

Zur Struktur und Expression des Gens für
das 5'-(CGG)_n-3'-bindende Protein bei Mensch und Maus

Inaugural-Dissertation
zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln
vorgelegt von

Frauke Naumann
aus Hamburg

Köln im März 2003

Berichterstatter: Prof. Dr. Walter Doerfler
Prof. Dr. Diethard Tautz
Tag der mündlichen Prüfung 16. Mai 2003

Meiner Familie

Inhaltsverzeichnis

Abbildungsverzeichnis.....	X
Tabellenverzeichnis.....	XI
Abkürzungsverzeichnis.....	XII

A Einleitung

A.1 Repetitive Sequenzen im menschlichen Genom.....	1
A.2 Mikrosatelliten.....	2
A.2.1 Polymorphismus von Mikrosatelliten.....	3
A.2.2 Instabilität von Mikrosatelliten in Tumoren.....	3
A.2.3 Mögliche Folgen der Instabilität von Mikrosatelliten.....	4
A.3 Fragile Stellen.....	4
A.4 Amplifikation von Mikrosatelliten bei Erbkrankheiten des Menschen.....	5
A.4.1 Amplifikation von Polyglutamin-Ketten.....	6
A.4.2 Amplifikation von Mikrosatelliten in nicht-translatierten Regionen von Genen.....	7
A.4.3 Dynamische Mutation pathogener Mikrosatelliten.....	8
A.5 Amplifikation von 5'-d(CGG) _n -3'-Trinukleotid-Folgen.....	9
A.5.1 Das Fragile-X-Syndrom oder Martin-Bell-Syndrom.....	10
A.5.1.1 Vorkommen und Symptome.....	10
A.5.1.2 Eine polymorphe 5'-d(CGG) _n -3'-Wiederholung in der 5'-nicht- translatierten Region.....	10
A.5.1.3 Inaktivierung des FMR1-Gens.....	10
A.5.1.4 Methylierung des FMR1-Gens.....	11
A.5.1.5 Funktionen des FMR1-Proteins.....	12
A.5.1.6 Auswirkungen einer Prämutation im FMR1-Gen.....	12
A.5.1.7 Zeitpunkt der Amplifikation.....	12
A.5.1.8 Faktoren der Instabilität von 5'-d(CGG) _n -3'-Wiederholungen im Fragilen-X-Syndrom.....	13
A.5.1.8.1 Keimbahnpassage.....	13
A.5.1.8.2 Länge der Repetition.....	13
A.5.1.8.3 Unterbrechungen der Repetition.....	14
A.5.1.8.4 <i>Cis</i> - und <i>trans</i> -aktive Faktoren.....	14

A.5.1.8.5 DNA-Methylierung.....	15
A.6 Überlegungen zum Mechanismus der pathogenen Amplifikation von Mikrosatelliten.....	15
A.6.1 DNA <i>slippage</i> und Sekundärstrukturen.....	15
A.6.2 Interaktionen von Proteinen mit repetitiver DNA.....	17
A.7 Das 5'-d(CGG) _n -3'-bindende Protein CGGBP1.....	17
A.7.1 Reinigung des Proteins und Klonierung des CGGBP1-Gens.....	17
A.7.2 Eigenschaften der Bindungssequenz von CGGBP1.....	18
A.7.3 Eigenschaften des CGGBP1.....	19
A.7.4 Mögliche Funktion von CGGBP1.....	20
A.8 Ziel dieser Arbeit.....	21
B Material	
B.1 Chemikalien.....	22
B.2 Radioisotope.....	24
B.3 Plasmide.....	24
B.4 Lambda-DNA.....	25
B.5 Oligodesoxyribonukleotide.....	25
B.6 Nukleinsäuren (DNA).....	25
B.7 Bakterienstämme.....	25
B.8 Enzyme.....	25
B.9 Restriktionsenzyme.....	25
B.10 Sonstige Enzyme.....	25
B.11 Kits.....	26
B.12 Gewebe.....	27
B.13 Zellen.....	27
B.14 Lösungen und Medien für die Zellkultur.....	27
B.14.1 Allgemeine Zellkultur.....	27
B.14.2 ES-Zellkultur.....	28
B.15 Lösungen und Puffer.....	30
B.16 Datenverarbeitung.....	33
B.17 Sonstiges Labormaterial.....	33

C Methoden

C.1 Zellkultur.....	36
C.1.1 Allgemeine Techniken.....	36
C.1.1.1 Zählen von Zellen.....	36
C.1.1.2 Einfrieren von Zellen.....	36
C.1.1.3 Auftauen von Zellen.....	36
C.1.2 Kultur transformierter Zelllinien „Monolayer-Kulturen“	37
C.1.2.1 Passage und Ernte	37
C.1.2.2 Transfektion transformierter Zelllinien – NukleofektionTM.....	37
C.1.3 Embryonale Stammzellkultur.....	38
C.1.3.1 Gewinnung des LIF-Proteins.....	38
C.1.3.2 Kultur von embryonalen Fibroblasten (EF).....	39
C.1.3.2.1 Gewinnung und Expansion.....	39
C.1.3.2.2 Mitomycin C Behandlung.....	39
C.1.3.3 Arbeiten mit embryonalen Stammzellen.....	40
C.1.3.3.1 Einfrieren und Auftauen ES-Zellen.....	40
C.1.3.3.1.1 Gewebekulturschalen (Ø 6 cm).....	40
C.1.3.3.1.2 Einfrieren in 96-Loch Platten.....	40
C.1.3.3.2 Expansion von ES-Zellen.....	41
C.1.3.3.3 Transfektion embryonaler Stammzellen mit einem Geninaktivierungs-Vektor.....	41
C.1.3.3.3.1 Vorbereitung des Geninaktivierungs-Vektors.....	41
C.1.3.3.3.2 Elektroporation von embryonalen Stammzellen.....	41
C.1.3.3.3.3 Selektion von ES-Zellen.....	42
C.1.3.3.3.4 Isolierung von ES-Zellkolonien.....	42
C.1.3.3.4 DNA-Schnellpräparation aus embryonalen Stammzellen.....	43
C.2 Arbeiten mit Proteinen.....	43
C.2.1 Proteinextraktion aus transfizierten Zellen.....	43
C.2.1.1 Erntezeitpunkt.....	43
C.2.1.2 Rohextrakte aus Zellen aus Gewebekulturschalen (Ø 10 cm).....	43
C.2.1.3 Lyse der Zellen in der Gewebekulturschale (Ø 3 cm).....	44
C.2.2 Bestimmung der Proteinkonzentration.....	44
C.2.3 Luciferase-Aktivitätsbestimmungen.....	44
C.2.4 Gelretardation.....	45

C.3 Präparation von Nukleinsäuren.....	45
C.3.1 Präparation von genomischer DNA.....	45
C.3.1.1 Präparation von genomischer DNA aus kultivierten Zelllinien.....	45
C.3.1.2 Präparation von DNA aus Organen.....	46
C.3.1.3 Reinigung von DNA mit Phenol/Chloroform.....	46
C.3.2 Präparation von Plasmid-DNA.....	46
C.3.2.1 Schnellpräparation von Plasmid-DNA.....	47
C.3.2.1.1 Zahnstocher-Methode.....	47
C.3.2.1.2 Kochlyse.....	47
C.3.2.1.3 Alkalische Lyse.....	47
C.3.2.2 Präparation größerer Mengen von Plasmid-DNA.....	48
C.3.2.3 Synthetische Oligodesoxyribonukleotide.....	48
C.3.2.4 Isolation von DNA-Fragmenten aus Gelen.....	49
C.3.2.4.1 Isolation von DNA-Fragmenten aus Agarosegelen.....	49
C.3.2.4.2 Isolation von DNA-Fragmenten aus Polyacrylamidgelen und UV-Shadowing.....	49
C.3.3 Präparation von RNA.....	49
C.3.3.1 Präparation von Gesamt-RNA aus kultivierten Zellen.....	49
C.3.3.2 Präparation von mRNA aus RNA-Präparationen.....	50
C.3.4 Ethanolpräzipitation von Nukleinsäuren.....	50
C.3.5 Bestimmung der Nukleinsäurekonzentration.....	50
C.3.5.1 Photometrische Bestimmung der Nukleinsäurekonzentration.....	50
C.3.5.2 Bestimmung der DNA-Konzentration durch ein Tüpfelgel.....	51
C.4 Genomisches Sequenzieren nach der Bisulfit-Methode	51
C.4.1 Behandlung der DNA mit Bisulfit.....	51
C.4.2 PCR im Rahmen des genomischen Sequenzierens.....	52
C.5 Enzymatische Reaktionen mit Nukleinsäuren.....	53
C.5.1 Restriktionsspaltung von DNA.....	53
C.5.2 Polymerase Kettenreaktion.....	53
C.5.3 <i>In-vitro</i> -Methylierung von DNA.....	54
C.5.4 Modifikation von DNA-Enden.....	54
C.5.5 Einführen von fortlaufenden Deletionen.....	54
C.5.6 Dephosphorylierung.....	55
C.5.7 Radioaktive Markierung von DNA.....	55

C.5.7.1 Oligo-Markierung.....	55
C.5.7.2 Kinasierung von DNA-Enden.....	56
C.5.8 Sequenzierung von DNA mit der Kettenabbruchstechnik.....	56
C.5.9 RT-PCR.....	57
C.6 Elektrophoretische Trennung von Nukleinsäuren.....	57
C.6.1 Agarosegele.....	57
C.6.1.1 Elektrophoretische Trennung von DNA in Agarosegelen.....	57
C.6.1.2 Qualitätskontrolle von RNA in Agarosegelen.....	58
C.6.2 Elektrophoretische Trennung von DNA in Polyacrylamidgelen.....	58
C.7 DNA-Transfer-Hybridisierung „Southern Blot“.....	59
C.8 RNA-Transfer-Hybridisierung „Northern Blot“.....	59
C.9 Hybridisierung radioaktiv markierter DNA-Sonden an membrangebundene Nukleinsäuren.....	60
C.9.1 Hybridisierung der membrangebundenen DNA-Fragmente.....	60
C.9.2 Entfernen radioaktiver Sonden von hybridisierten Membranen.....	60
C.10 Herstellung einer subgenomischen Bibliothek.....	60
C.10.1 Größenselektion durch Zonenzentrifugation im Saccharosegradienten.....	60
C.10.2 Präparation der zu klonierenden DNA-Fragmente.....	61
C.10.3 Präparation des Vektors λ -DASH II.....	61
C.10.4 Ligation und Verpackung.....	62
C.10.5 Analyse der subgenomischen Bank – Plaquehybridisierung.....	62
C.10.6 Arbeiten mit λ -DASH II.....	62
C.10.6.1 Präparation der Wirtsbakterien XL1 blue MRA (P2).....	62
C.10.6.2 Isolierung einzelner Plaques.....	63
C.10.6.3 Präparation von λ -DNA.....	63
C.11 Klonierungstechniken.....	64
C.11.1 Ligation.....	64
C.11.1.1 Präparation und Ligation der Vektor- und Fremd-DNA.....	64
C.11.1.2 Ligation von PCR-Produkten in den pGEM-T-Vektor.....	64
C.11.2 Herstellung elektrokompeter Bakterien.....	65
C.11.3 Transformation von Bakterien.....	65
C.11.3.1 Elektroporation.....	65
C.11.3.2 Transformation durch Hitzeschock.....	65
C.11.4 Identifizierung rekombinierter Bakterienkolonien „Blau-Weiß-Selektion“.....	66

C.11.5 Identifizierung rekombinanter Bakterienkolonien durch Kolonie-Hybridisierung.....	66
C.12 Sequenzvergleiche in Datenbanken mit dem BLAST Programm.....	66
D Ergebnisse	
D.1 Vorarbeiten.....	68
D.1.1 Der genomische Klon des CGGBP1-Gens des Menschen.....	68
D.1.2 Das CGGBP1-Gen in anderen Spezies.....	69
D.1.3 Chromosomale Lokalisation des CGGBP1-Gens.....	70
D.2 Klonierung des cggbp1-Gens des Mausstamms C57BL/6.....	70
D.2.1 Herstellung und Durchsuchen einer genomischen Bank.....	71
D.2.2 PCR-Klonierung der 5'-flankierenden Sequenz des cggbp1-Gens.....	72
D.2.3 Kontrolle des isolierten cggbp1-Gens aus C57BL/6-DNA.....	73
D.2.4 Restriktionskartierung des genomischen Klons.....	75
D.2.5 Sequenzierung des isolierten cggbp1-Gens.....	75
D.3 Expressionsanalysen von CGGBP1 und cggbp1.....	76
D.3.1 Expression von CGGBP1 und cggbp1 in Zelllinien.....	77
D.3.2 Expression von CGGBP1 in Geweben.....	78
D.3.2.1 Expression von CGGBP1 in adulten Geweben	78
D.3.2.2 Expression von CGGBP1 in verschiedenen Gehirnregionen.....	79
D.3.2.3 Expression von CGGBP1 in fötalen Geweben des Menschen.....	80
D.3.3 Alternatives Spleißen von CGGBP1-Transkripten.....	80
D.3.4 Transkription des cggbp1-Gens während der Embryonalentwicklung.....	82
D.4 Struktur und Konservierung des CGGBP1-Gens.....	83
D.4.1 Identifizierung von CGGBP1-Genen in anderen Spezies.....	83
D.4.1.1 Identifizierung des CGGBP1-Gen der Ratte – r-cggbp1.....	84
D.4.1.2 EST-Klone des CGGBP1-Gens anderer Spezies.....	88
D.4.2 Identifizierung von Domänen in der Aminosäuresequenz des CGGBP1	88
D.4.3 Allgemeine Struktur der CGGBP1- und cggbp1-Gene.....	89
D.4.3.1 Das offene Leseraster der CGGBP1-Gene.....	89
D.4.3.2 Die nicht-translatierten Regionen der CGGBP1-Gene.....	91
D.4.3.2.1 Die 5'-nicht-translatierte Region (5'-UTR).....	91
D.4.3.2.2 Die 3'-nicht-translatierte Region – alternative Polyadenylierung der CGGBP1-Gene.....	95
D.4.4 Analyse des CGGBP1-Promotors in Transfektions-Experimenten.....	97

D.4.4.1 Herstellung der Reportergenkonstrukte.....	98
D.4.4.1.1 Ableitung des Transkriptionsstartpunktes.....	98
D.4.4.1.2 Klonierung.....	98
D.4.4.2 Eigenschaften der Reportergenkonstrukte.....	100
D.4.4.2.1 Eigenschaften der Reportergenkonstrukte mit der 5'-flankierenden Sequenz des CGGBP1-Gens.....	100
D.4.4.2.2 Eigenschaften der Reportergenkonstrukte mit der 5'-UTR des CGGBP1-Gens	103
D.4.4.3 Nukleofektion der Promotortestplasmide in HeLa-Zellen.....	104
D.4.4.3.1 Etablierung der Transfektionsmethode der Nukleofektion.....	104
D.4.4.4 Der minimale Promotor des CGGBP1-Gens.....	105
D.4.5 Methylierungssensitivität des CGGBP1-Promotors.....	107
D.4.6 Methylierungsstatus des CGGBP1-Gens.....	109
D.4.6.1 Lokalisierung von CpG-Dinukleotiden im genomischen Klon des CGGBP1-Gens.....	109
D.4.6.2 Genomisches Sequenzieren nach der Bisulfitmethode.....	109
D.4.6.3 Keine CpG-Methylierung im CGGBP1-Gen in primären Zellen und transformierten Zelllinien.....	110
D.5 Gezielte Geninaktivierung des CGGP1 Gens.....	113
D.5.1 Technische Grundlagen.....	113
D.5.1.1 Konventionelle Geninaktivierung.....	113
D.5.1.2 Konditionale Geninaktivierung.....	115
D.5.2 Grundlegende Überlegungen zur gezielten Geninaktivierung des cggbp1-Gens.....	116
D.5.2.1 Art der Geninaktivierung.....	116
D.5.2.2 Art der Mutation.....	116
D.5.3 Strategie der Geninaktivierung des cggbp1-Gens.....	117
D.5.3.1 Wahl der Screening-Methode für die gezielte Inaktivierung des cggbp1-Gens.....	117
D.5.3.2 Nachweis von Rekombinanten im cggbp1-Genlokus.....	117
D.5.3.3 Detektion von homologen Rekombinanten nach der Transfektion von ES-Zellen.....	120
D.5.3.3.1 Detektion der LoxP-Sequenz C.....	120

D.5.3.3.2 Klassifizierung der Allele nach der Rekombination durch die Cre-Rekombinase.....	121
D.5.3.4 Anwendbarkeit der Strategie im Maustamm 129.....	122
D.5.4 Klonierung des Geninaktivierungs-Vektors.....	123
D.5.4.1 Der Ausgangsplasmid easyflox.....	123
D.5.4.2 Klonierung des offenen Leserasters des cggbp1-Gens in easyflox (EF-I).....	123
D.5.4.3 Klonierung des SacII/PstI-Fragmentes in EF-I (EF-II und III).....	123
D.5.4.4 Klonierung des SnaBI/BamHI-Fragmentes in EF-III (EF-IV).....	125
D.5.5 Transfektion des LoxP-flankierten cggbp1-Gens in ES-Zellen.....	126
D.5.5.1 Transfektion in C57BL/6 ES-Zellen.....	126
D.5.5.2 Transfektion in 129-ES-Zellen.....	127
D.5.6 Modifizierung des Geninaktivierungs-Vektors.....	127
D.6 Charakterisierung der Bindungssequenz von CGGBP1.....	129
D.6.1 Bindung an 5'-d(CGG) _n -3' (n ≤ 8).....	129
D.6.2 Bindung an unterbrochene 5'-d(CGG) _n -3'-Trinukleotid-Folgen.....	131
E Diskussion	
E.1 Der genomische Klon des cggbp1-Gens.....	133
E.2 Zur Expression des CGGBP1-Gens.....	134
E.3 Das 5'-d(CGG) _n -3'-bindende Protein.....	135
E.3.1 Eigenschaften des offenen Leserasters.....	135
E.3.2 Funktionelle Überlegungen.....	136
E.4 Eigenschaften der 5'-nicht translatierten Regionen der CGGBP1-Gene.....	137
E.5 Alternative Polyadenylierung im CGGBP1-Gen.....	139
E.6 Der Promotor des CGGBP1-Gens.....	139
E.7 Zur Methylierung des CGGBP1-Gens.....	140
E.8 Zur konditionalen Inaktivierung des cggbp1-Gens.....	142
E.9 Die Bindung von CGGBP1 an kurze und unterbrochene 5'-(CGG) _n -3'-Wiederholungen.....	144
E.10 Ausblick.....	144
F Zusammenfassung	146
G Summary	149
H Literatur	151

I Anhang

I.1 Datenbankeintrag von CGGBP1 - Zugriffsnummer AJ000258.....	179
I.2 Verwendete Oligodesoxyribonukleotide.....	181
I.2.1 Sequenzierung von m20-9.....	181
I.2.2 Oligodesoxyribonukleotide zur Sequenzierung/Klonierung des Geninaktivierungs-Vektors.....	182
I.2.3 Oligodesoxyribonukleotide für RT-PCR und Klonierung/Sequenzierung der Promotortestkonstrukte von g20.....	183
I.2.4 Oligodesoxyribonukleotide für das genomische Sequenzieren nach der Bisulfit-Methode.....	183
I.3 Sequenz – Zugriffsnummer AY189895.....	184
I.4 Plasmidkarten m20-9 und m20-10.....	189
I.5 EST-Klone.....	190
I.6 Quellinformation der Northern Blots der Firma Clontech.....	191
I.7 Liste der zitierten Internetseiten.....	192
I.8 Zugriffsnummern der zitierten U1-Zink-Finger-Proteine.....	192
I.9 Kurzzusammenfassung.....	193
I.10 Abstract.....	194
I.11 Lebenslauf.....	195
I.12 Erklärung.....	196

