Blot Mitglieder der Subgruppe U3III in adulten Mäusen (BALB Stamm) in allen Geweben außer Muskel und Gehirn, der Subgruppe U3IV im Herz, Muskel, Lunge, Leber und Testis exprimiert gefunden werden (*Nilsson et al 1994*).

6.8.2.3. Expression von EST01 und mögliche Funktion von VL30-Elementen

Über die biologische Funktion der VL30-Elemente bei Hypoxie und Anoxie (aber auch in anderen pathologischen Zuständen) ist bisher kaum etwas bekannt. Insgesamt sind nur etwa 2-3 % aller VL30-Elemente in der Maus unmethyliert und damit in einem transkriptionell aktiven Zustand (*Carter et al 1986*). Ihre Expression wird während der Embryogenese zeit- und gewebespezifisch in Amnion, Dottersack und Leber - nicht jedoch im Gehirn - reguliert (*Norton et al 1988*). Mit Ausnahme des Gehirns konnte im Northern Blot mit einer VL30-spezifischen Sonde in fast allen Organen adulter BALB-Mäuse eine VL30-Expression gezeigt werden (*Nilsson et al 1994*). In „*wasted*" Mäusen, die eine autosomal rezessive Mutation (wst) tragen und phänotypisch neurologische Abnormitäten, eine fehlerhafte DNA-Reparatur in Lymphozyten, eine Immundefizienz, sowie Funktions- und Entwicklungsstörungen in Milz, Gehirn und Thymus zeigen, wurde eine Induktion von VL30 in Gehirn, Milz und Thymus detektiert. Wahrscheinlich handelt es sich dabei zum großen Teil um die LTR`s und weniger um die internen Sequenzen der VL30-Elemente (*Gavinski et al 1989*).

In der vorliegenden Arbeit wurde eine massive Induktion von VL30-RNA im postischämischen Hirngewebe gefunden. Beginnend bereits 2 h nach 2-stündiger fokaler cerebraler Ischämie im Maushirn wurde eine etwa 10-12-fache Induktion gezeigt mit einem Expressionsmaximum nach etwa 16-24 h um mindestens den Faktor 25. Interessanterweise berichtet auch MacManus in einer kürzlich erschienenen Arbeit (*MacManus et al 2004*), dass sie in ihrem Maus-Ischämiemodel ebenso eine ausgeprägte Induktion von VL30-RNA gefunden hätten und diese auch mit PCR bestätigt, die entsprechende Ergebnisse aber bisher noch nicht publiziert hätten. Weitergehend wird in der Literatur über die Induktion von kompletten VL30-Elementen in Maus NIH3T3- und Rattenfibroblasten durch Anoxie berichtet. Dies scheint jedoch nicht einfach ein Überlebensmechanismus zu sein, da Zellen die VL30 nicht exprimieren genauso anoxietolerant sind (*Anderson et al 1989*). Die Induktion der VL30-Elemente durch Anoxie ist nicht Bestandteil der Streßantwort und benötigt keine Proteinsynthese (*Anderson et al 1993*). Auch durch andere Einflüsse, wie Serum, Wachstumsfaktoren, Calcium, Peptidhormone, Glucocorticoide, *retinoic acid*, 12-O-Tetradecanoylphorbol-13-Acetat (TPA), Cycloheximid, ionisierende Strahlen, Transformation (durch SV40 Infektion oder das Ha-ras Oncogene) kann das Transkriptionsniveau der VL30-Elemente modifiziert werden (*Nilson et al 1994*).

Trotz vielfach beschriebener differentieller Expression wurde bislang jedoch noch kein Translationsprodukt der VL30-Elemente gefunden und es ist unklar, ob es überhaupt Mitglieder der VL30-Familie gibt, die für ein Protein kodieren können. Die Sequenzanalyse

der bisher vollständig sequenzierten VL30-Klone NVL3 und Bvl1 ergab zwar das Vorhandensein gag und pol-ähnlicher Regionen, jedoch spricht das Vorhandensein zahlreicher Stopcodons in jedem Leserahmen gegen ein funktionelles Protein (*French et al 1997*). Gegen die früher in der Literatur diskutierte These, dass die Expression der VL30-Elemente ein Evolutionsresiduum ohne funktionelle Relevanz darstellt (*French et al 1997*) sprechen jedoch inzwischen mehrere Publikationen, die eine Funktion der VL30-Elemente insbesondere bei der Genregulation nahe legen. So soll z.B. eine Untergruppe der Ratten VL30-Elemente, die SVL30, welche durch Anoxie 50-300-fach induziert werden, möglicherweise für ein Protein p61 mit Sequenzverwandtschaft zu den ets-Proteinen (*Laudet et al 1999*) kodieren (*Anderson et al 1993*). Dieses Protein könnte eine transkriptionsregulierende Rolle spielen und andere Anoxie-Gene oder weitere SVL30 induzieren. Es wurde gezeigt, dass der Verlust eines 1,7 kb langen Ratten-SVL30-Bereiches in rekombinierten Harvey Sarkoma Viren der Ratte zum Verlust der Fähigkeit, solide Tumoren in vivo zu induzieren führte (*Anderson et al 1993*). Andere Autoren vermuteten eine Funktion der VL30 m-RNA als nicht kodierendes Transkript bei der Metastasierung von humanen Melanomzellen in einem Modell einer immundefizienten Maus (*Song et al 2002*). In transfizierten humanen Tumorzellen konnte eine transkriptionsregulierende Funktion der VL30-RNA durch Proteinbindung bei der Steroidogenese und Metastasierung gezeigt werden (*Song et al 2004*). Die gleiche Arbeitsgruppe fand später weiterhin, dass durch Bindung einer nichtkodierenden VL30-RNA an die RNA-bindende Domäne eines Tumorsupressor-Proteins beim Säuger (PSF), dessen transkriptionsinhibierende Funktion am Zielgen aufgehoben werden kann (*Song et al 2005*).

6.8.2.4. Ausblick

Ob zukünftig eine spezifische Zuordnung des Gens EST01 zu einem spezifischen VL30-Element gelingt, erscheint fragwürdig. Das wäre unter anderem durch Hybridisierungstechniken mit sehr kurzen spezifischen Oligonucleotiden möglich (*Nilsson et al 1994*), wobei dies technisch sicherlich sehr aufwendig ist. Möglicherweise handelt es sich bei der beobachteten Induktion der VL30-RNA auch um das Resultat der gleichzeitigen Induktion einer Vielzahl von homologen Sequenzen. Weitergehende Fragestellungen ergeben sich hinsichtlich Lokalisation der Expression und möglicher funktioneller Relevanz beim ischämischen Gewebsschaden. Interessant wäre dabei zu untersuchen, ob die VL30-RNA auch in der Ischämie als nicht kodierendes Transkript eine Rolle bei der Genregulation spielt, insbesondere bei der Regulation neuroprotektiver Gene.