Abbildungsverzeichnis

Abbildung 1	Position und Sequenz der pathogenen Mikrosatelliten und der kodierten Aminosäuren in einem schematisierten menschlichen Gen.....	6
Abbildung 2	Übersichtskarte des genomischen Klon g20-2 des CGGBP1-Gens aus HeLa-DNA.....	68
Abbildung 3	Hybridisierung von genomischer DNA verschiedener Spezies mit der cDNA des menschlichen CGGBP1-Gens.....	69
Abbildung 4	Chromosomale Lokalisation des CGGBP1-Gens des Menschen auf dem kurzen Arm von Chromosom 3(cen).....	70
Abbildung 5	Restriktionsanalyse von C57BL/6 DNA.....	71
Abbildung 6	Kontrolle der 5'-flankierenden Sequenz.....	72
Abbildung 7	Rückhybridisierung.....	73
Abbildung 8	DNA-Transfer-Analyse von genomischer Maus-DNA und den klonierten Fragmenten m20-1.1 und m20-9.....	74
Abbildung 9	Restriktionskartierung des genomischen Klon des cggbp1-Gens.....	75
Abbildung 10	Restriktionskarte der genomischen Sequenz.....	76
Abbildung 11	Northern-Transfer-Hybridisierung von Gesamt-RNA und Poly-A-vorselektionierter mRNA verschiedener Zelllinien des Menschen und der Maus.....	77
Abbildung 12	Hybridisierung von RNA aus adulten Geweben des Menschen.....	78
Abbildung 13	Hybridisierung von RNA aus verschiedenen Gehirnregionen des Menschen.....	79
Abbildung 14	Hybridisierung von RNA aus fötalen menschlichen Geweben.....	80
Abbildung 15	Alternatives Spleißen in der 5'-UTR des CGGBP1-Gens.....	81
Abbildung 16	Hybridisierung von RNA aus Mausembryonen unterschiedlicher Entwicklungsstadien.....	82
Abbildung 17	Genomische Organisation des CGGBP1-Gens des Menschen und der Maus.....	85
Abbildung 18	Ähnlichkeiten in den genomischen Sequenzen.....	87
Abbildung 19	Sequenzvergleich von CGGBP1 (Aminosäure 41-91) mit dem C2H2-Zinkfingermotiv der U1-ähnlichen Zinkfingerproteine.....	89
Abbildung 20	Aminosäuresequenz des CGGBP1-Gens in verschiedenen Spezies.....	90
Abbildung 21	Die 5'-UTR der CGGBP1-Gene.....	92
Abbildung 22	Sequenzvergleich der Peptid- und Nukleotidsequenzen im ersten Minicistron der CGGBP1-Gene.....	95
Abbildung 23	Klonierung der Reporter-genkonstrukte.....	99
Abbildung 24	Sequenzvergleich der Promotorregion des CGGBP1-Gens.....	102-3

Abbildung 25	Bestimmung des Erntezeitpunktes.....	105
Abbildung 26	Funktionale Analyse des CGGBP1-Promotors in HeLa-Zellen.....	106
Abbildung 27	Überprüfung der <i>in-vitro</i> -Methylierung	107
Abbildung 28	Methylierungssensitivität des CGGBP1-Promotors	108
Abbildung 29	Genomisches Sequenzieren nach der Bisulfit-Methode.....	110
Abbildung 30	Genomisches Sequenzieren nach der Bisulfit-Methode: Karte des CGGBP1-Promotors.....	112
Abbildung 31	Strategie der Geninaktivierung.....	118
Abbildung 32	Southern-Transfer-Analyse mit den Testsonden.....	119
Abbildung 33	DNA-Transfer-Hybridisierung von DNA des Mausstamms 129Sv.....	122
Abbildung 34	Klonierung des Gen-Inaktivierungs-Vektors.....	124
Abbildung 35	Restriktionsanalyse des Geninaktivierungs-Vektors EF-IV.....	125
Abbildung 36	DNA-Transfer-Hybridisierung mit DNA aus transfizierten ES-Zellen des C57BL/6-Mausstamms.....	126
Abbildung 37	Modifizierung des Geninaktivierungs-Vektors.....	128
Abbildung 38	Restriktionsanalyse des Geninaktivierungs-Vektors EF-V.....	128
Abbildung 39	Kompetitions-Experimente.....	130
Abbildung 40	Gelretardation mit CGGBP1.....	132
Abbildung 41	<i>In-vitro</i> -Mutationen von CGGBP1.....	136
Abbildung 42	Plasmidkarten m20-9 und m20-10.....	189

Tabellenverzeichnis

Tabelle 1	Spezifische Bindung von CGGBP1 an Oligodesoxyribonukleotide.....	19
Tabelle 2	dN/dS-Werte.....	91
Tabelle 3	Sequenzumgebung der Startcodons in den CGGBP1-Genen in fünf Säugetierspezies.....	94
Tabelle 4	Mögliche Polyadenylierungssignale in CGGBP1 des Menschen, der Maus und der Ratte.....	97
Tabelle 5	Genomisches Sequenzieren nach der Bisulfit-Methode: Primer und amplifizierte Regionen.....	111
Tabelle 6	Nachweis für homologe Rekombinanten mit der 5'-externen Sonde.....	120
Tabelle 7	Nachweis für LoxP C mit allen Sonden.....	121
Tabelle 8	Nachweis der Allele nach Rekombination durch Cre mit der 3'-internen Sonde.....	121
Tabelle 9	Sequenz der in EMSA verwendeten doppelsträngigen Oligodesoxyribonukleotide.....	129

Abkürzungsverzeichnis

A	Adenosin
Abb.	Abbildung
ATG	Startcodon der Translation (bei der Besprechung von cDNAs)
b	Base
bp	Basenpaar
BSA	Rinderserumalbumin (<i>bovine serum albumin</i>)
BPB	Bromphenolblau
C2H2	zwei Cystein- und zwei Histidinreste
cDNA	<i>complementary DNA</i>
CGGBP1	5'-(CGG) _n -3'-bindendes Protein 1
CHO	<i>chinese hamster ovary</i>
Ci	Curie
dbEST	Datenbank der EST-Klone am NCBI
DM	Myotone Dystrophie
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
dNTP	Desoxyribonukleotid-Triphosphat
ddNTP	Didesoxyribonukleotid-Triphosphat
E. coli	<i>Escherichia coli</i>
EF	embryonale Fibroblasten
EMSA	Gelretardierungs-Experiment (<i>electrophoretic mobility shift assay</i>)
ES	embryonale Stammzellen
EST	exprimierte sequenzmarkierte Stelle (<i>expressed sequence tag</i>)
et al.	und andere
EtBr	Ethidiumbromid
ExoIII	Exonuklease III
FCS	fötale Kälberserum (<i>fetal calf serum</i>)
FraX	Fragile Stelle A auf dem X-Chromosom (Fragiles X-Syndrom)
Fra(Zahl)	Fragile Stelle auf Chromosom (Zahl)
G	Guanosin
g20-2	genomischer Klon des CGGBP1-Gens des Menschen
GCG	Genetics Computer Group (Programmpaket zur Sequenzanalyse der Universität Wisconsin)
GFP	grün-fluoreszierendes Protein
IHGSC	<i>International Human Genome Sequencing Consortium</i>
kb	Kilobasen(paar)
λ	Lambda
LIF	Leukämie inhibierender Faktor
LINE	<i>long interspersed elements</i> (Familie repetitiver Elemente)
LTR	<i>long terminal repeats</i> (Familie repetitiver Elemente)
mA	Milli-Ampere
M	Mol
MBI	Firma MBI Fermentas, Vilnius
MGSC	Mouse Genome Sequencing Consortium
µg	Mikrogramm

µl	Mikroliter
m20-1.1	genomischer Klon des CGGBP1-Gens der Maus in λ kloniert
m20-9/10	genomischer Klon des CGGBP1-Gens der Maus in Plasmid kloniert
ml	Milliliter
mRNA	Boten-RNA (<i>messenger RNA</i>)
NCBI	<i>National Center for Biotechnology Information</i> (www.ncbi.nlm.nih.gov)
neo	Gen für die Resistenz gegenüber Geneticin und Neomycin
nt	Nukleotid
NTH	Northern Transfer Hybridisierung
OD	optische Dichte
ORF	<i>open reading frame</i>
PAA	Polyacrylamid
PAGE	Polyacrylamidgelelektrophorese
PBS(d)	<i>phosphate buffered saline (deficient)</i>
pBSII	Plasmid Bluescript II
PCR	Polymerase-Kettenreaktion (<i>polymerase chain reaction</i>)
pfu	<i>plaque forming units</i>
PWBZ	periphere weiße Blutzellen
RNA	Ribonukleinsäure
RNase	Ribonuklease
SAM	S-Adenosyl-Methionin
SINE	<i>short interspersed elements</i> (Familie repetitiver Elemente)
SSC	<i>standard saline citrate</i>
T	Thymidin
Thr	Threonin
TK	Thymidin-Kinase
u.a.	unter anderem
uORF	<i>upstream open reading frame</i>
U.p.m.	Umdrehungen pro Minute
U	Uridin
UTR	nicht translatierte Region einer mRNA bzw. cDNA (<i>untranslated region</i>)
U	Einheit der Enzymaktivität
v.a.	vor allem
z.B.	zum Beispiel
z.T.	zum Teil

A Einleitung

A.1 Repetitive Sequenzen im menschlichen Genom

Eukaryontische Genome unterscheiden sich drastisch in ihrer Größe, ohne daß die Menge an DNA dabei mit der Komplexität eines Organismus zu korrelieren scheint (Mirsky und Ris, 1951; Thomas, 1971; Hartl, 2000). Die grundlegenden Beobachtungen zu diesem sogenannten C-Wert-Paradox stammen noch aus Zeiten, in denen die Struktur der DNA unbekannt war. Nach Experimenten zur Renaturierungskinetik von genomischer DNA wurde klar, daß DNA repetitive Bereiche enthielt, und daß das Ausmaß dieser repetitiven Bereiche in verschiedenen, auch verwandten Organismen sehr unterschiedlich ausfiel (Britten und Kohne, 1968; Bonner et al., 1973; Davidson et al., 1973). Die weitgehende Sequenzierung mehrerer eukaryontischer Genome macht es nun möglich, die Anteile kodierender DNA am Genom quantitativ zu erfassen. Während bei der Hefe *Saccharomyces cerevisiae* etwa 70% des Genoms aus proteinkodierenden offenen Leserastern (ORF) besteht (Goffeau et al., 1996), umfassen Strukturgene und regulatorische Bereiche nur etwa 5% des menschlichen Genoms (International Human Genome Sequencing Consortium, 2001).

Etwa die Hälfte des menschlichen Genoms läßt sich eindeutig verschiedenen Familien repetitiver Elemente zuordnen, von denen die meisten von Transposons oder Retrotransposons abgeleitet sind (International Human Genome Sequencing Consortium, 2001). Solche *short* oder *long interspersed elements* (SINEs oder LINEs) weisen eine charakteristische Verteilung innerhalb der bisher sequenzierten Genome von Mensch und Maus auf (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). Obwohl SINEs in beiden Spezies in den Regionen hoher Gendichte angereichert zu sein scheinen, unterscheiden sich insgesamt jedoch die Zusammensetzung und der Anteil bestimmter repetitiver Elemente in beiden Genomen (Mouse Genome Sequencing Consortium, 2002).

Anfangs wurde vorgeschlagen, daß es sich bei repetitiver DNA um Abfall (*junk*) (Ohno, 1972), oder um eigennützige (*selfish*) DNA (Doolittle und Sapienza, 1980; Orgel und Crick, 1980) handeln könnte. Angesichts des formenden Einflusses von repetitiven Elementen auf Genome erscheint diese Interpretation nun als voreilig. Die Rekombination von verstreut liegenden repetitiven Elementen ist wahrscheinlich an der Neuverknüpfung genetischer Information beteiligt. Die sich durch reverse Transkription und Transposition vermehrenden

repetitiven Elemente waren aufgrund ihrer Mobilität vermutlich auch bei der Schaffung von neuen Genen, oder der Neuorganisation von Genen bzw. Proteindomänen, beteiligt (International Human Genome Sequencing Consortium, 2001).

SINEs und LINEs gehören zu den über das gesamte Genom verteilten repetitiven Sequenzen, die für die Evolution von Genomen zwar möglicherweise von Bedeutung waren, denen *per se* jedoch keine Funktion zugeordnet werden kann. Im Genom werden aber auch funktionelle Regionen mehrfach wiederholt: Die rRNA-Gene liegen in tandem-artiger Wiederholung auf den kurzen Armen der akrozentrischen Chromosomen 13, 14, 15, 21 und 22 und bilden die die Nukleoli organisierenden Zentren (Wellauer und Dawid, 1979). Die sogenannte Satelliten-DNA an den Telomeren und den Centromeren eukaryontischer Chromosomen besteht aus Wiederholungen kurzer DNA-Sequenzen, die für die Struktur und Stabilität von Chromosomen von großer Bedeutung sind (Blackburn, 1990). Auf eine bestimmte Klasse repetitiver DNA, nämlich die der Mikrosatelliten, soll im folgenden genauer eingegangen werden.

A.2 Mikrosatelliten

Bei Wiederholungen von einer Sequenzeinheit zwischen 1 bis 6 Basen spricht man von Mikrosatelliten oder von einfachen Sequenz-Wiederholungen (*simple sequence repeat, SSR*) (Tóth et al., 2000; Katti et al., 2001; International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). Sie kommen in allen bisher untersuchten eukaryontischen und – zu einem geringeren Anteil – auch in prokaryontischen Genomen vor (Tautz et al., 1986; Richard et al., 1999). SSRs machen allein etwa 3% des menschlichen Genoms aus (International Human Genome Sequencing Consortium, 2001). Im Genom der Maus treten sie bis zu viermal häufiger auf und die Einheiten werden z.T. auch öfter wiederholt als im menschlichen Genom (Mouse Genome Sequencing Consortium, 2002; Love et al., 1990; Beckmann und Weber, 1992).

Die Häufigkeit bestimmter SSRs hängt von der Länge der wiederholten Einheit und ihrer Sequenzzusammensetzung ab (Tóth et al., 2000). Man beobachtet außerdem, daß sich die Häufigkeit des Auftretens bestimmter SSRs in verschiedenen Spezies unterscheidet (Beckmann und Weber, 1992; Tautz und Schlötterer, 1994; Lagercrantz et al., 1993). SSRs sind je nach Sequenz und Länge der wiederholten Einheit in Bezug auf ihre Position in Exons, Introns, intergenischen Regionen und auch Chromosomen ungleich verteilt (Katti et al., 2001). Die meisten Mikrosatelliten befinden sich in Positionen, in denen Längenveränderungen wahrscheinlich keine funktionalen Folgen haben (Cox und Mirkin, 1997; Metzgar et al., 2000). Wenn SSR in proteinkodierenden Exons auftreten, so sind Trinukleotid-Wieder-

holungen, gefolgt von Hexanukleotid-Wiederholungen, am häufigsten (Tóth et al., 2000), was angesichts der Anforderungen des Leseraster nicht erstaunlich ist. CGG/CCG-Trinukleotide scheinen bei Vertebraten in Exons, vor allem aber in intergenischen Regionen angereichert zu sein (Tóth et al., 2000). Möglicherweise treten sie bevorzugt am 5'-Ende von Genen auf (Debrauwere et al., 1997). Die Position von SSRs in kodierenden Regionen ist häufig bei verschiedenen Spezies konserviert, so daß ihnen in diesen Fällen eventuell eine wichtige Rolle bei der Proteinfunktion zugeschrieben werden kann (Übersicht in Künzler et al., 1995; Djian et al., 1995; Rubinsztein et al., 1995a; Eichler et al., 1995).

A.2.1 Polymorphismus von Mikrosatelliten

Eine charakteristische Eigenschaft von SSRs erleichterte Kartierungsprojekte und das *DNA fingerprinting*, das in der Forensik oder bei Vaterschaftstests von großer Bedeutung ist (Tautz, 1989; Weissenbach, 1993; Jeffreys et al., 1991). Die Mutationsrate von SSRs liegt bei 10^{-3} pro Generation und Genlocus (Weber und Wong, 1993; Brinkmann et al., 1998). SSRs sind deshalb in der Bevölkerung polymorph und daher als Marker sehr gut geeignet. Meist werden bei einer Mutationen nur eine oder wenige Einheiten an die bestehende Wiederholung addiert bzw. deletiert (zur Übersicht Ellegren, 2000). Man geht davon aus, daß diese Längenänderung von SSRs auf *DNA slippage* beruht (Kornberg et al., 1964; Streisinger et al., 1966). Ein Strang der DNA dissoziiert dabei während der DNA-Replikation oder der DNA-Reparatur von seinem Gegenstrang und lagert sich anschließend, um eine oder wenige Einheiten versetzt, wieder an. In einer enzym- und sequenzspezifischen Reaktion ließen sich sowohl Di- als auch Trinukleotid-Sequenzen auf diese Art *in vitro* verlängern (Schlötterer und Tautz, 1992; Behn-Krappa und Doerfler, 1994). Ein ungleiches *crossing over* während der Rekombination gilt als der andere mögliche Mechanismus, aufgrund dessen die Länge von SSRs variieren könnte (Smith, 1974; Levinson und Gutman, 1987). Bei der Mutation von Minisatelliten findet auch Genkonversion statt, die auch bei der Ausdehnung von Mikrosatelliten eine Rolle spielen könnte (Jeffreys et al., 1994).

A.2.2 Instabilität von Mikrosatelliten in Tumoren

In einigen spontan auftretenden Tumoren sowie in erblichen Nicht-polyposen Kolonkarzinomen (HNPCC) beobachtet man genomweite Instabilitäten von Mikrosatelliten (Thibodeau et al., 1993; Ionov et al., 1993; Sturzeneker et al., 2000). Diese erhöhte Mutationsrate beruht wahrscheinlich auf Funktionsverlusten von DNA-Reparaturenzymen (Duval und Hamelin, 2002a). Tumore mit Instabilitäten von Mikrosatelliten können von denen, in denen Mikrosa-

telliten stabil bleiben, durch ihre klinisch-pathologischen Eigenschaften unterschieden werden (Duval und Hamelin, 2002b). Die genomweiten Veränderungen der Länge von Mikrosatelliten gehen wahrscheinlich mit weiteren inaktivierenden Mutationen in anderen Genen einher, durch die der Tumorverlauf beeinflusst werden kann.

A.2.3 Mögliche Folgen der Instabilität von Mikrosatelliten

Je nach Anzahl der Wiederholungen von SSR innerhalb von translatierten Exons entstehen verschieden lange Aminosäure-Ketten. Vor allem von einigen Transkriptionsfaktoren ist bekannt, daß sie Polyglutamin- oder Polyprolin-Ketten enthalten. Die Aktivierung der durch diese Transkriptionsfaktoren regulierten Gene wird durch die Veränderung der Anzahl dieser Aminosäuren beeinflusst (Gerber et al., 1994; Übersicht in Künzler et al., 1995). *DNA slippage* trägt möglicherweise auch zu der Erweiterung des Antikörper-Repertoires bei, da Trinukleotid-Wiederholungen in den variablen Regionen der schweren Immunglobulin-Ketten angereichert sind (Lantto und Ohlin, 2002).

In Promotoren können SSRs in Abhängigkeit ihrer Länge die Aktivität dieser Promotoren beeinflussen (Hamada et al., 1984; Übersichten in Kashi et al., 1997; Rockman und Wray, 2002; Borrmann et al., 2003). So reprimiert z.B. eine 5'-d(CTG)₂₅-3'-Folge den Metallothionein-III-Promotor der Maus (Imagawa et al., 1995), während eine 5'-(TC)_n-3'-Wiederholung die Expression des HMGA2-Promotors in menschlichen Zelllinien verstärkt (Borrmann et al., 2003). SSRs können als Bindungsstellen für Proteine dienen und so die Zusammensetzung von Transkriptionsfaktoren an Promotoren verändern. Die Bindung von Proteinen könnte dabei sequenzspezifisch oder strukturspezifisch erfolgen, da einige SSRs DNA-Strukturen ausbilden können, die von der B-DNA-Struktur deutlich abweichen (Taillandier et al., 1984; Sinden, 1994; Cox und Mirkin, 1997). Zudem verändern einige SSRs die lokale Chromatinstruktur, da bei repetitiver DNA z.B. die Affinität zu Nukleosomen DNA stark sequenzabhängig ist (Wang und Griffith, 1995; Godde et al., 1996). Die Transkriptionsrate eines benachbarten Gens könnte durch diese strukturellen Effekte moduliert werden (zur Übersicht in Brahmachari et al., 1995).

A.3 Fragile Stellen

Eine starke Ausdehnung von SSRs kann unter bestimmten Zellkulturbedingungen zur Ausbildung von lichtmikroskopisch sichtbaren, nicht anfärbbaren Bereichen auf Chromosomen, den sogenannten fragilen Stellen, führen (Sutherland und Richards, 1995). Man unterscheidet häufige und seltene fragile Stellen, die zudem anhand der Stoffe, die sie auslösen, weiter klassifi-

ziert werden. Die molekulare Ursache der häufigen fragilen Stellen ist bisher nicht bekannt (Sutherland und Richards, 1999). Fragile Stellen scheinen Regionen genomischer Instabilität aufzuzeigen, da sie mit erhöhten Raten von Schwesterchromatid-Austauschen, mit Verlusten der Heterozygotie in Tumoren oder mit viralen Integrationen assoziiert sind (zur Übersicht Glover, 1998; Sutherland und Baker, 2000; Sutherland und Richards, 1999).

Die bisher klonierten seltenen, durch Folatmangel induzierten, fragilen Stellen beruhen auf der starken Amplifikation und Methylierung von CGG-Wiederholungen. So führt die fragile Stelle A auf dem X-Chromosom zum Fragilen-X-Syndrom, von dem später noch die Rede sein wird (A.5.1.). Die seltene, CGG-induzierte fragile Stelle B auf Chromosom 11 (Abb. 1) wird mit Chromosomenbrüchen im Jacobsen-Syndrom in Verbindung gebracht (Jones et al., 1995).

A.4 Amplifikation von Mikrosatelliten bei Erbkrankheiten des Menschen

Die Amplifikation bestimmter Mikrosatelliten kann beim Menschen eine Reihe von Erbkrankheiten auslösen. Bislang sind mehrere Loci bekannt, in denen die Verlängerung einer repetitiver DNA-Folge ein Genprodukt bzw. dessen Expression beeinträchtigt (Übersicht in Margolis und Ross, 2001; Margolis et al., 1999; Ranum und Day, 2002; Cummings und Zoghbi, 2000). Meist handelt es sich beim amplifizierten Element um Trinukleotid-Wiederholungen. Es gibt jedoch auch Erkrankungen, die mit der Amplifikation von Tetra- und Pentanukleotidwiederholungen sowie einer Dodecamer-Wiederholung assoziiert sind (Abb. 1). Im Vergleich zur der genomweiten Instabilität von Mikrosatelliten in Tumoren (A.2.2) ist bei den pathogenen Amplifikationen von Mikrosatelliten in einem Gen-Locus eine drastische Zunahme der Anzahl der wiederholten einfachen Sequenzeinheit die Regel (Wells, 1996).

In einigen Proteinen beobachtet man eine geringfügige Verlängerung oder auch Verkürzung von Polyalanin- oder Polyaspartat-Folgen (Mundlos et al., 1997; Muragaki et al., 1996; Brais et al., 1998; Délot et al., 1999; Utsch et al., 2002; Nakamoto et al., 2002). Da diese Krankheiten wahrscheinlich nicht zu den Amplifikations-Krankheiten gehören, wird hier nicht weiter auf sie eingegangen.

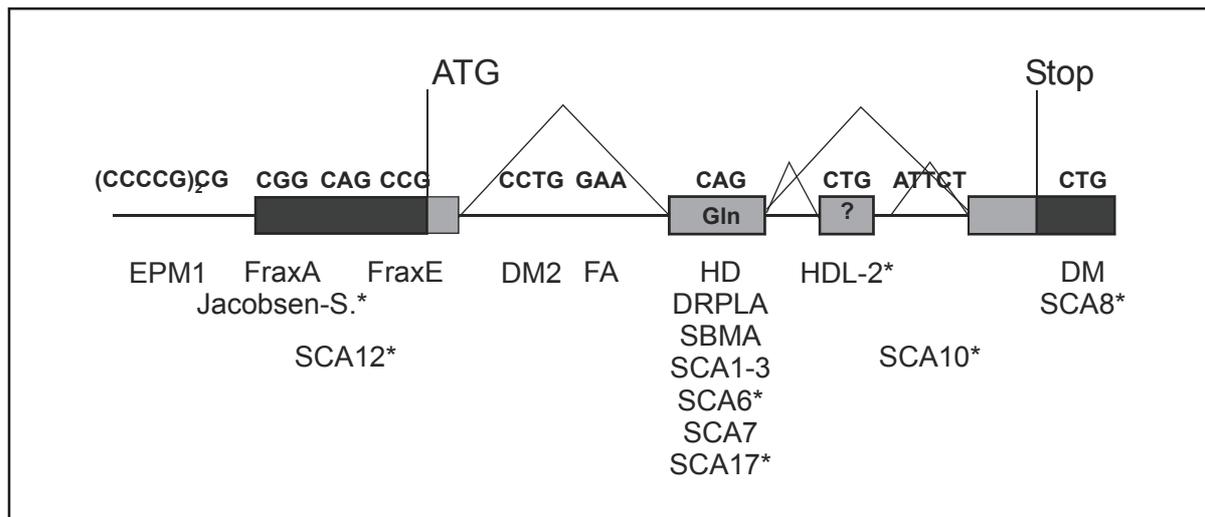


Abbildung 1 Position und Sequenz der pathogenen Mikrosatelliten und der kodierten Aminosäuren in einem schematisierten menschlichen Gen.

Die nicht-translatierten Regionen der Exons (Rechtecke) sind schwarz, die translatierten Bereiche grau gekennzeichnet. Dreieckige Verbindungslinien zwischen den Exons deuten auf Spleißen hin. Die Konsequenz der Amplifikation der 5'-d(CTG)_n-3'-Wiederholung bei der HDL-2 ist bislang aufgrund der Lage des Mikrosatelliten in einem alternativ gespleißten Exon, das nur bedingt translatiert wird, ungeklärt (Ranum und Day, 2002). Bei den mit Stern gekennzeichnete Krankheiten sind Amplifikationen bislang nur sehr selten als pathogene Ursache festgestellt worden (Margolis und Ross, 2001). Die Abkürzungen der Amplifikations-Krankheiten sind unterhalb der Sequenz des Mikrosatelliten eingetragen. DM Myotone Dystrophie; DRPLA Dentatorubrale pallidoluyische Atrophie; EPM1 progressive myoklone Epilepsie Typ 1; FraXA/E Fragiles-X-Syndrom Typ A oder E; FA Friedreichsche Ataxie; HD Huntingtonsche Krankheit; HDL-2 HD-ähnliche Erkrankung Typ 2; SBMA Spinobulbäre Muskelatrophie; SCA Spinozerebelläre Ataxie. (Abbildung verändert nach Sinden et al., 2002)

A.4.1 Amplifikation von Polyglutamin-Ketten

Man unterscheidet je nach Position der repetitiven DNA innerhalb eines Gens zwei Klassen von Amplifikations-Krankheiten (Paulson und Fischbeck, 1996). Bei den Krankheiten des Typs I liegen die amplifizierte Trinukleotide, meist CAG, im translatierten Bereich des Gens und führen zu einem veränderten Protein. Neben der Entdeckung, daß eine instabile 5'-d(CAG)_n-3'-Wiederholung für Chorea Huntington (The Huntington's Disease Collaborative Research Group, 1993) verantwortlich ist, wurden ähnlich instabile Trinukleotide als ursächlich für die Spinobulbäre Muskelatrophie (SBMA; La Spada et al., 1992), die Dentatorubrale Pallidoluyische Atrophie (DPRLA; Koide et al., 1994) und mehrere Subtypen der Spinozerebellären Ataxie (SCA) beschrieben. Von insgesamt 17 Subtypen solcher Ataxien wurden SCA1 (Orr et al., 1993); SCA2 (Imbert et al., 1996); SCA3 (Kawaguchi et al., 1994), SCA6 (Zhuchenko et al., 1997), SCA7 (David et al., 1997) und wahrscheinlich auch SCA17 (Nakamura et al., 2001) mit 5'-d(CAG)_n-3'-Wiederholungen in Verbindung gebracht. Eine

sehr seltene Ausdehnung einer 5'-d(CTG)_n-3'-Wiederholung im Junctophilin-Gen 3 bewirkt wahrscheinlich nur bei Menschen afrikanischer Abstammung einen Huntington-ähnlichen Phänotyp (Huntington *disease-like 2*; HDL-2) (Holmes et al., 2001; Bauer et al., 2002).

Neben der Trinukleotid-Wiederholung 5'-d(CAG)_n-3', die für Polyglutamin kodiert, weisen die betroffenen Proteine keine Ähnlichkeiten auf (Cummings und Zoghbi, 2000). Dennoch gleichen sich die ausgelösten Symptome, so daß man eine gemeinsame Pathogenese vermutet. Bei allen Polyglutamin-Erkrankungen beobachtet man eine Neurodegeneration, die typischerweise erst später im Verlauf des Lebens beginnt (Koshy und Zoghbi, 1997). Je öfter Glutamin wiederholt ist, desto jünger sind die Betroffenen bei den ersten Symptomen (Rubinsztein et al., 1993). Zudem korreliert die Schwere der Symptome mit der Länge der Polyglutamin-Kette. Die Krankheiten der Klasse I werden vermutlich nicht durch den Verlust der jeweiligen Proteinfunktion ausgelöst, sondern beruhen wahrscheinlich auf einer durch die Amplifikation der Polyglutamin-Folge erworbenen zytotoxischen Eigenschaft der veränderten Proteine (*gain-of-function*-Mutation) (Zoghbi und Orr, 1999; Margolis und Ross, 2001).

A.4.2 Amplifikation von Mikrosatelliten in nicht-translatierten Regionen von Genen

Die Amplifikations-Erkrankungen der Klasse II sind heterogener als die der Klasse I. Neben den verschiedenen Ätiologien unterscheiden sich die amplifzierten Mikrosatelliten in ihrer Sequenz, in der Länge der wiederholten Einheit und in ihrer Position innerhalb des betroffenen Gens (Abb. 1). Zunächst wurde die Amplifikation einer 5'-d(CGG)_n-3'-Wiederholung mit dem Fragilen-X-Syndrom (FraXA) in Verbindung gebracht (Fu et al., 1991; Verkerk et al., 1991). Es handelt sich dabei um ein Syndrom, das wie das FraXE-Syndrom auch (Knight et al., 1993), mit geistiger Behinderung der Betroffenen einher geht. Des weiteren gehören die Myotonen Dystrophien 1 und 2 (Brook et al., 1992; Liquori et al., 2001), die Friedreichsche Ataxie (Campuzano et al., 1996), die Spinozerebellären Ataxien des Typs 8 und 12 (Koob et al., 1999; Holmes et al., 1999) und die Progressive Myoklonische Epilepsie Typ 1 (La Freniere et al., 1997) zu dieser Klasse von Amplifikations-Erkrankungen.

Allen Klasse II-Erkrankungen ist gemeinsam, daß das vom betroffenen Gen kodierte Protein nicht von der Amplifikation der Tri-, Tetra-, Penta- und Dodecamernukleotid-Wiederholungen betroffen ist (Margolis et al., 1999; Ranum und Day, 2002). Zudem resultieren krankheitsauslösende Mutationen aus Amplifikationen von sogenannten Prämutationen, die bereits im Vergleich zu den normalen Allelen amplifiziert vorliegen, jedoch nicht die Symptome einer starken Expansion nach sich ziehen (Cummings und Zoghbi, 2000). In der

Regel werden bei den Amplifikationskrankheiten der Klasse II die Sequenz-Wiederholungen stärker amplifiziert, als bei den Polyglutamin-Erkrankungen (La Spada, 1997; Margolis et al., 1999).

Die Folgen der Amplifikation für das betroffene Gen unterscheiden sich ebenfalls bei allen Klasse II-Erkrankungen. Im Falle der Fragilen-X-Syndrome A und E beobachtet man eine Hypermethylierung der Promotorregionen der Gene FMR1 und FMR2 (*fragile X mental retardation gene 1* und 2) und einen Verlust der Genexpression (Oberlé et al., 1991; Sutcliffe et al., 1992). Bei der Friedreichschen Ataxie beeinträchtigt eine 5'-d(GAA)_n-3'-Wiederholung im Intron des Frataxin-Gens die Elongation bei der Transkription (zur Übersicht Grabczyk et al., 2001). Die Amplifikation bei beiden Typen der Myotonen Dystrophie scheint die Prozessierung des Primärtranskriptes zu beeinträchtigen oder sogar die normale Interaktion von RNA-bindenden Proteinen mit ihren Ziel-RNAs dominant-negativ zu verändern (Davis et al., 1997; Tapscott et al., 1998; zur Übersicht Ranum und Day, 2002). Man spricht hierbei von einem *gain-of-function*-Effekt auf RNA-Ebene, einem bislang unbekanntem pathogenen Mechanismus bei Erbkrankheiten des Menschen (Galvao et al., 2001). Bei der Myotonen Dystrophie Typ 1 könnte es sich zudem um eine Störung der Expression mehrerer benachbarter Gene handeln (Harris et al., 1996). Ähnliche toxische RNA-Effekte und eine gleichzeitige Störung der Expression der betroffenen Gene werden auch bei den Spinocerebellären Ataxien SCA Typ 8, 10 und 12 vermutet (Ranum und Day, 2002).

A.4.3 Dynamische Mutation pathogener Mikrosatelliten

Mutationen verändern die genomische Information eines Organismus. Im Falle einer Mutation in der Keimbahn wird die Mutation an die nächste Generation weitergegeben. In der Regel sind Mutationen statisch, d.h. die DNA einer Zelle, die eine Mutation trägt, mutiert mit der gleichen Wahrscheinlichkeit wie die DNA der Zelle, von der sie abstammt. Bei den krankheitsrelevanten Mikrosatelliten erhöht sich dagegen die Wahrscheinlichkeit einer weiteren Amplifikation drastisch, sobald die normale Länge eines Allels überschritten ist. Es wurde daher der Begriff der dynamischen Mutation für die Ausdehnung der pathogenen Mikrosatelliten geprägt (Richards et al., 1992; Richards und Sutherland, 1992).

Schon bei der Vererbung von nicht-pathogenen Mikrosatelliten beobachtet man eine zunehmende Mutationsanfälligkeit von solchen Allelen, in denen die Einheit oft wiederholt wird (Wierdl et al., 1997; Brinkmann et al., 1998). Das unbegrenzte Wachstum der meisten Mikrosatelliten wird offenbar verhindert durch eine mit der Länge der Wiederholung zunehmenden Wahrscheinlichkeit der Kontraktion des Allels (Xu et al., 2000; Ellegren, 2000; Harr und

Schlötterer, 2000). Es ist bisher nicht bekannt, was die pathogenen Mikrosatelliten von anderen Mikrosatelliten unterscheidet und warum der Prozeß der Amplifikation bei ihnen weiter fortschreitet.

Die Zunahme der Anzahl der wiederholten Einheiten lieferte die Erklärung für das Phänomen der Antizipation, das typisch ist für viele der oben beschriebenen Amplifikations-Erkrankungen. Vor allem bei den Erkrankungen des Typs I nimmt das Manifestationsalter in der Folge der Generationen ab und die Schwere der Erkrankung zu (Rubinsztein et al., 1993; Cummings und Zoghbi, 2000; Zoghbi und Orr, 1999; Ashley und Warren, 1995).

Bei dem Fragilen-X-Syndrom (A.5.1.) gibt es keine vergleichbare Genotyp/Phänotyp-Korrelation. Diese X-chromosomal vererbte Erkrankung zeichnet sich jedoch durch einen ungewöhnlichen Vererbungsmodus (Sherman et al., 1984; 1985) aus, der als Sherman-Paradox bezeichnet wurde (Opitz et al., 1986). Vom Fragilen-X-Syndrom selbst nicht betroffene Überträger vererben die krankheitsauslösende Mutation ausschließlich über ihre Töchter an ihre Enkel. Bei diesen Überträgern liegt die polymorphe 5'-d(CGG)_n-3'-Wiederholung bereits über einen Schwellenwert hinaus verlängert, also prämutiert, vor (Fu et al., 1991). Nach der Passage in der weiblichen Keimbahn (A.5.1.8.1.) wird eine Prämutation meistens wiederum über einen Schwellenwert hinaus verlängert, so daß in der folgenden Generation das Krankheitsbild des Fragilen-X-Syndroms manifest wird (Verkerk et al., 1991). Auf die Mutation dieser 5'-d(CGG)_n-3'-Wiederholung wird in den folgenden Kapiteln genauer eingegangen (A.5.1.).

A.5 Amplifikation von 5'-d(CGG)_n-3'-Trinukleotid-Folgen

Auf dem X-Chromosom sind drei durch Folatmangel induzierte fragile Stellen (A, E und F) bekannt, die auf der Amplifikation von 5'-d(CGG)_n-3'-Wiederholungen beruhen. Vor der Entdeckung der molekularen Ursache des Fragilen-X-Syndroms (Verkerk et al., 1991) diente die fragile Stelle A als diagnostischer Marker für das Syndrom (Turner et al., 1978). Während die fragile Stelle E mit einer milden Form der geistigen Behinderung assoziiert ist (Knight et al., 1993), wurde die fragile Stelle F bisher mit keiner Erbkrankheit in Verbindung gebracht (Parrish et al., 1994; Ritchie et al., 1994). Weitere Expansionen von 5'-d(CGG)_n-3' liegen auf den Chromosomen 11 und 16. Das Jacobsen-Syndrom (Fra11B) hängt möglicherweise mit einer schweren Form der geistigen Behinderung zusammen (Jones et al., 1995). Negative Folgen der Fra16A sind nicht bekannt (Nancarrow et al., 1994). Im folgenden soll das Fragile-X-Syndrom exemplarisch für eine pathogene Amplifikation von 5'-d(CGG)_n-3'-Wiederholungen vorgestellt werden.

A.5.1 Das Fragile-X-Syndrom oder Martin-Bell-Syndrom

A.5.1.1 Vorkommen und Symptome

Das Fragile-X-Syndrom (FraXA) war die erste Erbkrankheit beim Menschen, bei der eine massive Amplifikation einer Trinukleotid-Folge beobachtet wurde (Fu et al., 1991; Verkerk et al., 1991). Es tritt bei 1:4000 Männern und 1:7000 Frauen auf und zählt zu den häufigsten Ursachen für vererbte geistige Behinderung bei Männern (Crawford et al., 2001). Weitere Symptome sind Makroorchidie und eine charakteristische Physiognomie mit einem schmalen Gesicht, großen Ohren und einer vorstehenden Stirn (Martin und Bell, 1943; zur Übersicht O'Donnell und Warren, 2002).

A.5.1.2 Eine polymorphe 5'-d(CGG)_n-3'-Wiederholung in der 5'-nicht-translatierten Region

In der 5'-nicht-translatierten Region (5'-UTR) des FMR1-Gens liegt eine polymorphe 5'-d(CGG)_n-3'-Trinukleotid-Folge (Ashley et al., 1993). Man kennt vier Allel-Klassen des FMR1-Gens: Normale Allele weisen zwischen 5 und etwa 44 (Maddalena et al., 2001), intermediäre FMR1-Allele zwischen 45 und 60 5'-d(CGG)-3'-Einheiten auf (Nolin et al., 2003). Die Klasse der Prämutations-Allele umfaßt etwa 55 bis 200 5'-d(CGG)_n-3'-Wiederholungen. Intermediäre und prämutierte Allele können erst anhand ihres Verhaltens bei der Vererbung unterschieden werden. Während Prämutations-Allele bei der Übertragung über die weibliche Keimbahn sehr instabil sind (A.5.1.8.1), bleiben intermediäre Allele in der Regel stabil (Nolin et al., 2003). Vollmutations-Allele weisen über 200 Trinukleotid-Einheiten auf. Die Länge dieser Triplet-Wiederholung kann sich in den verschiedenen Zellen und Geweben eines Individuums unterscheiden (Nolin et al., 1994; zur Übersicht Brown, 2002). Während in Spermien von Vollmutationsträgern fast immer nur die Prämutation vorkommt (Reyniers, 1993), sind die Gewebe von etwa 40% der Vollmutationsträger Mosaik aus Zellen mit unterschiedlich langen Trinukleotid-Wiederholungen („Längenmosaizismus“) (zur Übersicht Brown, 2002).

A.5.1.3 Inaktivierung des FMR1-Gens

In den meisten Vollmutationsträgern ist das FMR1-Gen komplett methyliert (Pieretti et al., 1991; Sutcliffe et al., 1992), das umgebende Chromatin liegt in einer inaktiven Konformation vor (Coffee et al., 1999; 2002) und das Gen wird nicht mehr transkribiert (Pieretti et al., 1991). In einigen Patienten wird noch FMR1-mRNA gebildet, jedoch in einem geringeren Maß als bei Normalpersonen (Tassone 2000a; 2000b; 2001). Die Transkripte werden bei den

Patienten nicht translatiert, wahrscheinlich weil die lange Trinukleotid-Folge in der 5'-UTR die Translation des Transkriptes verhindert (Feng et al., 1995). Deletionen des FMR1-Gens oder Punktmutationen im offenen Leseraster (ORF) führen ebenfalls zum Fragilen-X-Syndrom, so daß davon ausgegangen wird, daß der Verlust des FMR1-Proteins ursächlich für die Erkrankung ist (DeBouille et al., 1993; Hammond et al., 1997). Eine *knock-out*-Maus des murinen *fmr1*-Gens zeigt zudem einige Aspekte der Symptome des Fragilen-X-Syndroms (The Dutch-Belgian Fragile X-Consortium, 1994; Kooy, 2003).

A.5.1.4 Methylierung des FMR1-Gens

Bislang ist nicht geklärt, wie die Amplifikation der 5'-d(CGG)_n-3'-Wiederholung mit der Inaktivierung des FMR1-Gens zusammenhängt. Während bei fast allen Vollmutationsträgern das FMR1-Gen methyliert vorliegt und keine Proteinexpression stattfindet (Oberlé et al., 1991; Hornstra et al., 1993), sind auch Vollmutationsträger bekannt, die nicht geistig behindert sind (Smeets et al., 1995; zur Übersicht Hagerman, 2002). Bei diesen sogenannten *high functioning males* ist das FMR1-Gen nicht methyliert, es wird transkribiert und das Transkript wird auch translatiert (Hagermann et al., 1994; Tassone 1999). Zusätzlich zu dem Mosaizismus der Längen der 5'-d(CGG)_n-3'-Wiederholung in verschiedenen Geweben beobachtet man auch einen Methylierungsmosaizismus, d.h. das FMR1-Gen ist nicht in allen Zellen komplett oder im gleichen Muster methyliert (Stöger et al., 1997; Genç et al., 2000; Tassone et al., 2002). Die Amplifikation der 5'-d(CGG)_n-3'-Folge reicht also alleine als Signal für eine komplette Methylierung, bzw. Inaktivierung des Gens nicht aus. Analysen mit *in-vitro*-methylierten Promotorkonstrukten bzw. Reaktivierungsversuche mit demethylierenden Agenzien zeigten, daß die Methylierung der Promotorregion das FMR1-Gen inaktiviert (Sutcliffe et al., 1992; Genç et al., 2000; Chiurazzi und Neri, 2001). Die Assoziation des FMR1-Gens mit deacetylierten Nukleosomen (Chiurazzi et al., 1999; Coffee et al., 1999; 2002) bestätigt den transkriptionell inaktiven Zustand des FMR1-Gens in Vollmutationsträgern (El-Osta und Wolffe, 2000).

Ungeklärt ist bislang, was die Methylierung des FMR1-Gens auslösen könnte. Es wird diskutiert, daß die DNA-Methyltransferasen eine hohe Affinität zu Haarnadelstrukturen besitzen könnten, wie sie von einzelsträngigen 5'-d(CGG)_n-3'-Wiederholungen angenommen werden können (Smith et al., 1994; Chen et al., 1995; Chen et al., 1998, Bacolla et al., 2001). Möglicherweise wird die amplifizierte Trinukleotid-Folge auch als fremd erkannt und deshalb *de novo* methyliert (Doerfler, 1996; Doerfler et al., 2001).

A.5.1.5 Funktionen des FMR1-Proteins

Das FMR1-Protein (FMRP) enthält mehrere RNA-bindende Domänen und bindet bevorzugt an RNAs, die eine Sekundärstruktur aus vier Guanosinen ausbilden, das sogenannte G-Quartett (Verheij et al., 1993; Siomi et al., 1993; 1994, Darnell et al., 2001; zur Übersicht Sung et al., 2000). FMRP ist mit aktiv translatierenden Polyribosomen assoziiert (Corbin et al., 1997). Da es zudem eine Kernlokalisationssequenz sowie ein Kernexportsignal enthält, schreibt man FMRP Aufgaben beim Transport bestimmter RNAs aus dem Zellkern und bei der Regulation der Translation dieser Transkripte zu (zur Übersicht Bardoni und Mandel, 2002).

A.5.1.6 Auswirkungen einer Prämutation im FMR1-Gen

Eine Prämutation des FMR1-Gens führt zwar nicht zum Fragilen-X-Syndrom, ist jedoch nicht bei allen Trägern asymptomatisch (zur Übersicht Hagerman und Hagerman, 2002). Bei etwa 20% der Prämutationsträgerinnen wird z. B. eine verfrüht einsetzende Menopause diagnostiziert (Allingham-Hawkins et al., 1999). In Abhängigkeit der Länge der Triplet-Wiederholung beobachtet man zudem eine verstärkte Transkription des FMR1-Gens (Tassone et al., 2000c). Das Transkript wird dabei jedoch weniger effizient translatiert (Tassone et al., 2000c; Kenneson et al., 2001; Primerano et al., 2002), so daß insgesamt nicht mehr Protein exprimiert wird als Zellen mit normalen Allelen. Man vermutet wegen der verstärkten Transkription des Gens und wegen intranukleärer Einschlüsse, die in einer Gruppe von Prämutationsträgern mit neurodegenerativen Symptomen beobachtet wurden (Hagerman et al., 2001; Greco, 2002; Primerano et al., 2002), bei der Prämutation einen toxischen RNA-Effekt (A.4.2) (zur Übersicht Bardoni und Mandel, 2002; Hagerman und Hagerman, 2002). Ein Dosiseffekt des FMRPs ist angesichts der nicht überlappenden Symptome bei Prämutationsträgern und Vollmutationsträgern unwahrscheinlich.

A.5.1.7 Zeitpunkt der Amplifikation

Vollmutationen gehen fast immer auf eine Prämutation in der Mutter zurück (zur Übersicht Maddalena et al., 2001). Es stellt sich die Frage, wann die Amplifikation des Prämutations-Allels stattfindet: während der meiotischen Teilung in der Oogenese oder bei den mitotischen Zellteilungen im Embryo. Ein Mosaik aus Zellen mit unterschiedlich langen Trinukleotid-Folgen am FMR1-Gen (A.5.1.2.) könnte sowohl auf mehreren Amplifikationen eines Prämutations-Allels, als auch auf mehreren Kontraktionen eines bereits amplifizierten Vollmutations-Allels beruhen (Nolin et al., 1994; Nelson und Warren, 1993). Zur

Beantwortung dieser Frage müßten Oozyten von Prämutationsträgerinnen untersucht werden. Zur Zeit gelten sowohl eine meiotische Instabilität während der Oogenese als auch eine zeitlich begrenzte mitotische Instabilität während der Embryogenese als wahrscheinlich (Reyniers et al., 1993, Wöhrle et al., 1993; Malter et al., 1997; Moutou et al., 1997).

A.5.1.8 Faktoren der Instabilität von 5'-d(CGG)_n-3'-Wiederholungen im Fragilen-X-Syndrom

Seit der Entdeckung der molekularen Ursache des Fragilen-X-Syndroms im Jahr 1991 wurden stabile sowie instabile Transmissionen dieser Trinukleotid-Wiederholung auf Gesetzmäßigkeiten hin untersucht. Im folgenden Kapitel über die bisher identifizierten Parameter liegt der Schwerpunkt auf der Amplifikation der 5'-d(CGG)_n-3'-Wiederholung im FMR1-Gen. Erkenntnisse aus den Studien anderer instabiler Mikrosatelliten werden nur zur Verdeutlichung bestimmter Konzepte erwähnt.

A.5.1.8.1 Keimbahnpassage

Bei der Übertragung eines Prämutations-Allels von der Mutter auf ihre Kinder finden fast ausschließlich Expansionen der 5'-d(CGG)_n-3'-Wiederholung statt (Nolin et al., 1996; Ashley-Koch et al., 1998). In den meisten Fällen amplifiziert eine Prämutation dabei zur Vollmutation (A.4.2). Im Gegensatz dazu kann eine vergleichbar lange Trinukleotid-Folge bei der Vererbung durch den Vater geringfügig expandieren, kontrahieren oder unverändert weitergegeben werden (Nolin et al., 1996). Eine genomische Prägung der FMR1-Allele wird derzeit diskutiert (Sullivan et al., 2002; Nolin et al., 2003; Hundscheid et al., 2000).

In einigen Polyglutamin-Erkrankungen, z.B. bei der Spinozerebellären Ataxie 1, treten krankheitsauslösende Zunahmen der Allelgrößen verstärkt bei der Vererbung durch den Vater auf (Orr et al., 1993). Eine generelle Aussage über eine höhere Stabilität von Mikrosatelliten in der maternalen oder paternalen Keimbahn sind somit nicht möglich (La Spada, 1997).

A.5.1.8.2 Länge der Repetition

Je länger die 5'-d(CGG)_n-3'-Wiederholung am FMR1-Locus ist, desto instabiler verhält sie sich während der Transmission (Snow et al., 1993; Fisch et al., 1995; Nolin et al., 1996; Ashley-Koch et al., 1998). Dabei sind Kontraktionen der 5'-d(CGG)_n-3'-Wiederholung seltene Ausnahmen (Brown et al., 1996; Nolin et al., 2003). Die Wahrscheinlichkeit einer Amplifikation zu einer Vollmutation liegt bei Prämutations-Allelen mit 59 Trinukleotid-Wiederholungen noch bei 4%, während sie bei Prämutations-Allelen mit mehr als 80 Trinukleotiden auf 60% ansteigt. Allele mit mehr als 140 repetierten Einheiten amplifizieren immer zu einer

Vollmutation (Nolin et al., 2003). Die kleinste Prämutation, die sich bisher innerhalb einer Generation zu einer Vollmutation entwickelte, wies 59 ununterbrochene 5'-d(CGG)_n-3'-Wiederholungen auf (Nolin et al., 2003). Der Übergang von einem normalen Allel zu einem Prämutations-Allel wurde noch nie beobachtet (Nolin et al., 2003).

Die Abhängigkeit der Instabilität von der Anzahl der wiederholten Einheit trifft auch für die anderen Amplifikations-Krankheiten zu. Der Schwellenwert zur Instabilität und die Zahl der pathogenen Wiederholungen ist bei allen bisher bekannten Amplifikations-Erkrankungen unterschiedlich (zur Übersicht Margolis et al., 1999).

A.5.1.8.3 Unterbrechungen der Repetition

Die Anzahl der Trinukleotide ist aber nicht alleine für die Instabilität einer 5'-d(CGG)_n-3'-Wiederholung verantwortlich. Sequenzunterbrechungen innerhalb der Trinukleotid-Folge gelten als wichtigstes Stabilitätskriterium (Eichler et al., 1994; Hirst, 1995; Kunst und Warren, 1994). Ein 5'-d(AGG)-3'-Trinukleotid unterbricht die 5'-d(CGG)_n-3'-Folge meist nach jedem 9. oder 10. Triplet. Die Anzahl der Unterbrechungen und ihre Position(en) innerhalb der Trinukleotid-Folge trägt zu der hohen Polymorphie des FMR1-Gens bei (zur Übersicht Sherman, 2002). Intermediäre Allele ($\sim 46 < n < 60$) werden meist stabil vererbt, sobald jedoch 34 bis 50 Trinukleotide ohne Unterbrechung tandem-artig wiederholt vorliegen, kann man bei Transmission durch beide Elternteile bereits Veränderungen um wenige Einheiten beobachten (Eichler et al., 1994; Nolin et al., 1996). Die Unterbrechung der reinen 5'-d(CGG)_n-3'-Wiederholungen durch 5'-d(AGG)-3' scheint somit einen stabilisierenden Einfluß auf die Trinukleotid-Folge auszuüben. Ähnliche stabilisierende Unterbrechungen der repetitierten Einheit finden sich auch bei anderen Amplifikations-Erkrankungen, z.B. bei den Spinozerebellären Ataxien Typ 1 und 2 (zur Übersicht La Spada, 1997).

A.5.1.8.4 *Cis*- und *trans*-aktive Faktoren

Aus Kopplungsanalysen beim Fragilen-X-Syndrom ergeben sich Hinweise auf eine Anreicherung amplifizierter Allele in bestimmten Haplotypen, die bei gesunden Kontrollpersonen seltener vorkommen (Gunter et al., 1998; Crawford et al., 2000). Die Natur des deshalb vermuteten *cis*-aktiven Faktors auf die Instabilität ist noch nicht geklärt. Außerdem unterscheidet sich das Ausmaß der Instabilität bei Familienangehörigen weniger als bei nicht verwandten Personen (Nolin et al., 1996; Burman et al., 2000). Ein ebenfalls noch nicht identifizierter *trans*-aktiver Faktor könnte dafür verantwortlich sein.

Der Vergleich von flankierenden Sequenzen mehrerer instabiler 5'-d(CAG)_n-3'-Loci deutet auf einen Zusammenhang zwischen Instabilität und dem GC-Gehalt der flankierenden Sequenzen hin (Brock et al., 1999). Die Instabilität der 5'-d(CAG)_n-3'-Folge bei der Huntingtonschen Erkrankung ist möglicherweise durch die der Repetition direkt benachbarten Sequenz beeinflusst (Rubinsztein et al., 1993). Die amplifizierte Allele der Myotonen Dystrophie Typ 1 sind außerdem alle mit einem spezifischen Haplotypen assoziiert (Imbert et al., 1993; Neville et al., 1994), so daß ein dort *cis*-aktiver Einfluß flankierender Sequenzen auf die Instabilität wahrscheinlich ist (Mirkin und Smirnova, 2002).

A.5.1.8.5 DNA-Methylierung

Der Einfluß des Methylierungszustandes auf die Stabilität von amplifizierten 5'-d(CGG)_n-3'-Wiederholungen wurde in primären Fibroblasten und in Zellhybriden untersucht. Methylierte, amplifizierte Allele des FMR1-Locus waren während ihrer Kultur *in vitro* häufig – aber nicht immer (Burman et al., 1999) – stabiler als nicht methylierte Allele (Wöhrle et al., 1993; Gläser et al., 1999; Salat et al., 2000; Nichol und Pearson, 2002).

A.6 Überlegungen zum Mechanismus der pathogenen Amplifikation von Mikrosatelliten

Der Mechanismus der Amplifikation von pathogenen Mikrosatelliten ist bis heute nicht geklärt. In Mäusen wurden vergleichbare Amplifikationen von Trinukleotid-Wiederholungen bei ihrer Vererbung meist nicht beobachtet (Bontekoe et al., 1997; Lavedan et al., 1997; van den Broek et al., 2002; Mangiarini et al., 1997; Sato et al., 1999). Nach ihrer Transmission weisen selbst sehr lange Allele nur eine geringe Änderung der Allelgrößen auf. Der Schwellenwert zur Instabilität scheint bei Mäusen wesentlich höher zu sein als beim Menschen (Bontekoe et al., 2001; Lorenzetti et al., 2000; Fortune et al., 2000).

Die Replikation von einigen Trinukleotid-Wiederholungen *in vitro*, in Bakterien oder in Hefen lieferten jedoch Hinweise auf mögliche Faktoren, die die meiotische bzw. mitotische Instabilität von Trinukleotiden beeinflussen könnten.

A.6.1 DNA *slippage* und Sekundärstrukturen

Die massive Amplifikation repetitiver Elemente wurde erst nach der Entdeckung des Fragilen-X-Syndroms als neuer Mutationsmechanismus erkannt. Sie war in einfachen Modellsystemen bis dahin nicht beobachtet worden (zur Übersicht Sinden et al., 2002). als wahrscheinlichster Mechanismus für die Zunahme an repetierten Einheiten gilt ein Verrutschen eines DNA-Stranges während der Replikation, der Reparatur oder

Rekombination (*DNA slippage*, A.2.1) (Kornberg et al., 1964: 1965; Streisinger et al., 1966). Beim Menschen haben die pathogenen Mikrosatelliten eine starke Neigung zur Expansion. Bei Bakterien oder Hefen werden vergleichbare Sequenzen häufiger deletiert als amplifiziert (La Spada et al., 1997; Jin und Warren, 2000).

Die Häufigkeit und Richtung der Längenänderungen von Trinukleotid-Folgen ist abhängig von der Orientierung der Repetition in Relation zum Replikationsursprung. Daraus wurde auf Unterschiede in der Stabilität zwischen der kontinuierlichen und der diskontinuierlichen DNA-Synthese geschlossen (Shimizu et al., 1996; Hirst und White, 1998; Iyer und Wells, 1999). Die meisten repetitiven Elemente, deren Amplifikation beim Menschen zu Erbkrankheiten führt, können alternative DNA-Strukturen annehmen, wie z.B. Haarnadelstrukturen oder Tripel- und Tetra-Helices (zur Übersicht Sinden et al., 2002; Darlow und Leach, 1998; Mitas, 1997; Fry und Loeb, 1994). Häufig unterscheidet sich bei den Einzelsträngen einer repetitiven Sequenz die Tendenz, eine alternative DNA-Struktur auszubilden, was die Abhängigkeit von der Orientierung zum Replikationsursprung erklären könnte (Wells, 1996). In *in-vitro*- und in *in-vivo*-Replikations-Experimenten blockieren bzw. behindern solche Sekundärstrukturen die DNA-Polymerisation (Kang et al., 1995; Usdin und Woodford, 1995; Samadashwily et al., 1997). Ein Pausieren der DNA-Polymerase an repetitiver DNA und ein mehrfaches Wiederaufnehmen der Polymerisation könnte auch eine Vervielfachung der Anzahl der wiederholten Einheiten durch *DNA slippage* erklären, die in einem Schritt nur zu einer Verdopplung der Anzahl der repetierten Einheit führen kann (Sinden et al., 1002; McMurray, 1995; 1999).

In neueren Experimenten zur Instabilität von 5'-d(CAG)_n-3'-Wiederholungen in Säugerzellen spielte zudem der Abstand der repetitiven Sequenz vom Replikationsursprung eine Rolle (Cleary et al., 2002). Diese Tatsache stellt einen Ansatz für die Erklärung der unterschiedlichen Stabilität von repetitiven DNA-Elementen in der Keimbahn und in Körperzellen dar. Durch den Gebrauch von verschiedenen Replikationsursprüngen in Keimbahn- oder Somazellen könnte sich die Lage eines Okazaki-Fragments relativ zur repetitiven DNA verändern (Mirkin und Smirnova, 2002).

In *Saccharomyces cerevisiae* führen Mutationen in Proteinen, die bei dem Prozessieren von Okazaki-Fragmenten eine Rolle spielen, zu einer Destabilisierung von repetitiver DNA (Gordenin et al., 1997; White et al., 1999; Spiro et al., 1999). Repetitive DNA wird durch Mutationen in Enzymen für die DNA-Reparatur oder DNA-Replikation in Hefe und Bakterien

abhängig von ihrer Sequenz ebenfalls destabilisiert (zur Übersicht Jin und Warren, 2000; Rolfsmeier et al., 2001; Pelletier et al., 2003). Methylierung stabilisiert 5'-d(CGG)_n-3'-Elemente in Bakterien (Nichol und Pearson, 2002).

A.6.2 Interaktionen von Proteinen mit repetitiver DNA

Die Analyse von Proteinen, die repetitive Elemente binden, könnte zur Aufklärung der Funktion von repetitiven Elementen und ihrer (in)stabilen Vererbung beitragen. In unserem Labor wurde deshalb ein 5'-d(CGG)_n-3'-bindendes Protein (CGGBP1) gereinigt und charakterisiert (Deissler et al., 1996; 1997; Müller-Hartmann et al., 2000; Naumann et al., 2001). Proteine, die andere repetitive Elemente binden, sind ebenfalls beschrieben worden (Richards et al., 1993; Yano-Yanagisawa et al., 1995; Epplen et al., 1996; Mäueler et al., 1999). Ein im Zellkern lokalisiertes Protein bindet an 5'-d(GAA)_n-3'-Wiederholungen, die bei der Friedreichschen Ataxie amplifiziert vorliegen (Mäueler et al., 2001). Außerdem erkennt das Fehlpaarungs-Reparaturprotein MSH2 Haarnadelstrukturen, die von 5'-d(CTG)_n-3'-Elementen eingenommen werden (Pearson et al., 1997).

Für das Verständnis eines toxischen RNA-Effektes (A.4.2) liefert die Analyse von RNA-bindenden Proteinen einen Ansatz. Die identifizierten 5'-(CUG)_n-3'-bindenden Proteine 1 und 2 werden durch die amplifizierte Trinukleotid-Wiederholung bei der Myotonen Dystrophie möglicherweise weggefangen oder in ihrer subzellulären Lokalisation gestört (Bhagwati et al., 1996; Timchenko, 1999). RNA-bindende 5'-(CAG)_n-3'- oder 5'-(CGG)_n-3'-spezifische Proteine sind ebenfalls bekannt (McLaughlin et al., 1996; Rosser et al., 2002).

A.7 Das 5'-d(CGG)_n-3'-bindende Protein CGGBP1

A.7.1 Reinigung des Proteins und Klonierung des CGGBP1-Gens

In Gelretardations-Experimenten (*electrophoretic mobility shift assays*; EMSA) wurde eine 5'-d(CGG)_n-3'-bindende Aktivität in Kernextrakten von Zelllinien verschiedener Säugetiere sowie in primären menschlichen Zellen nachgewiesen (Deissler et al., 1996). Aufgrund der hohen Sequenzspezifität der Bindung wurde das Protein mittels DNA-Affinitätschromatographie gereinigt und p20 oder CGGBP1 genannt (Deissler et al., 1996). South-Western-Experimente zeigten, daß das 20 kDa große Protein allein an der Komplexbildung mit der repetitiven DNA beteiligt war (Deissler et al., 1997). Tryptische Peptide des aufgereinigten Proteins wurden durch Tandem-Massenspektroskopie sequenziert und die so bestimmte Sequenz für Datenbankrecherchen eingesetzt. Es wurde ein *expressed sequence tag* (EST) Klon (Boguski et al., 1993) identifiziert, der alle drei Peptide in einem Leseraster

enthielt. Die Expression in *E. coli* und die anschließende Analyse der Bindungseigenschaften zeigte, daß der isolierte EST-Klon 269133 (siehe I.1) tatsächlich für das 5'-d(CGG)_n-3'-bindende Protein CGGBP1 kodierte (Deissler et al., 1997).

A.7.2 Eigenschaften der Bindungssequenz von CGGBP1

CGGBP1 erkennt und bindet sequenzspezifisch die Triplet-Folge 5'-(CGG)_n-3' in doppelsträngiger DNA. Im Überschuß angebotene Trinukleotide anderer Sequenz oder andere GC-reiche Sequenzen verhindern den Komplex aus DNA und Protein nicht. Das Protein toleriert nur wenige spezielle Veränderung seiner repetitiven Zielsequenz (Tab. 1). Der Austausch jedes zweiten Desoxycytidins durch ein Desoxythymidin oder eine 5'-d(AGG)-3'-Unterbrechung nach jedem 9. Triplet führen im Gegensatz zu anderen Abweichungen weiterhin zu spezifischen Komplexen zwischen CGGBP1 und ³²P-markierten Oligodesoxyribonukleotiden (Deissler et al., 1996; 1997). Die biologische Funktion der Komplexbildung wurde überprüft, indem die Trinukleotid-Folge im genomischen Sequenzzusammenhang des FMR1-Gens in EMSA als Sonde eingesetzt wurde (Deissler et al., 1996).

Der Verlust der oft im FMR1-Gen vorhandenen 5'-d(AGG)-3'-Unterbrechungen gilt als ein Schritt bei der pathogenen Amplifikation der Trinukleotid-Wiederholung (Eichler et al., 1994; Kunst und Warren, 1994; Dombrowski et al., 2002). Bei einer Analyse des Einflusses von verschiedenen Basenfehlpaarungen auf die Bindung durch CGGBP1 stellte sich heraus, daß CGGBP1 einzig eine A/G-Basenfehlpaarung in jedem zweiten Triplet duldet (Deissler et al., 1997).

Um die Häufigkeit der Bindungssequenz im menschlichen Genom zu demonstrieren, wurden Metaphasechromosomen mit einem Fusionsprotein aus CGGBP1 und dem grün fluoreszierenden Protein (GFP) inkubiert (Müller-Hartmann et al., 2000). Eine Lokalisation des Fusionsproteins wurde durch das Bestrahlen der Metaphasechromosomen mit Licht der Wellenlänge von 475 nm visualisiert. Daraus ergab sich, daß Bindungssequenzen für CGGBP1 auf allen Chromosomen vorhanden waren. Das Fusionsprotein lagerte sich jedoch bevorzugt an die 5'-d(CGG)_n-3'-Wiederholungen der tandem-artig wiederholten 28 S rRNA-Gene auf den akrozentrischen Chromosomen 13, 14, 15, 21 und 22 an.

Bei dem Fragilen-X-Syndrom sind amplifizierte 5'-d(CGG)_n-3'-Wiederholungen in der Regel methyliert (Oberlé et al., 1991), weshalb auch die Methylierungssensitivität von CGGBP1 untersucht wurde. Es wurde keine Komplexbildung mit methylierter DNA beobachtet, wenn gereinigtes CGGBP1 verwendet wurde (Deissler et al., 1996). Zudem ließen sich spezifische

Komplexe des Proteins mit einem ^{32}P -markierten 5'-d(CGG) $_{17}$ -3'-Oligodesoxyribonukleotid als Sonde nicht durch einen Überschuß an vollständig methylierter DNA beeinflussen. Partiiell methylierte DNA wurde durch CGGBP1 jedoch gebunden.

Die nicht unterbrochenen 5'-d(CGG) $_n$ -3'-Wiederholungen	
(CGG) $_8$	5'-(CGG CGG) $_4$ -3' 3'-(GCC GCC) $_4$ -5'
(CGG) $_{12}$	5'-(CGG CGG) $_6$ -3' 3'-(GCC GCC) $_6$ -5'
(CGG) $_{17}$	5'-(CGG CGG) $_8$ -CGG-3' 3'-(GCC GCC) $_8$ -GCC-3'
Die unterbrochenen 5'-d(CGG) $_n$ -3'-Wiederholungen	
CGG8Tds	5'-(CGG TGG) $_8$ -CGG-3' 3'-(GCC ACC) $_8$ -GCC-5'
CGG10AGGds	5'-(CGG) $_3$ AGG(CGG) $_9$ AGG(CGG) $_3$ -3' 3'-(GCC) $_3$ TCC(GCC) $_9$ TCC(GCC) $_3$ -5'
Sequenz der methylierten Oligodesoxyribonukleotide	
8MCGGds	5'-(CGG m CGG) $_8$ CGG-3' 3'-(GCC G m CC) $_8$ -GCC-5'
4MCGGds	5'-(CGG) $_6$ (m CGG) $_4$ (CGG) $_7$ -3' 3'-(GCC) $_6$ (G m CC) $_4$ (GCC) $_7$ -5'
Sequenz der Oligodesoxyribonukleotide mit Basenfehlpaarung	
CGG8A/ (CGG) $_{17}$	5'-(CGG AGG) $_8$ -CGG-3' 3'-(GCC GCC) $_8$ -GCC-5'
CGG10AGG (CGG) $_{17}$	5'-(CGG) $_3$ AGG (CGG) $_9$ AGG (CGG) $_3$ -3' 3'-(GCC) $_3$ GCC (GCC) $_9$ GCC (GCC) $_3$ -5'
5'-d(CGG) $_n$ -3'-Wiederholungen im genomischen Kontext	
FMR1-(CGG) $_{16}$	5'-CGT GCG GCA GCG (CGG) $_{16}$ CTG GGC CTC GAG-3' 3'-GCA CGC CGA CGC (GCC) $_{16}$ GAC CCG GAG CTC-5'
28S rRNA-(CGG) $_7$	5'-CGT GTG GTG TGC GTC GGA GGG (CGG) $_7$ GGG TGT GGG GTC CTT CCC-3' 3'-GCA CAC CAC ACG CAG CCA CCC (GCC) $_7$ CCC ACA CCC CAG GAA GGG-5'

Tabelle 1 Spezifische Bindung von CGGBP1 an Oligodesoxyribonukleotide.

A.7.3 Eigenschaften des CGGBP1

Die Primärsequenz gab keinen Hinweis auf die mögliche Funktion(en) des CGGBP1. Bei Datenbankrecherchen wurde nur eine mögliche Kernlokalisationssequenz gefunden. Analysen mit Deletionsmutanten belegten die Funktionalität dieses Signals in Zellkultur (Müller-Hartmann et al., 2000). Deletionen sowie bestimmte Aminosäureaustausche im Bereich dieser basischen Region im Protein beeinträchtigten zudem die DNA-Bindung des Proteins (siehe E.3.1).

Bei der repetitiven Natur der Bindungssequenz und der geringen Größe des CGGBP1 war es denkbar, daß mehrere CGGBP1-Moleküle nebeneinander an DNA banden. In EMSA wurde die Größe des DNA-Protein-Komplexes bestimmt (Müller-Hartmann et al., 2000). Es wurden dabei bei zu vier CGGBP1-Molekülen pro 5'-d(CGG)₁₇-3'-Wiederholung gefunden.

A.7.4 Mögliche Funktion von CGGBP1

Die Überexpression von CGGBP1 inhibiert den FMR1-Promotor in HeLa-Zellen (Müller-Hartmann et al., 2000). Bei Cotransfektion eines CGGBP1-Expressionsplasmids mit Promotortest-Vektoren, die das Reportergen Luciferase unter der Kontrolle des FMR1-Promotors exprimierten, wurde die Luciferase-Expression reprimiert, wenn das Testkonstrukt eine 16fache 5'-d(CGG)-3'-Wiederholungen enthielt. Während das Hinzufügen einer 160-fachen 5'-d(CGG)-3'-Wiederholung die Repression verstärkte, war der Promotor nach der Deletion der endogenen Trinukleotid-Wiederholung weniger empfindlich gegenüber der Überexpression von CGGBP1 (Müller-Hartmann et al., 2000). Bisher ist nicht bekannt, wieviele andere zelluläre Gene, die mit 5'-d(CGG)_n-3'-Wiederholungen assoziiert sind, durch CGGBP1 reguliert werden.

Bei der Sequenzierung des offenen Leserasters des CGGBP1-Gens bei Trägern der Prä- oder Vollmutation des Fragilen-X-Syndroms wurden keine Mutationen festgestellt. Das CGGBP1 wurde in peripheren weißen Blutzellen von Patienten auch unverändert exprimiert. Eine auslösende Rolle des CGGBP1 bei der Pathogenese des Fragilen-X-Syndroms ist somit unwahrscheinlich.

A.8 Ziel dieser Arbeit

In einer Reihe von menschlichen Krankheiten beobachtet man die pathogene Verlängerung bestimmter repetitiver Trinukleotid-Sequenzen. Der Mechanismus dieser Verlängerung ist noch weitgehend ungeklärt. Die Untersuchung von Proteinen, die diese Sequenz spezifisch erkennen, könnte zum Verständnis der Replikation und Variabilität dieser instabilen Elemente beitragen. Deshalb wurde in unserem Labor ein CGG-bindendes Protein (CGGBP1) aus Kernextrakten gereinigt und teilweise charakterisiert, das in HeLa-Zellen bei Überexpression den FMR1-Promotor inhibiert (Deissler et al., 1996 und 1997; Müller-Hartmann et al., 2000). In der vorliegenden Studie sollte die genomische Organisation des CGGBP1-Gens bei Mensch und Maus aufgeklärt werden. Den genomischen Klon des menschlichen Gens hatte ich bereits während der Diplomarbeit isoliert (Naumann, 1998). Die Klonierung des CGGBP1-Gens der Maus war Thema dieser Arbeit. Die Expression des CGGBP1-Gens in menschlichen Zellen und Geweben und während der Embryonalentwicklung der Maus sollte mittels RNA-Transfer- und RT-PCR-Experimenten untersucht werden. Der Promotor des menschlichen Gens sollte in Transfektionsstudien identifiziert und seine Methylierungssensitivität analysiert werden. Durch die Bisulfit-Methode des genomischen Sequenzierens sollte dann der Methylierungszustand des CGGBP1-Gens in der DNA aus Leukozyten und aus HeLa-Zellen ermittelt werden. Um zukünftig die Erforschung der Funktion des *cggbp1*-Gens durch Geninaktivierungs-Experimente in der Maus zu ermöglichen, sollte eine Strategie für seine konditionale Inaktivierung entworfen und der Geninaktivierungs-Vektor konstruiert werden. Außerdem sollte geklärt werden, ob CGGBP1 an kurze und von Trinukleotiden anderer Sequenz unterbrochene 5'-d(CGG)_n-3'-Wiederholungen bindet, wie sie in den FMR1- und CGGBP1-Genen des Menschen und der Maus vorkommen.

B Material

B.1 Chemikalien

Während der hier beschriebenen Arbeit wurden einige Firmen von anderen Firmen aufgekauft und haben somit ihren Namen gewechselt. Hier werden die Firmen genannt, von denen die Chemikalien erworben wurden, auch wenn die Firmen mittlerweile anders heißen.

Acrylamid/Bisacrylamid (19:1) 40%	Roth, Karlsruhe
Agar	Fluka, Buchs, CH
Agarose	Seakem, Biozym Diagnostik, Hameln Life Technologies, Paisley, Schottland
Ammoniumacetat	Merck, Darmstadt
Ammoniumpersulfat	Sigma, Steinheim
Ampicillin	Grünenthal, Stolberg
Bacto-Tryptone	Difco, Detroit, MI, USA
Borsäure	Merck, Darmstadt
Bromphenolblau (BPB)	Merck, Darmstadt
Calciumchlorid	Baker, Deventer, Niederlande
Caesiumchlorid	KMF Laborchemie, St. Augustin
Chromatographiewasser	Merck, Darmstadt
Desoxyribonukleosid-5'-triphosphate	Sigma, Steinheim
Dextransulfat	Pharmacia Biotech, Uppsala, Schweden Biomol Feinchemikalien, Hamburg
N, N'-Dimethylformamid (DMF)	Merck, Darmstadt
Dimethylsulfoxid (DMSO)	Fluka, Buchs SG, Schweiz
D,L-Dithiothreitol (DTT)	Sigma, Steinheim
DNA aus Heringspermien	Sigma, Steinheim
Dodecylsulfat, Na-Salz (SDS, ultrarein)	Merck, Darmstadt
EDTA (TitrplexIII)	Merck, Darmstadt
Essigsäure	Merck, Darmstadt
Ethanol	Merck, Darmstadt
Ethidiumbromid	Calbiochem, La Jolla, CA, USA
Formamid (p.a.)	Merck, Darmstadt
Formaldehyd	Roth, Karlsruhe
Glukose (wasserfrei)	Merck, Darmstadt

Glycerin, 87%	Merck, Darmstadt
Größenmarker für DNA-Fragmente	Fermentas MBI, Vilnius, Litauen
Größenmarker für RNA-Moleküle	Promega, Madison, WI, USA
Harnstoff	Merck, Darmstadt
HEPES	Merck, Darmstadt
Hexa-Oligodesoxyribonukleotide	Boehringer Mannheim, Mannheim
Hydrochinon	Sigma, Steinheim
Isopropyl- β -D-thiogalactopyranosid (IPTG)	Diagnostics Chemicals, Charlottetown, Kanada
Isoamylalkohol	Merck, Darmstadt
Kaliumacetat	Roth, Karlsruhe
Kaliumchlorid	Merck, Darmstadt
Kanamycin	Sigma, Steinheim
Magnesiumchlorid	Merck, Darmstadt
Magnesiumsulfat	Merck, Darmstadt
2-Mercaptoethanol	Roth, Karlsruhe
Magermilchpulver	Nestlé, Frankfurt am Main
Mineralöl	Sigma, Steinheim
MOPS 3-[N-morpholino]-2-Hydroxypropansulfonsäure	Roth, Karlsruhe
Natriumacetat	Merck, Darmstadt
Natriumchlorid	Merck, Darmstadt
Natriumcitrat	Merck, Darmstadt
Natriumhydrogenphosphat (Na_2HPO_4)	Merck, Darmstadt
Natriumhydroxid (Plätzchen)	Roth, Karlsruhe
Natrium-N-lauroylsarcosinat	Fluka, Buchs, CH
NZ-Amine	Otto Aldag, Hamburg
Penicillin	Hoechst, Frankfurt am Main
Phenol	J.T.Baker, Deventer, Niederlande Roth; Karlsruhe
PolydAdT	Pharmacia
2-Propanol	Merck, Darmstadt
Protein Assay	Biorad, Hercules, CA, USA
Rinderserumalbumin	New England Biolabs, Beverly, MA, USA
RNA-Größenmarker	Promega, Madison, WI, USA
Saccharose (ultrarein)	ICN Biomedicals Inc., Cleveland, Ohio, USA

S-Adenosyl-Methionin (SAM)	New England Biolabs, Beverly, MA, USA
Salzsäure (rauchend)	Merck, Darmstadt
Select-Yeast Extract	Gibco BRL, Life Technologies, Paisley, Schottland
Sephadex G-50	Pharmacia Biotech, Uppsala, Schweden
Spermidin	Sigma, Steinheim
Streptomycin	Hoechst, Frankfurt am Main
TEMED (N,N,N,N'-Tetra-Methyl-ethylendiamin)	Biorad, Hercules, CA, USA
Tetracyclin	Sigma Steinheim
Trichlormethan	Riedel-de Haën, Seelze
Tris(hydroxymethyl)aminomethan	Merck, Darmstadt
Triton X-100	Sigma, Steinheim
X-gal (5-Bromo-4-chloro-3-indolyl- α -D-galactopyranosid)	Biomol Feinchemikalien, Hamburg
Xylencyanol	Fluka, Buchs SG, Basel, Schweiz

B.2 Radioisotope

Desoxyribonucleosid-5'-[α - 32 P]-Triphosphate spezifische Aktivität > 3000 Ci/mmol	DuPont NEN, Boston, MA, USA
Desoxyribonucleosid-5'-[γ - 32 P]-Triphosphate spezifische Aktivität > 3000 Ci/mmol	DuPont NEN, Boston, MA, USA

B.3 Plasmide

easyflox	war ein Geschenk von Manolis Pasparakis
EST-Klon 269133 (CGGBP1)	war ein Geschenk von Heidrun Deissler
g20-143	Naumann, 1998
g20-66	Naumann, 1998
pBSII KS (-)	war ein Geschenk von Dr. Jennifer Hertz
pEGFP-C1	Promega, Madison, WI, USA
pGEM-T Vektor	Promega, Madison, WI, USA
pGL2basic	Promega, Madison, WI, USA
pGL2control	Promega, Madison, WI, USA
pGL2-Derivate	diese Arbeit
pCMV-p20	Herbert MüllerHartmann, 2000

B.4 Lambda-DNA

g20-2	Naumann, 1998
λ -DASH II	Stratagene, La Jolla, CA, USA

B.5 Oligodesoxyribonukleotide

Alle Desoxyribonukleotide, die in dieser Arbeit verwendet wurden, wurden bei Gibco BRL, MWG Biotech und Sigma-ARK synthetisiert und in Chromatographiewasser gelöst.

B.6 Nukleinsäuren (DNA)

genomische DNA von HeLa und PWBZ behandelt mit Bisulfit	war ein Geschenk von Dr. Ralph Remus
genomische DNA von C57BL/6	war ein Geschenk von Dr. Jennifer Hertz
β -Aktin	Clontech, Palo Alto, CA, USA

B.7 Bakterienstämme

XL-1 Blue MRA (P2)	Stratagene, La Jolla, CA, USA
XL-1 Blue MRF'	Stratagene, La Jolla, CA, USA
JM110	war ein Geschenk von Prof. Müller-Hill

B.8 Enzyme

B.9 Restriktionsenzyme

Die in der vorliegenden Arbeit verwendeten Restriktionsenzyme stammen von folgenden Firmen: Boehringer Mannheim, Mannheim; New England Biolabs, Beverly, MA, USA, Roche, Mannheim, Gibco BRL, Life Technologies, Paisley, Schottland und Fermentas MBI, Vilnius, Litauen. Die Restriktionen wurden mit den von den Firmen gelieferten Puffern gemäß der angegebenen Bedingungen durchgeführt.

B.10 Sonstige Enzyme

Alkalische Phosphatase*	Roche, Mannheim
DNA-Polymerase I (Klenow-Fragment)	Boehringer Mannheim, Mannheim Fermentas MBI, Vilnius, Litauen
DNaseI*	Promega, Madison, WI, USA
Exonuclease III*	Fermentas MBI, Vilnius, Litauen
Lysozym	Calbiochem, La Jolla, CA, USA
Methylasen (HhaI-M; HpaII-M, SssI-M)*	New England Biolabs, Beverly, MA, USA

Proteinase K	Merck, Darmstadt
Pwo-DNA Polymerase*	Roche, Mannheim
Pfu-DNA Polymerase*	Stratagene, La Jolla, CA, USA
Reverse Transkriptasen*	Promega, Madison, WI, USA Qiagen, Hilden Gibco BRL, Life Technologies, Paisley, Schottland
RNase A	Sigma, Steinheim
T4 DNA-Ligase *	New England Biolabs, Beverly, MA, USA Promega, Madison, WI, USA Fermentas MBI, Vilnius, Litauen
T4-5'-Polynukleotidkinase *	New England Biolabs, Beverly, MA, USA Promega, Madison, WI, USA
T4-DNA-Polymerase*	New England Biolabs, Beverly, MA, USA
S1-Nuklease*	Fermentas MBI, Vilnius, Litauen New England Biolabs, Beverly, MA, USA
Shrimp Alkaline Phosphatase*	United States Biochemical Corporation, Cleveland, OH, USA
Trypsin 1-300 (aus Schweinepankreas)	United States Biochemical Corporation, Cleveland, OH, USA
Taq DNA Polymerase *	Promega, Madison, WI, USA Sigma, Steinheim Qiagen, Hilden

* Bei diesen Enzymen wurde der vom Hersteller mitgelieferte Puffer für die Reaktionen verwendet

B.11 Kits

ABI PRISM Big Dye Terminator Sequencing Kit	PE Applied Biosystems, Foster City, CA, USA
RT-PCR-Kits	Promega, Madison, WI, USA Qiagen, Hilden Gibco BRL, Life Technologies, Paisley, Schottland
ExoIII/S1 Deletion Kit	Fermentas MBI, Vilnius, Litauen
Gelextraktion	Macherey-Nagel, Düren Qiagen, Hilden
Luciferase Assay System	Promega, Madison, WI, USA
mRNA Isolierung	Roche Diagnostics, Mannheim Macherey-Nagel, Düren
Nucleofection Kit R	amaxa, Köln

NucleoSpin RNA L	Macherey-Nagel, Düren
pGEM-T Vektor System	Promega, Madison, WI, USA
Plasmidpräparations-Kits	Gibco BRL, Life Technologies, Paisley, Schottland
	Qiagen, Hilden

B.12 Gewebe

Organe von Mäusen (C57BL/6; 129)	wurden von Maria Palka-Santini zur Verfügung gestellt
----------------------------------	--

B.13 Zellen

3T3 (Maus)	(Todaro et al., 1963) ATCC CCL-92
ES-Zellen (C57BL/6) Bruce 4	(Köntgen et al., 1993) Geschenk von Prof. Rajewsky, ehemals Institut für Genetik, Köln
CHO 8/24 720 LIF D(.1)	Genetics Institute Cambridge, Mas. USA Geschenk von Prof. Rajewsky, ehemals Institut für Genetik, Köln
primäre embryonale Fibroblasten (EF)	Geschenk von Prof. Rajewsky, ehemals Institut für Genetik, Köln
EF-neo Mausstamm	(Müller et al., 1991)
HEK293	menschliche Zellen in Kultur mit einem Fragment von Adenovirus 5 transformiert (Graham et al., 1977) ATCC CRL-1573
HeLa	Zellen eines menschlichen Gebärmutterkarzinoms; American Type Culture Collection (ATCC)Rockville, USA Nummer ATCC CCL2
KB	Zellen eines menschlichen Mundschleimhautkarzinoms; American Type Culture Collection (ATCC) Rockville, USA Nummer ATCC CCL17

B.14 Lösungen und Medien für die Zellkultur

B.14.1 Allgemeine Zellkultur

Dulbecco modifiziertes Eagle Medium (DMEM)	Bablanian et al., 1965 Lagerung bei 4°C
--	--

Einfriermedium für Zelllinien (Stammlösung)	Kulturmedium mit 20% FCS 10% Glycerin Lagerung bei -20°C
Fötales Kälberserum (FCS)	Cytogen, Berlin
Kulturmedium (Zelllinien)	Dulbecco oder Eagles mit 10% FCS
PBS (Phosphate Buffered Saline)	140 mM NaCl 3 mM KCl 8 mM Na ₂ HPO ₄ x 2 H ₂ O 1 mM KH ₂ PO ₄ 1 mM MgCl ₂ x 6 H ₂ O 1 mM CaCl ₂ x 2 H ₂ O pH auf 7,4 eingestellt
PBS-d	PBS ohne CaCl ₂ und MgCl ₂
Tris-Saline	14 mM NaCl 5 mM KCl 0,3 mM Na ₂ HPO ₄ 25 mM Tris-HCl pH 7,4 0,1% Glukose (w/v) 0,03% Penicillin (w/v) 0,02% Streptomycin (w/v)
Trypsin Lösung	0,25% Trypsin in PBS-d 1,35 mM EDTA 0,03% Penicillin (w/v) 0,02% Streptomycin (w/v)

B.14.2 ES-Zellkultur

Die Medien und die Zusätze für die ES-Zellkultur wurden, wenn nicht anders aufgeführt, bei Gibco BRL, Life Technologies, Paisley, Schottland bezogen.

DMEM für ES-Zellen	500 ml DMEM (4500 mg Glukose, ohne Natriumpyruvat) dem folgende Mengen weiterer Produkte von Gibco zugesetzt wurde 75-90 ml FCS (vorher getestet, hitzeinaktiviert) 6 ml 100 mM Natriumpyruvat 6 ml 200 mM Glutamin 6 ml nicht essentielle Aminosäuren (100x) 1,2 ml 50 mM β-Mercaptoethanol 6 ml Penicillin/Streptomycin 1,5 - 3 ml "LIF"
--------------------	---

DMEM für EF-Zellen	500 ml DMEM (Glutamax-1) 4500 mg Glukose, ohne Natriumpyruvat 60 ml FCS (hitzeinaktiviert) 6 ml 100 mM Natriumpyruvat 6 ml Penicillin/Streptomycin
Dimethylsulfoxid (DMSO)	Sigma, Steinheim
Einfriermedium für ES/EF-Zellen (1x)	Kulturmedium mit 10% DMSO
Fötale Kälberserum (FCS)	PAN, Südafrika
Fötale Kälberserum, dialysiert	Gibco BRL, Life Technologies, Paisley, Schottland
Hühnerserum	Life Technologies, Paisley, Schottland
Gancyclovir	Cymeven [®] , Syntex
Geneticin (G418)	Life Technologies, Paisley, Schottland
Gelatine	Sigma, Steinheim
L-Glutamin	Life Technologies, Paisley, Schottland
LIF-Medium für CHO 8/24 720 LIF D(.1)	500 ml α -MEM ohne Nukleoside / Desoxynucleoside 50 ml FCS, dialysiert, hitzeinaktiviert 5 ml Penicillin/Streptomycin (100x) 5 ml 100 mM Natriumpyruvat
"LIF" (Leukämie inhibierender Faktor)	Überstand der LIF-produzierenden Zelllinie CHO 8/24 720 LIF D(.1)
2-Mercaptoethanol	Gibco BRL, Life Technologies, Paisley, Schottland
Methotrexat	Sigma, Steinheim
Mineralöl (auf Embryos getestet)	Sigma, Steinheim
Mitomycin C	Sigma, Steinheim
Mitomycin C-Stammlösung	1 mg/ml in PBS mit 5% DMSO für die Zellkultur Lagerung bei -20°C, lichtempfindlich
Natriumpyruvat (100 mM)	Gibco Life Technologies, Paisley, Schottland
Penicillin/Streptomycin 100x	Gibco Life Technologies, Paisley, Schottland
PBS ohne Calcium und Magnesium	PAA, Linz, Österreich
Trypsin-EDTA (10x)	Gibco, Life Technologies, Paisley, Schottland 10x mit PBS auf 1x verdünnen und für den Einsatz bei ES-Zellen mit 2% Hühnerserum versetzen. Lagerung bei 4°C

B.15 Lösungen und Puffer

CsCl-Lösung (λ -DNA-Präparation)	3 M, 4 M, 5 M und 7,2 M CsCl in 10 mM MgSO ₄ 10 mM Tris-HCl pH 8 0,1 mM EDTA
Chloroform zur DNA-Extraktion	Chloroform und Isoamylalkohol im Verhältnis 24:1
DNA-Ladepuffer	0,25% Bromphenolblau 0,25% Xylencyanol 30% Glycerin in H ₂ O
Elutionspuffer (PAGE)	0,3 M NaAc, pH 7,5 –7,8 2 mM EDTA
ES-Zell-Lysispuffer	10 mM Tris/HCl pH 7,5 10 mM NaCl 10 mM EDTA 0,5% Sarcosyl 0,4 – 1 mg Proteinase K (frisch zusetzen)
ES-Restriktionspuffer	1x Restriktionspuffer 100 μ g/ml Rinderserumalbumin 1 mM DTT 50 μ g/ml RNase A 1 mM Spermidin 50 U Enzym pro Ansatz
Ethidiumbromidlösung	2 μ g/ml in H ₂ O
Glukoselösung (Elektroporation)	36,04 g Glukose in 100 ml H ₂ O sterilfiltrieren, Lagerung bei 4°C
10x Hexanukleotidpuffer	62,5 A ₂₆₀ U/ml Hexanukleotide 2 mg/ml BSA 0,5 M Tris-HCl pH 7,5 0,1 M MgCl ₂ 1 mM DTT Lagerung bei -20°C
HNDG (Deissler et al., 1996)	20mM HEPES pH 7,9 50 mM NaCl 0,5 mM DTT 10% Glyzerin
Hybridisierungslösung (Southern)	2x SSC 1% SDS 0,5% Milchpulver (fettarm) 0,5 mg/ml Heringsspermien-DNA 10% Dextransulfat Lagerung bei 4°C

Isopropylthio- β -D-galactosid (IPTG)	100 mM in H ₂ O (sterilfiltriert, Lagerung in Aliquots bei -20°C)
Ladepuffer für Retardierungsgele 5x	0,5x TE 50% Glycerin 1% (w/v) Bromphenolblau
Lambda-Verdünnungspuffer	10 mM MgSO ₄ 10 mM Tris-HCl pH 7,5
LB Medium (pro Liter)	10 g Bacto-Tryptone 5 g Select-Yeast Extract 10 g NaCl
LB Agar (pro Liter)	LB Medium mit 20 g Bacto-Agar
Luciferase-Puffer	Promega, Madison, WI, USA
Lysis Reagenz 5x (Zellkultur)	125 mM Tris pH 7.8 mit H ₃ PO ₄ ; 10 mM CDTA, 10 mM DTT, 50% Glycerin, 5% Triton X-100 (Promega, Madison, WI, USA)
Lysispuffer (Alkalische Lyse)	0,2 N NaOH, 1% SDS; frisch ansetzen wobei Ausgangslösungen auch nicht älter als 1-2 Monate sein dürfen
Lysispuffer (Zahnstocher Methode)	50 mM NaOH, 5 mM EDTA, 0,5% SDS, loading buffer (final 2x). Lagerung bei -20°C.
Lysozymlösung	10 mg/ml in STET-Puffer (frisch angesetzt)
2M Mg ²⁺ Lösung (Elektroporation)	20,33 g MgCl ₂ x 6 H ₂ O 24,65 g MgSO ₄ x 7 H ₂ O mit H ₂ O auf 100 ml auffüllen, autoklavieren
MOPS-Puffer 5x	0,2 M MOPS pH 7 0,05 M Natriumacetat 0,005 M EDTA
NZY Medium (pro Liter)	5 g NaCl 2 g MgSO ₄ x 7 H ₂ O 5 g Select-Yeast Extract 10 g NZ-Amine
NZY-Agar (pro Liter)	NZY Medium mit 20 g Bacto-Agar
NZY-Top Agarose	NZY Medium mit 0,7% Agarose
Phenol	0,1% 8-Hydroxyquinolin gesättigt mit 0,1 M Tris-HCl pH 7,5
Prähybridisierungslösung (Southern)	2x SSC 1% SDS 0,5% fettarmes Milchpulver 0,5 mg/ml Heringsspermien DNA Lagerung bei 4°C

Proteinase K Stammlösung	10 mg/ml in 10 mM Tris-HCl pH 8,0 Lagerung bei -20°C
Proteinase K Puffer	100 mM Tris-HCl pH 8,5 5 mM EDTA 0,2% SDS 200 mM NaCl
RNase A (Stammlösung)	10 mg/ml in H ₂ O nach Lösung 15 min kochen, bei Raumtemperatur abkühlen lassen, Lagerung bei -20°C
RNA-Ladepuffer	50% Glycerin 1 mM EDTA 0,4% Bromphenolblau
RNA-Probenpuffer	10 ml deionisiertes Formamid 3,5 ml 37% Formaldehyd 2 ml MOPS 5x
Saccharose in TE	10% (10 g in 100 ml) 40% (40 g in 100 ml)
SM Puffer (pro Liter)	5,8 g NaCl 2 g MgSO ₄ x 7 H ₂ O 50 mM Tris-HCl pH 7,5 5 ml 2% Gelatine
SOB Medium (pro Liter)	20 g Bacto-Tryptone 5 g Select-Yeast Extract 0,5 g NaCl
SOC Medium	SOB mit 20 mM Glukose und 20 mM Mg ²⁺ Lösung vor Gebrauch frisch ansetzen
1x SSC (Standard Saline Citrate)	150 mM NaCl 15 mM Na-Citrat mit 1 M Zitronensäure auf pH 7,0 einstellen
STET-Puffer	50 mM Tris-HCl, pH 8,0 50 mM EDTA 8% Saccharose 0,5% Triton X-100 Lagerung bei 4°C
TE	10 mM Tris-HCl 1 mM EDTA pH 8.0
Waschpuffer I (Southern)	2x SSC 0,1% SDS
Waschpuffer II (Southern)	0,1x SSC 0,1% SDS
WB (Elektroporation)	10% Glycerin (v/v) in ddH ₂ O

X-Gal (5-Bromo-4-Chloro-3-Indolyl- β -D-Galaktosid)	20 mg/ml in Dimethylformamid (Lagerung bei -20°C)
2x YT-Agar (pro Liter)	2 xYT-Medium 18 g Agar
2x YT-Agar (pro Liter)	5 g NaCl 10 g Hefe-Extrakt 16 g Bacto-Trypton

B.16 Datenverarbeitung

Folgende Programme wurden für Sequenzanalysen verwendet. *Wisconsin Sequence Analysis Package* der *Genetic Computer Group* (GCG) Version 10.x, Genedoc (Nicholas K.B., Nicholas H. B. Jr., 1997) Clonemanager. Für die Literaturrecherche wurde vor allem die PubMed-Datenbank am NCBI genutzt.

B.17 Sonstiges Labormaterial

Bakterieninkubator	Memmert, Schwabach
Blaukappröhrchen, 50 ml	Becton Dickinson Labware, Franklin Lakes, NJ, USA
Blaukappröhrchen, 15 ml	Becton Dickinson Labware, Franklin Lakes, NJ, USA
Cryoröhrchen (1,8 ml)	Greiner, Frickenhausen
Dialyseschläuche	Medicell, London, Großbritannien
Einweg-Halbmikroküvetten Nr. 1201	Ratiolab, Dreieich,
Einwegspritzen	Braun, Melsungen
Einwegkanülen	Terumo, Leuven, Belgien
Elektroporationsküvetten	Invitrogen
Entwicklermaschine Curix 60	Agfa-Gavaert, Mortsel, Belgien
Fotozubehör	Hama, Monheim
Frischhaltefolie	Melitta, München
Gel Blotting Papier GB003	Schleicher und Schuell, Dassel
Geneamp PCR-System 9600	Perkin Elmer, Norwalk, CT, USA
Gene pulser	Bio-Rad Laboratories, Hercules, CA, USA
Gewebekulturflaschen	Becton Dickinson (Falcon), Oxnard, CA, USA
Glaskapillaren	Brand, Wertheim
Hybridisierungsröhrchen	Bachofer, Reutlingen
Inkubator Cytoperm Typ 8080	Heraeus, Düsseldorf

Kieselgel mit Fluoreszenz-Indikator UV ₂₅₄	Macherey und Nagel, Düren
Luminometer Lumat LB9501	Berthold, Bad Wildbad
Nylonmembran (positiv geladen und amyphoter)	Roche, Mannheim PALL
Oak Ridge-Röhrchen, 35 ml	Nalgene, Rochester, NY, USA
Packaging extract Gigapack II	Stratagene, La Jolla, CA, USA
PCR Reaktionsgefäße <i>Microamp</i> (200 µl, 8er Streifen)	PE Applied Biosystems, Foster City, CA, USA
Petrischalen	Sarstedt, Nümbrecht
pH-Indikatorstäbchen, nicht blutend	Merck, Darmstadt
Photometer DU640	Beckman
Pipetten	
0-10 µl, 10-100 µl	Eppendorf, Hamburg
1-20µl, 10-200µl	Gilson, medical electronics, Villiers-le-Bel, Frankreich
50-250, 250-1000	Brand, Wertheim
Pipettierhilfe pipump	Glasfirm, Gießen
Pipettierpumpe Fortuna Labopet 240	Greiner, Frickenhausen
Pipettenspitzen	Eppendorf, Hamburg
Pipettenspitzen mit Filter	SLG, Gauting
Pulse controller	Bio-Rad Laboratories, Hercules, CA, USA
Reaktionsgefäße (1,5 ml und 2 ml)	Brand, Wertheim Eppendorf, Hamburg
Röntgenfilme	Eastman Kodak Company, Rochester, NY, USA X-OMAT AR5
Röntgenfilmkassetten (Suprema) zur Autoradiographie mit zwei Verstärkerfolien	Dr. Goos
Sequenziergerät (ABI Prism® 377)	PE Applied Biosystems, Foster City, CA, USA
Spektrophotometer Ultrospec 2000	Pharmacia Biotech, Uppsala, Schweden
Sterilbank	Bio Gard ;Hood Baker Company, Sanford, CT, USA Antair BSK, Anthos Mikrosysteme, Köln
Sterilfilter-Einmalfilterhalter	Schleicher & Schuell, Dassel (0,22 µm und 0,45 µm Porengröße)
Szintillationszähler LS 1801	Beckman, Palo Alto, CA, USA
Whatman-Blotting Papier	Schleicher&Schüll
Vortex Genie 2	Bender&Hobein, Zürich, Schweiz

Zählkammer (Neubauer)	Brand, Wertheim
Zellkulturinkubatoren	Steri-Cult 200; Forma Scientific, Marietta, OH, USA Cytoperm; Heraeus Instruments, Düsseldorf
Zentrifugen	
Laborzentrifugen	Cryofuge 5000; Heraeus, Düsseldorf Doctor Wifug, Stockholm, Schweden Labofuge GL; Heraeus Sepatech, Hanau Medifuge; Heraeus Sepatech, Hanau MC 12 V; RC5B; TC6; Sorvall Instruments, Du Pont, Bad Homburg
Ultrazentrifugen	L7 55; L8 55 M; L8 70; Beckman Instruments Inc., Palo Alto, CA, US
Vakuumzentrifuge	Speed Vac Concentrator; Eppendorf, Hamburg
Zentrifugenröhrchen, Ultra-Clear (SW 60)	Beckman Instruments Inc., Palo Alto, CA, USA

C Methoden

Einige in diesem Kapitel aufgeführte Techniken sind aus meiner Diplomarbeit aus dem Jahr 1998 entnommen und wurden für diese Arbeit z. T. modifiziert.

C.1 Zellkultur

Es wurde immer unter einer Sterilbank (Antair BSK) gearbeitet. Das bei 4°C gelagerte Medium und der Waschpuffer wurden auf 37°C, das Serum und die Trypsinlösung auf Raumtemperatur (RT) vorgewärmt.

C.1.1 Allgemeine Techniken

C.1.1.1 Zählen von Zellen

Die Anzahl von Zellen wurde mittels einer Zählkammer nach Neubauer (Freshney, 1987) bestimmt. Vor der Verwendung der Zählkammer wurde darauf geachtet, daß an den Kontaktstellen zwischen dem Spezialdeckglas und der Neubauer-Zählkammer Newtonsche Ringe zu erkennen waren. Die Zellsuspension (trypsinisierte Zellen in Kulturmedium) wurde gut durchmischt, bevor mit einer 200- μ l-Pipette soviel Zellsuspension entnommen wurde, um einen Tropfen am Rand der Zählkammer zu bilden. Der Tropfen wurde durch Kapillarkraft in die Zählkammer gesogen. Es wurden 4 große Quadrate ausgezählt, der Mittelwert bestimmt und schließlich mit 10^4 multipliziert, um die Zellzahl pro ml zu erhalten.

C.1.1.2 Einfrieren von Zellen

Zum Einfrieren wurden noch nicht konfluent gewachsene Zellen verwendet. Trypsinisierte Zellen wurden in Wachstumsmedium bei 1.000 U.p.m. (Labofuge GL) bei 4°C 5 Minuten zentrifugiert und der Überstand anschließend entfernt. Das Zellsediment wurde in frischem, kaltem Kulturmedium mit 10% Glycerin (HeLa) bzw. 10% DMSO (ES- und EF-Zellen) resuspendiert (Endvolumen 1,8 ml). Nach vorsichtigem Homogenisieren wurde die Zellsuspension in ein Kryoröhrchen überführt, das in einem Styroporbehälter bei -80°C eingefroren wurde, bevor es in flüssigem Stickstoff gelagert wurde.

C.1.1.3 Auftauen von Zellen

Die in flüssigem Stickstoff gelagerten Cryoröhrchen wurden direkt in ein mit 37°C warmem Wasser gefülltes Becherglas überführt. Die Zellen wurden nach dem Auftauen in 10 ml auf 37°C vorgewärmtes Kulturmedium mit 10%–15% FCS aufgenommen und durch mehrfaches Invertieren des Gefäßes gut durchmischt. Die Lösung wurde 5 Minuten bei 1.000 U.p.m. (Labofuge GL) zentrifugiert, bevor das Zellsediment in frischem, 10%–15% FCS enthalten-

dem Medium resuspendiert und in einer Gewebekulturflasche ausgesät wurde. Das Medium wurde nach dem Anwachsen der Zellen, spätestens am nächsten Tag, erneuert.

C.1.2 Kultur transformierter Zelllinien „Monolayer-Kulturen“

Die Zellen wurden in Kunststoff-Gewebekulturflaschen (25 cm², 75 cm², 175 cm²) als Einzelzellschicht-Kultur gehalten. Die 25-cm²-Flaschen wurden für frisch aufgetaute Zellen verwendet, die 75-cm²- und 175-cm²-Flaschen dienten dem Heranziehen der Zellen für Präparationen. Dem Kulturmedium (10, 35 bzw. 70 ml für die verschiedenen Flaschengrößen) wurde 10% fötales Rinderserum (FCS) zugesetzt.

C.1.2.1 Passage und Ernte

Zur Ernte bzw. Passage von konfluent gewachsenen Zellen (nach 3 bis 4 Tagen) wurde zunächst das verbrauchte Medium abgegossen und die Zellen zweimal mit einigen ml PBS-d gewaschen. Es wurde so lange mit 2–4 ml Trypsinlösung inkubiert, bis man unter dem Mikroskop das Ablösen der Zellen beobachten konnte. Dann wurde das 2,5-fache an Wachstumsmedium mit 10% FCS zugegeben, um die Reaktion von Trypsin zu stoppen. Die Zellen wurden durch vorsichtiges Aspirieren vereinzelt und im Falle einer Passage wurde die gewünschte Menge in frische Gewebekulturflaschen mit vorgelegtem Medium überführt. Die Zellen wurden im Inkubator in 5%iger CO₂-Atmosphäre bei 37°C und 80% Luftfeuchtigkeit gehalten.

C.1.2.2 Transfektion transformierter Zelllinien – Nukleofektion™

Bei der Transfektion von Zellen nach dem von der Firma amaxa entwickelten Verfahren der Nukleofektion werden Zellen in einem speziellen Puffer einem kurzen starken elektrostatischen Feld ausgesetzt. Weder die Zusammensetzung der Puffer noch die elektrischen Parameter wurden bislang veröffentlicht. Weitere Angaben zur Nukleofektion von Zellen stehen auch unter der Internetadresse <http://www.amaxa.com/products/protocols> zur Verfügung.

Pro Transfektion wurden 1x10⁶ Zellen aus einer etwa 70% konfluenten Gewebekulturschale verwendet. Die Zellkulturen wurden am Tag vor der Transfektion expandiert. Nach der Ernte wurde die Zellkonzentration nach der Trypsinisierung in einer Neubauer-Zählkammer bestimmt. Die benötigte Zellmenge wurde dann bei 150 g 7 bis 10 min bei RT zentrifugiert und der Überstand danach vollständig entfernt. Die sedimentierten Zellen wurden durch vorsichtiges Schnippen am Röhrchens im nicht entfernbaren minimalen Restvolumen resuspendiert. Anschließend wurden 100 µl der auf RT erwärmten Nukleofector Solution R pro Transfektionsansatz gemäß den Herstellerangaben zugegeben. Da eine längere Inkubation (mehr als 20 min) der Zellen in der Nucleofector Solution R zu geringerer Transfektionseffizienz und stärkerem Zellsterben geführt hätte, waren vor der eigentlichen Transfektion

einige Vorbereitungen nötig. Pro Transfektion wurden jeweils 0,5 ml Medium mit 10% FCS in einem 1,5 ml Zentrifugenröhrchen auf 37°C angewärmt. Außerdem wurde pro Transfektion je eine Vertiefung einer 6-Lochplatte mit 1 ml Medium mit 10% FCS gefüllt und die Platten im CO₂-Inkubator aufbewahrt. Auch die zu transfizierende DNA-Lösung (1–20 µg DNA) war in 5 µl TE in 1,5 ml Zentrifugenröhrchen vorgelegt. Schließlich wurden 100 µl der gut durchmischten Zellsuspension mit der DNA-Lösung durch einmaliges Auf- und Abpipettieren vermischt, in die Elektroporationsküvette überführt und das Programm A28 gestartet. Mit den vorbereiteten 0,5 ml Medium wurden die Zellsuspension in den Küvetten sofort nach der Transfektion verdünnt und die verdünnte Lösung mittels einer Plastikpipette aus den Küvetten entfernt. Eine rasche Entfernung der Zellsuspension aus den Küvetten war maßgeblich für eine erfolgreiche Transfektion. Bis zum Ausplattieren in jeweils einer Vertiefung einer 6-Lochplatte wurde die Zellsuspension in den 1,5 ml Zentrifugenröhrchen im 37°C Heizblock aufbewahrt. Die Zellen wurden dann in 5%iger CO₂-Atmosphäre mit 80% Luftfeuchtigkeit bis zur Ernte im Inkubator belassen.

C.1.3 Embryonale Stammzellkultur

Embryonale Stamm- (ES-) Zellen sind pluripotente Zellen, die aus Blastocysten gewonnen und in Zellkultur überführt worden sind (Evans und Kaufmann, 1981). In Kultur können sie manipuliert und selektioniert werden (Gossler et al., 1986; Robertson et al., 1986) und anschließend in eine Blastocyste injiziert oder mit einer Morula aggregiert werden (D.5.1). Bei der Zellkultur müssen Bedingungen eingehalten werden, die eine Differenzierung der ES-Zellen verhindern und im folgenden genauer beschrieben werden. Hier wurden die Kolonien der ES-Zellen auf einer Einzellschicht teilungsunfähiger embryonaler Fibroblasten gehalten und täglich mit einem Phasenkontrastmikroskop betrachtet. Dem Wachstumsmedium, das täglich gewechselt wurde, wurde das Cytokin LIF (Leukämie inhibierender Faktor) (Smith et al., 1992) beigelegt.

Die hier aufgeführten Techniken wurden dem Buch Laboratory Protocols for Conditional Gene Targeting (Torres und Kühn, 1998) entnommen und für diese Arbeit leicht verändert (Angela Egert und Manolis Pasparakis, ehemalige Mitarbeiter der Abteilung von Prof. Rajewsky, persönliche Mitteilung). Für die ES-Zellkultur wurden ausschließlich Chemikalien und Lösungen verwendet, die in der Arbeitsgruppe Rajewsky optimiert worden waren. Sie sind im Kapitel Material gesondert aufgeführt. Alle im folgenden Abschnitt genannten Zelltypen wurden in einem Inkubator mit 9%iger CO₂-Atmosphäre bei 37°C und 95% Luftfeuchtigkeit kultiviert.

C.1.3.1 Gewinnung des LIF-Proteins

Das Cytokin LIF inhibiert die Differenzierung embryonaler Stammzellen während ihrer Kultivierung *in vitro*. Das Wachstumsmedium einer LIF-produzierenden CHO-Zelllinie (CHO

8/24 720 LIF D(1)) wurde hier anstelle des gereinigten Proteins verwendet. In dieser transgenen Zelllinie wird das LIF Protein exprimiert und in das Wachstumsmedium sezerniert. Zellen, die das LIF-Protein in besonders hohem Maße produzieren, werden durch Methotrexat selektioniert. Die CHO-Zellen wurden wie in C.1.1.2 beschrieben aufgetaut und bis zur Konfluenz kultiviert. Für die Selektion mit Methotrexat wurde es dem Medium in einer Endkonzentration von 1 μM beigelegt. Nach Erreichen der Konfluenz (2–3 Tage) wurden die Zellen zweimal mit PBS gewaschen, um das Methotrexat zu entfernen. Nach einer zweistündigen Inkubation in frischem Medium ohne Methotrexat wurden die Zellen fünfmal mit PBS gewaschen, trypsinisiert und auf einer doppelt so großen Oberfläche ausgesät. Der Überstand, der dann das LIF-Protein enthielt, wurde nach 48 bis 66 Stunden abgenommen. Zelltrümmer wurden vor dem Sterilfiltrieren aus diesem Wachstumsmedium durch Zentrifugation abgetrennt. Vor dem Einsatz in der ES-Zellkultur wurde der Überstand auf ES-Zellen getestet, um die benötigte Menge zu bestimmen, die eine Differenzierung der ES-Zellen verhindert. Dieser Titer der so hergestellten Lösung wurde freundlicherweise von Angela Egert aus der Abteilung Prof. Rajewsky, ehemals Universität zu Köln, bestimmt.

C.1.3.2 Kultur von embryonalen Fibroblasten (EF)

C.1.3.2.1 Gewinnung und Expansion

ES-Zellen, v.a. die des Mausstammes C57BL/6, neigen bei Kontakt mit Gewebekulturschalen zur Differenzierung. Deshalb wurden embryonale Stammzellen auf einer Einzellschicht von teilungsunfähigen embryonalen Fibroblasten (EF) kultiviert. Die EF-Zellen waren resistent gegen das bei Geninaktivierungs-Experimenten eingesetzte Geneticin. Primäre embryonale Fibroblasten wurden freundlicherweise von Angela Egert aus transgenen Mäuseembryonen gemäß Torres und Kühn (1997) präpariert und zur Verfügung gestellt. Sie wurden mit 1×10^6 Zellen pro 10 cm Schale ausgesät und bei 37°C im CO₂-Inkubator bis zum Erreichen der Konfluenz inkubiert. Ab dem zweiten Tag nach dem Auftauen wurde das Medium täglich gewechselt. Vor der Behandlung der Zellen mit 4 ml Trypsin pro 10-cm-Schale wurden die Zellen mit etwa 5 ml PBS gewaschen. Die Zellen wurden mit Trypsin etwa 2–3 min bei 37°C inkubiert und die Reaktion anschließend mit 7 ml EF-Medium gestoppt. Die Zellen wurden durch Auf- und Abpipettieren vereinzelt. Die Zellen wurden gezählt und anschließend 5 min bei 150 g zentrifugiert. Anschließend wurden die Zellen entweder in ihrem Kulturmedium mit 10% DMSO eingefroren oder passagiert.

C.1.3.2.2 Mitomycin C Behandlung

Bevor ES-Zellen auf der EF-Zellschicht ausgesät werden können, müssen die EF-Zellen mitotisch inaktiv sein. Dafür behandelt man die Zellen mit dem Zytostatikum Mitomycin C. Eine konfluente Schicht von EF-Zellen wurde mit 4 ml Medium überschichtet, das 10 μg Mitomycin C/ml enthielt und 2–6 Stunden im Inkubator belassen. Anschließend wurde das

Mitomycin C gründlich entfernt. Die Zellen wurden vor der Behandlung mit 4 ml Trypsin mindestens 3 mal mit 5–10 ml PBS gewaschen. Trypsin wurde durch Zugabe von Medium gestoppt, die Zellzahl der Suspension bestimmt und die Zellen zu $5,2\text{--}5,7 \times 10^4$ Zellen/cm² in gelatinisierten Gewebekulturschalen ausplattiert. Zur Gelatinisierung wurde eine sterile 10%ige Gelatinelösung auf den Gefäßboden der Gewebekulturschalen pipettiert, die dann mehrere Stunden im Inkubator belassen wurde. Die wäßrige Lösung wurde vor dem Ausplattieren der Zellen entfernt. Nach einer Inkubation üN in frischem EF-Medium wurden diese EF-Zellen dann für die ES-Zellkultur verwendet.

C.1.3.3 Arbeiten mit embryonalen Stammzellen

ES-Zellen wurden täglich im inversen Phasenkontrastmikroskop betrachtet und bei zu hoher Zellkoloniedichte passagiert. Undifferenzierte ES-Zellen wachsen als ovale Kolonie mit scharfen, im Phasenkontrast hell erscheinenden Rändern. Bei Kontakt mit dem Gefäßboden von Gewebekulturschalen neigen embryonale Stammzellen zur Differenzierung. Deshalb wurden embryonale Stammzellen auf embryonalen Fibroblasten in gelatinisierten runden Gewebekulturschalen mit einem Durchmesser von 10 oder 15 cm oder gelatinisierten 96-Loch-Platten gehalten. Da ES-Zellen gegenüber Trypsin sehr empfindlich sind, wurde der Trypsinlösung vor ihrem Einsatz in der ES-Zellkultur 2% Hühnerserum zugesetzt. Der Proteingehalt des Hühnerserums hemmt kompetitiv das Trypsin, enthält aber im Gegensatz zu Serum von Säugetieren keinen Trypsininhibitor. Vor einer Trypsinisierung wurde 2–4 Stunden vorher das Wachstumsmedium gewechselt. ES-Zellen wurden immer mit einer minimalen Dichte von 1×10^6 Zellen pro Gewebekulturschale von 10 cm ausgesät. Bevor die Kolonien sich berühren konnten, wurden die Zellen umgesetzt oder eingefroren. ES-Zellen wurden immer bei RT und 5 bis 7 min lang bei maximal 150 g zentrifugiert.

C.1.3.3.1 Einfrieren und Auftauen ES-Zellen

C.1.3.3.1.1 Gewebekulturschalen mit einem Durchmesser von 6 cm

Das Einfrieren und Auftauen von ES-Zellen entspricht bis auf die verwendeten Medien und Gewebekulturschalen den oben aufgeführten Protokollen (C.1.1.2;C.1.1.2). Trypsinisierte ES-Zellen wurden nach der Zentrifugation zu je 3×10^6 Zellen direkt in auf 4°C gekühltes ES-Einfriermedium mit 10% DMSO aufgenommen. Ein Ansatz wurde nach dem Auftauen in einer Gewebekulturschale von 10 cm Durchmesser ausgesät.

C.1.3.3.1.2 Einfrieren in 96-Loch Platten

Zum Einfrieren wurden die auf EF-Zellen gewachsenen ES-Zellen zunächst trypsinisiert. Dafür wurden das Medium entfernt und die Zellen zweimal mit 100 µl PBS gewaschen. Pro Vertiefung wurde 50 µl Trypsin-Lösung zugegeben und 3–5 min bei 37°C inkubiert. Danach wurden 50 µl 2x Einfriermedium hinzugefügt und die Zellen durch Auf- und Abpipettieren

vereinzelt. Die Suspension wurde mit 100 µl sterilem Mineralöl überschichtet, die Platte mit Parafilm abgedichtet, in eine Styroporbox gegeben und bei -80°C eingefroren.

Zum Auftauen wurden die Platten in 38–40°C warmes, steriles Wasser getaucht, wobei das Wasser nicht zu den Zellen in den Vertiefungen gelangen durfte. Die aufgetauten Klone wurden dann vollständig in 2 ml Medium überführt und 5 min (150 g) zentrifugiert. Der Überstand wurde verworfen und die Zellen wurden in 0,5–1 ml ES-Medium in vorbereitete Vertiefungen einer 24- oder 48-Loch-Platte überführt.

C.1.3.3.2 Expansion von ES-Zellen

Vor der Expansion wurden die ES-Zellen zweimal mit 5 ml PBS gewaschen und anschließend 3–5 min bei 37°C mit 3 ml 1x Trypsinlösung mit 2% Hühnerserum inkubiert. Die Enzymreaktion wurde durch die Zugabe des zweifachen Volumens an ES-Medium gestoppt und die Zellen durch Auf- und Abpipettieren in einer 5-ml-Pipette verein und anschließend gezählt (C.1.1.1). Nach einer Zentrifugation wurden die Zellen in frischem ES-Medium aufgenommen und mit $0,8\text{--}1 \times 10^6$ Zellen pro 10 cm Platte auf EF-Zellen ausgesät. Nach 2–3 Tagen bei täglichem Mediumwechsel erreichten diese dann die Dichte für eine weitere Passage. Sollte im folgenden aus den ES-Zellen DNA präpariert werden, wurden die Zellen auf gelatinisierten Platten (C.1.3.2.2) ausgesät, um eine Verunreinigung der ES-DNA durch EF-DNA zu vermeiden.

C.1.3.3.3 Transfektion embryonaler Stammzellen mit einem Geninaktivierungs-Vektor

C.1.3.3.3.1 Vorbereitung des Geninaktivierungs-Vektors

Für die Transfektion von ES-Zellen mit dem Geninaktivierungs-Vektor wurden 25–30 µg Vektor mit dem Restriktionsenzym SacII linearisiert, anschließend mit Phenol/Chloroform extrahiert, mit Ethanol präzipitiert und in 70%igem Ethanol bei -20°C gelagert. Am Tag der Transfektion wurde das Präzipitat unter sterilen Bedingungen an der Luft getrocknet und in PBS in einer Konzentration von 0,5 µg/µl gelöst.

C.1.3.3.3.2 Elektroporation von embryonalen Stammzellen

ES-Zellen wurden am Tag vor der Transfektion passagiert und mit einer höheren Dichte ($1,5 \times 10^6$ Zellen/10 cm Gewebekulturschale) ausplattiert. Drei bis vier Stunden vor der Elektroporation wurde das Medium der subkonfluenten ES-Zellen erneuert. Zu Beginn der Transfektion wurden die Platten zweimal mit 10 ml PBS gewaschen und anschließend mit 4 ml 1x Trypsin 3–5 min bei 37°C inkubiert. Für den Erfolg der Elektroporation ist entscheidend, daß die Zellen in einer Einzellsuspension vorliegen. Hierzu wurden die Zellen durch Auf- und Abpipettieren nach Zugabe von 8 ml Medium vereinzelt. Die Zellen wurden

gezählt und das Trypsin enthaltende Medium wurde durch eine Zentrifugation abgetrennt. Nach Resuspension der Zellen in 10 ml PBS pro 10 cm Schale wurde nochmals die Zellzahl bestimmt, Zellen für die unten beschriebenen Kontrollen abgezweigt und erneut zentrifugiert. Schließlich wurden die Zellen zu $1,25 \times 10^7$ Zellen/ml in 0,8 ml PBS resuspendiert und mit den 50–60 μ l des gelösten Geninaktivierungs-Vektors vermischt. Nach 5 min bei RT wurde die Suspension in eine Elektroporationsküvette überführt und mit einem einzigen Puls bei 230 V und 500 μ F elektroporiert. Die Zellen wurden anschließend durch Aufschlagen der Küvette auf einen harten Untergrund aufgewirbelt und 5 min bei RT stehen gelassen. Der Inhalt einer Elektroporationsküvette wurde auf drei bis fünf 10 cm Schalen und den Kontrollplatten verteilt, die mit EF-Zellen bewachsen waren und frisches ES-Medium enthielten.

Kontrollen:

- A 5×10^5 nicht-transfizierter Zellen wurden auf einer 10 cm Schale ausgesät und mit Geneticin-haltigem Medium kultiviert. Diese Platte dient der Kontrolle der Stringenz der Positivselektion mit Geneticin (D.5.1), dabei dürfen keine Zellkolonien überleben.
- B 10^3 nicht-transfizierte Zellen wurden auf einer 6 cm Platte ausgesät und die Koloniezahl nach einer Woche mit derjenigen von Kontrolle C verglichen, um die Überlebensrate nach der Elektroporation zu bestimmen.
- C 10^3 transfizierte Zellen in einer 6 cm Platte
- D 10^6 transfizierte Zellen wurden auf einer 10 cm Platte ausgesät und nur in Anwesenheit von Geneticin kultiviert. Diese Platte diente der Berechnung der Anreicherung durch die Negativselektion durch Ganciclovir (D.5.1).

C.1.3.3.3 Selektion von ES-Zellen (Mansour et al., 1988)

24 Stunden nach der Elektroporation wurden die Zellen mit Ausnahme der Kontrollen B und C in Anwesenheit von Geneticin (153 μ g/ml) kultiviert. Die negative Selektion mit Ganciclovir mit einer Endkonzentration von 2×10^{-6} M erfolgte an den Tagen 5 bis 8 nach der Transfektion. Das Medium der Zellen wurde täglich gewechselt. Resistente Kolonien waren nach 7–8 Tagen nach der Transfektion sichtbar.

C.1.3.3.4 Isolierung von ES-Zellkolonien

Die für Geneticin und Ganciclovir resistenten ES-Zellkolonien wurden am Tag 11 nach der Transfektion isoliert. Dazu wurde das Medium entfernt und die Platte zunächst mit PBS gewaschen und die Zellen anschließend damit bedeckt. An einer Horizontalstrom-Sterilbank wurden einzelne Kolonien, die einen undifferenzierten Phänotyp zeigten, mittels eines Stereomikroskops identifiziert. Die Kolonien wurden dann mit einem kleinen Volumen (5–15 μ l) PBS von der EF-Schicht gelöst und in 25 μ l der 1x Trypsin-Lösung mit 2% Hühnerserum überführt und maximal 20 bei RT belassen bis für 3–5 min bei 37°C inkubiert wurde. Anschließend wurden 110 μ l ES-Zellmedium auf die trypsinisierten Zellen gegeben und die

Zellen durch energisches Auf- und Abpipettieren vereinzelt. Die Vertiefungen einer 96-Loch-Platte, die mit EF-Zellen bedeckt war, wurden mit jeweils 110 µl frischem ES-Zellmedium beschickt. Die trypsinisierten Zellen wurden nun mit einer Mehrkanalpipette in diese 96-Loch-Platte überführt und die Zellsuspension gut durchmischt. Um eine Replikplatte herzustellen, wurde nun die Hälfte der entstandenen Zellsuspension in eine weitere 96-Loch-Platte mit mitotisch inaktiven EF-Zellen überführt. Die Zellen wuchsen dann 2–3 Tage lang bei täglichem Mediumwechsel im Inkubator. Nach Erreichen der Konfluenz wurden die Klone aus je einer 96-Loch-Platte auf drei 96-Loch-Platten übertragen. Von diesen Platten wurden zwei im subkonfluenten Stadium eingefroren. Die Zellen aus der letzten Platte wurden auf drei gelatinisierte 96-Loch-Platten ohne EF-Zellen übertragen, aus denen im folgenden DNA präpariert wurde.

C.1.3.3.4 DNA-Schnellpräparation aus embryonalen Stammzellen (modifiziert nach Ramírez-Solis et al., 1992)

ES-Zell-DNA wurde konfluent gewachsenen ES-Zellkolonien in 96-Loch-Platten präpariert. Hierzu wurde zunächst das Medium entfernt und die Zellen mit 100 µl PBS gewaschen. Danach wurden je 50 µl ES-Zell-Lysis-Puffer mit frisch hinzugefügter Proteinase K (0,4–1 mg/ml) in alle Vertiefungen gegeben und die Platten üN bei 55°C in einer feuchten Kammer inkubiert. Nach Abkühlung der Platten auf RT wurde die genomische DNA durch Zugabe von 100 µl absolutem Ethanol gefällt und 1 bis 2 Stunden bei RT belassen. Der Überstand wurde dann durch vorsichtiges Invertieren der Platte abgegossen und die DNA dreimal mit 70%igem Ethanol gewaschen. Nach dem letzten Waschschrift wurde der Ethanol weitgehend entfernt und die DNA bei RT 15–20 min an der Luft getrocknet. Für die sich direkt anschließende Restriktion wurde die DNA in 35 µl ES-Restriktionspuffer gelöst und üN bei 37°C inkubiert.

C.2 Arbeiten mit Proteinen

C.2.1 Proteinextraktion aus transfizierten Zellen

C.2.1.1 Erntezeitpunkt

Transfizierte Zellen wurden zum Zeitpunkt der höchsten Aktivität geerntet. Der Verlauf der Aktivitätszunahme der transfizierten Eigenschaft wurde in Vorversuchen bestimmt. Bei Transfektionen mit einem für Luciferase kodierenden Plasmid wurde demnach nach 7 Stunden geerntet, bei Überexpression von CGGBP1 nach 24 Stunden.

C.2.1.2 Rohextrakte aus Zellen aus Gewebekulturschalen (Ø 10 cm)

Zunächst wurde das Medium der adhärent wachsenden Zellen entfernt. Nach zweimaligen Waschen mit PBSd wurden die Zellen mit einem Gummischaber in 1,5 ml PBSd vom Gefäßboden gelöst und die Zellsuspension in ein Zentrifugenröhrchen pipettiert. Nach einer

fünfminütigen Zentrifugation bei 2000 U.p.m. (MC 12 V) bei RT wurden die Zellen erneut mit PBSd gewaschen und zentrifugiert. Der Überstand wurde vollständig entfernt und die Zellen in dem minimalen Restvolumen resuspendiert. Anschließend wurden 10–50 µl 1x Lysispuffer zugegeben und 15 sec gevortext. Nach zehnminütiger Inkubation bei RT wurden das Lysat 10 min bei RT und 12.000 U.p.m. (MC 12 V) zentrifugiert. Der Überstand wurde sofort bei -20°C oder -80°C eingefroren.

C.2.1.3 Lyse der Zellen in der Gewebekulturschale (Ø 3 cm)

Zur Bestimmung der Luciferase-Aktivität wurden die Zellen direkt in der Gewebekulturschale lysiert. Nach Entfernung des Mediums wurden die Zellen zweimal mit etwa 10 ml PBSd gewaschen. Der Waschpuffer wurde vollständig entfernt und die Zellen *in situ* in 150 bis 200 µl 1x Lysispuffer 10 min bei RT inkubiert. Das Lysat wurde dann mittels eines Gummischabers vom Gefäßboden gelöst. Nach 15 sekündigem Vortexen wurden die Ansätze 15 sec bei maximaler Geschwindigkeit (MC 12 V) bei RT zentrifugiert. Der Überstand wurde bei -80°C gelagert.

C.2.2 Bestimmung der Proteinkonzentration (Bradford, 1976)

Die Proteinkonzentration wurde mit Coomassie Brilliant Blue G-250 nach Bradford (1976) bestimmt. Das Absorptionsmaximum dieses Farbstoffes verschiebt sich von $\lambda=495$ nm auf $\lambda=595$ nm nach Bindung an Proteine. Hierzu wurde die Coomassie Protein Assay-Lösung (Biorad) nach Vorschrift verdünnt. Für jede neue Verdünnung wurden Eichkurven mit Rinderserumalbumin erstellt. Da das im Lysispuffer enthaltene Triton X-100 eine Eigenextinktion bei 595 nm aufweist, die abhängig von der Triton X-100-Konzentration ist, wurden Eichkurven mit Rinderserumalbumin und dem entsprechenden zur Messung eingesetzten Volumen Lysispuffer erstellt. Für die Bestimmung der Proteinkonzentration in den Zell-extrakten wurden zwischen 5 und 20 µl der fraglichen Proteinlösung mit 1 ml der verdünnten Coomassie-Protein-Assay-Lösung versetzt. Nach gründlichem Durchmischen und mindestens 5 min Inkubation bei RT wurde die optische Dichte bei 595 nm gemessen. Als Nullprobe diente das selbe Volumen des Puffers ohne Protein. Zur Proteinquantifizierung wurden nur Meßwerte berücksichtigt, die im linearen Bereich der Eichkurve lagen.

C.2.3 Luciferase-Aktivitätsbestimmungen

Das Luciferasegen wurde in Transfektionsanalysen als Reporter gen verwendet. Luciferase (*luc*) aus *Photinus pyralis* wandelt in Anwesenheit von molekularem Sauerstoff, ATP und Magnesium Luciferin in Oxyluciferin und CO₂ um. Bei dieser Reaktion wird Licht von einer Wellenlänge von 562 nm emittiert, das in einem Luminometer gemessen werden kann. Von der gemessenen Lichtmenge kann man indirekt auf die Transkriptionsrate des Luciferase-Reporter gens und somit auf die Stärke eines Promotors schließen. Lyophilisiertes Substrat aus

dem Luciferase Assay System der Firma Promega wurde gemäß Herstellerangaben im mitgelieferten Puffer gelöst und bei -80°C gelagert. Nach der Transfektion von HeLa-Zellen mit dem Plasmid pGL2basic (Promega) und dessen Derivaten wurden Rohextrakte hergestellt (C.2.1) und $20\ \mu\text{l}$ davon mit Hilfe des Luminometers LB 9501 analysiert. Nach dem Auftauen der Proben auf Eis wurden $20\ \mu\text{l}$ in Meßröhrchen vorgelegt und mindestens 20 min auf RT äquilibriert. Die Schläuche im Luminometer wurden vor der Messung erst mit Wasser und anschließend mindestens zweimal mit Substratpuffer, der ebenfalls RT hatte, gespült. Im Rohdatenzklus des Luminometers wurden $100\ \mu\text{l}$ Substratpuffer automatisch zu den vorgelegten $20\ \mu\text{l}$ Proteinlösung gegeben und die Lichtemission 10 sec lang gemessen. Als Kontrollen dienten Rohextrakte von mock-transfizierten Zellen bzw. Rohextrakte von Zellen, die mit dem pGL2basic oder dem pGL2control-Plasmid transfiziert worden waren. Das pGL2control Plasmid trägt das luc-Gen unter der Kontrolle des SV40 *immediate early* Promotors, während der pGL2basic Vektor keinen eukaryontischen Promotor enthält.

C.2.4 Gelretardation

Protein-DNA-Interaktionen wurden mittels der Gelretardations-Technik (*electrophoretic mobility shift assay*; EMSA) dokumentiert. Protein-DNA-Komplexe weisen im elektrischen Feld eine geringere Mobilität auf, als freie DNA, von der sie dann gut unterschieden werden können. Reaktionsansätze von 5 bis 10 fmol radioaktiv markierten doppelsträngigen synthetischer Oligodesoxyribonukleotide und $1\ \mu\text{g}$ PolydAdT wurden mit 5 bis $10\ \mu\text{g}$ Zellrohextrakten in 1x HDNG-Puffer inkubiert (nach Deissler et al., 1996). In Kompetitionsexperimenten wurde nicht markierte doppelsträngige DNA in 30–300fachem molaren Überschuß zugegeben. Der Ansatz enthielt direkt den Ladepuffer. Die Ansätze wurden auf ein 3 bis 5%iges Polyacrylamidgel mit 0,5x TBE geladen, das bei 150–300 V bereits mindestens 30 min gelaufen war. Nach der Elektrophorese wurde das Gel auf 3MM-Blot Papier aufgezogen und unter Vakuum bei 80°C getrocknet. Anschließend wurde ein Röntgenfilm mit dem Gel bei Raumtemperatur exponiert.

C.3 Präparation von Nukleinsäuren

C.3.1 Präparation von genomischer DNA

C.3.1.1 Präparation von genomischer DNA aus kultivierten Zelllinien

Zellen, die in Gewebekulturflaschen bis zur Konfluenz gewachsen waren, wurden zweimal mit wenigen ml Tris-Saline gewaschen. Danach wurden die Zellen mit 3 bis 5 ml 37°C warmer Tris-Saline, die 2 mM EDTA enthielt, inkubiert. Man löste die Zellen durch Klopfen von der Gefäßwand und überführte sie in ein 13 ml-Reaktionsgefäß (Weißkapp). Verbleibende Zellen wurden mit weiteren 5 ml Tris-Saline aus der Gewebekulturflasche gespült und mit der zuerst entnommenen Zellsuspension vereinigt. Nach Zentrifugation

(Heraeus Labofuge GL, 2.500 U.p.m., 4°C, 10 min) wurde das Zellsediment in frischer Tris-Saline aufgenommen und erneut zentrifugiert. Dann wurden die Zellen in Proteinase-K-Puffer resuspendiert und entweder bei -20°C gelagert oder direkt mit 1/10 des Volumens an Proteinase-K-Stammlösung (10 mg/ml) versetzt. Mit Proteinase K wurde mindestens 3 Stunden unter Rotieren bei 50°C inkubiert. Anschließend wurde die Proteinase K durch 20-minütiges Erhitzen auf 68°C inaktiviert, da RNA im folgenden Schritt durch eine 20 minütige Inkubation mit RNase A (200 µg/ml) bei 37°C abgebaut wurde. Die DNA wurde durch eine Phenol/ Chloroform- und eine Chloroform-Extraktion gereinigt (C.3.1.3) und mit Ethanol präzipitiert (C.3.4). Die präzipitierte genomische DNA wurde auf eine zugeschmolzene Glas-kapillare aufgespult, in 70%igem Ethanol gewaschen und an der Luft getrocknet. Danach wurde die DNA bei 37°C in einem angemessenen Volumen TE (50–200µl) gelöst und bei 4°C gelagert.

C.3.1.2 Präparation von DNA aus Organen

Die verwendeten Gewebe mußten direkt nach Entnahme in flüssigem Stickstoff schockgefroren und bei -80°C gelagert werden, da die genomische DNA sonst degradierte. Die tiefgefrorenen Gewebe wurden mit einem Skalpell in Würfel geschnitten und in Proteinase-K-Puffer mit frisch zugesetztem Enzym überführt (C.3.1.1). Anschließend inkubierte man die Ansätze 16 bis 18 Stunden unter Rotieren bei 50°C. Die Aufreinigung der DNA ist in Abschnitt 3 beschrieben (C.3.1.3), außer daß in diesem Falle der Phenol/Chloroform-Extraktion ein reiner Phenolschritt vorausging, und – je nach Intensität der Interphase – die Phenol/Chloroform Extraktion zweimal wiederholt wurde.

C.3.1.3 Reinigung von DNA mit Phenol/Chloroform

Die Phenol/Chloroform-Extraktion dient der Abtrennung von Proteinen und Lipiden von der DNA. Dazu wurde die DNA-Lösung mit dem gleichen Volumen Phenol/Chloroform/Isoamylalkohol (25:24:1) versetzt und 10 Minuten im Rotator vermischt. Die Phasen wurden durch eine Zentrifugation bei 3.000 U.p.m (Medifuge, 10 Minuten, Raumtemperatur) getrennt. Die DNA-haltige Oberphase wurde ohne die Protein-haltige Interphase abgenommen und in ein neues Reaktionsgefäß überführt. Auf die gleiche Weise wurde mit einem Volumen Chloroform extrahiert, wobei Reste von Phenol aus der wäßrigen Phase entfernt wurden.

C.3.2 Präparation von Plasmid-DNA

Bakterien wurden in dYT oder LB gezüchtet. Entsprechend der vom jeweiligen Plasmid vermittelten Resistenz wurden dem Medium 100 µg Ampicillin oder 30 µg Kanamycin pro ml beifügt. Die Bakterienkulturen wurden immer von einer Einzelkolonie ausgehend angezüchtet.

C.3.2.1 Schnellpräparation von Plasmid-DNA

Die folgenden Methoden zur Schnellpräparation von Plasmid-DNA wurden hauptsächlich verwendet. In einigen Fällen, v.a. wenn die Plasmide anschließend mit salzempfindlichen Enzymen behandelt werden sollten, wurde das Qiaprep Spin Miniprep Kit gemäß den Herstellerangaben verwendet.

C.3.2.1.1 Zahnstocher-Methode (Jennifer Hertz, persönliche Mitteilung)

Bei einem deutlichen Größenunterschied der rekombinanten DNA vom selbstligierten Vektor konnte zur schnellen Überprüfung transformierter Bakterien die Zahnstocher-Methode verwendet werden. Sie diente vor allem der Anreicherung von positiven Klonen vor einer der im Folgenden beschriebenen Präparationsmethoden. Superspiralisierte Plasmid-DNA wurde aus einer Einzelkolonie präpariert, die mindestens eine Größe von 1 mm² aufweisen mußte. Die Zellen wurden mit einem Zahnstocher an der Wand eines 0,5 ml Gefäßes abgestrichen, wobei der gleiche Zahnstocher dazu verwendet wurde, die Kolonie auf einer neuen Platte zu rekultivieren oder Medium anzuimpfen. Durch Vortexen wurden die Zellen in 10 µl Lysispuffer resuspendiert und 15 min bei 65°C erhitzt. Anschließend wurden die Proben kurz zentrifugiert und auf ein Agarosegel aufgetragen, das erst nach einem kurzen Vorlauf mit Laufpuffer überschichtet wurde. Negative Klone liefen auf der gleichen Höhe wie der als Referenz aufgetragene Vektor ohne Insert.

C.3.2.1.2 Kochlyse (modifiziert nach Holmes und Quigley, 1981)

Eine Übernachtskultur (1,5 ml) wurde kurz bei 12.000 U.p.m bei Raumtemperatur zentrifugiert (MC 12V). Nach vollständiger Entfernung des Mediums wurden die Bakterien mit 0,5 ml STET-Puffer und 40 µl Lysozym-Lösung versetzt und mittels eines Vortex-Mischers vollständig resuspendiert. Nach 90 sec Inkubation in einem kochenden Wasserbad wurden die Ansätze 10 min bei Raumtemperatur bei 12.000 U.p.m. (MC 12 V) zentrifugiert. Der Überstand wurde in ein neues Reaktionsgefäß überführt und die darin enthaltene DNA mit 40 µl 3 M Na-Acetat und 0,5 ml Isopropanol präzipitiert und 12 min bei 12.000 U.p.m. (Medifuge 13000) pelletiert. Das DNA-Präzipitat wurde mit 70%igem Ethanol gewaschen, getrocknet und in 30 µl Chromatographiewasser gelöst, dem jeweils 10 µg RNase beigefügt waren. Davon reichten üblicherweise 1–3 µl für eine Restriktionsanalyse und eine Sequenzreaktion aus.

C.3.2.1.3 Alkalische Lyse (Indrikis Muiznieks, persönliche Mitteilung)

Eine Übernachtskultur (1,5 ml) wurde kurz bei 12.000 U.p.m zentrifugiert (MC 12V) und das Medium vollständig entfernt. Das Sediment wurde in 200 µl 10 mM EDTA, pH 8 aufgenommen und durch Pipettieren resuspendiert. Anschließend wurde das gleiche Volumen an frisch angesetztem Lysispuffer beigemischt, die Suspension durch vorsichtiges Invertieren

der Gefäße durchmischt und 2 min bei RT inkubiert. Mit 0,2 ml 3 M NaAcetat pH 4,5–5 wurden die Ansätze neutralisiert und dann 10–15 min auf Eis gelagert. Anschließend wurden die Proben (Plasmide < 7 kb) für 2 min bei 72°C erhitzt. Nach Zugabe von 0,2 ml 6 M KAcetat pH 8,0 wurden die Proben bei -20°C 1 bis 1,5 Stunden oder für 20-30 min bei -80°C eingefroren. Danach wurde 15–20 min bei 12.000 U.p.m. (MC 12V) zentrifugiert, wobei die Proben auftauten. Der Überstand wurde in ein neues Reaktionsgefäß überführt, in dem 0,7 ml Isopropanol vorgelegt worden war. Direkt danach wurde 10–15 min bei maximaler Geschwindigkeit zentrifugiert. Nach vollständiger Entfernung des Überstandes wurde das DNA-Präzipitat mit 70%igem Ethanol gewaschen, getrocknet und in 50 µl Chromatographiewasser gelöst. Für eine Restriktionsanalyse mit gleichzeitiger RNase A Behandlung wurden üblicherweise 5 µl eingesetzt.

C.3.2.2 Präparation größerer Mengen von Plasmid-DNA

Zur Präparation größerer Mengen von Plasmid-DNA wurden die Plasmid MIDI oder MAXI-Kits der Firmen Gibco Life Technologies oder Qiagen verwendet. Es wurden die vom Hersteller empfohlenen Mengen an Übernachtskultur eingesetzt. Das Protokoll beruht auf der Lyse der Bakterien unter alkalischen Bedingungen mit SDS. Unlösliche Zellbestandteile der neutralisierten Lysate wurden durch Zentrifugation abgetrennt. Der Überstand wurde anschließend auf eine Säule mit einem Anionenaustauscherharz aufgetragen. Nach Bindung der Plasmid-DNA wurden Verunreinigungen durch Waschschriffe entfernt. Zur Präzipitation der DNA wurde abweichend vom angegebenen Protokoll das Eluat in 50 ml Rörchen bei 5000 U. p. m. 1 Stunde lang bei 4°C zentrifugiert. Der Geninaktivierungs-Vektor für die Transfektion embryonaler Stammzellen wurde nach der CsCl-Methode (Sambrook und Russel, 2001) präpariert. Dies wurde freundlicherweise von Doris Renz durchgeführt.

C.3.2.3 Synthetische Oligodesoxyribonukleotide

Synthetische DNA-Einzelstränge wurden von den Firmen Gibco BRL, MWG Biotech und Sigma-ARK hergestellt und bei Bedarf 5'-phosphoryliert. Sie wurden in Chromatographiewasser rekonstituiert und bei -20°C gelagert. Manche Oligodesoxyribonukleotide wurden mit ihrem komplementären Gegenstrang hybridisiert, damit sie als Doppelstrang z.B. bei Klonierungen oder bei Gelretardierungs-Experimenten (C.2.4) eingesetzt werden konnten. Für Klonierungszwecke wurden komplementäre Einzelstränge im gleichen molaren Verhältnis aufgeköcht, wonach die Flüssigkeit langsam auf RT abkühlte. Oligodesoxyribonukleotide, die in Gelretardierungs-Experimenten verwendet wurden, wurden wie in Deissler et al., 1996 beschrieben, hybridisiert. Nachdem in einem Teil des Ansatzes die Hybridisierung mittels 5%iger NuSieve-Agarosegele kontrolliert worden war, wurden die Doppelstränge aus einem Polyacrylamidgel gereinigt (C.3.2.4.2). Die hybridisierten Oligodesoxyribonukleotide wurden bei -20°C gelagert.

C.3.2.4 Isolation von DNA-Fragmenten aus Gelen

C.3.2.4.1 Isolation von DNA-Fragmenten aus Agarosegelen

Die zu eluierende DNA wurde auf ein präparatives Agarosegel aufgetragen, um etwaige Nebenprodukte, wie Primer oder DNA-Fragmente anderer Größen abzutrennen. Die DNA wurde mit Ethidiumbromid angefärbt und die gewünschten Banden mit jeweils neuen Skalpellklingen ausgeschnitten. Die Agarosestücke wurden entweder bei -20°C gelagert oder direkt mit einem Gelextraktionskit weiterverarbeitet. Zur Aufreinigung von DNA wurde das Qiaquick Gelextraktionskit (Qiagen) oder das NucleoSpin Extract (Macherey und Nagel) den Herstellerangaben folgend verwendet. Meist wurde mit dem geringsten möglichen Volumen an 10 mM Tris, pH 8.5 oder Chromatographiewasser eluiert.

C.3.2.4.2 Isolation von DNA-Fragmenten aus Polyacrylamidgelen und *UV-Shadowing*

Mit dieser Technik wurden 51 bp lange doppelsträngige Oligodesoxyribonukleotide aufgereinigt. In 1–2 cm breiten Geltaschen wurden 8 bis 15 µg der hybridisierten Oligodesoxyribonukleotide aufgetragen und eine Elektrophorese mit einem 5%igen Polyacrylamidgel (1 mm Dicke) in 1x TBE bei ca. 200 V durchgeführt. Die Doppelstränge wurden durch UV-Shadowing visualisiert. Das Polyacrylamidgel wurde dabei auf eine mit einem Fluoreszenzindikator beschichtete Kieselgelplatte (Macherey und Nagel) gelegt, die von oben durch eine kurzwelligen UV-Lampe angestrahlt wurde und dann grün leuchtete. Der DNA-Doppelstrang absorbierte die UV-Strahlung, und die Kieselgelplatte darunter erschien dunkelrot. Mit einer Skalpellklinge wurde das entsprechende Polyacrylamidstück ausgeschnitten (ca. 0,5–1 cm²) und in 1,5 ml Eppendorfgefäßen mit ca. 500 µl Elutionspuffer üN bei 25°C eluiert. Zur Präzipitation wurde das zweifache Volumen an absolutem Ethanol zugegeben und üN bei -80°C inkubiert. Nach Waschen in 70%igem Ethanol wurden die aufgereinigten DNA-Fragmente in Wasser gelöst.

C.3.3 Präparation von RNA

C.3.3.1 Präparation von Gesamt-RNA aus kultivierten Zellen

Gesamt-RNA aus kultivierten Zellen wurde unter Verwendung des NucleoSpin-RNA-L-Kits der Firma Macherey und Nagel gemäß den Herstellerangaben gereinigt. Die Qualität der erhaltenen RNA wurde auf einem Agarosegel auf Degradation (C.6.1.2) und in RT-PCR-Experimenten (C.5.9) ohne Zusatz von Reverser Transkriptase auf Kontaminationen durch genomische DNA hin kontrolliert.

Sollte RNA aus undifferenzierten ES-Zellen präpariert werden (C.1.3.3), mußten die ES-Zellen von den EF-Zellen getrennt werden. Nachdem die Zellen trypsinisiert worden waren, wurde die ES/EF-Zellsuspension in Wachstumsmedium dreimal 35 min lang in einer

Zellkulturschale inkubiert. EF-Zellen haften innerhalb von etwa 20 min am Gefäßboden, während ES-Zellen sich erst nach etwa 45 min absetzen. Der Anteil von EF-Zellen konnte somit in der ES-Zellsuspension stark verringert werden.

C.3.3.2 Präparation von mRNA aus RNA-Präparationen

mRNA wurde unter Verwendung von mRNA-Isolationskits der Firmen Roche und Macherey und Nagel aus isolierter Gesamt-RNA angereichert. Beide Kits beruhen auf der Bindung von Oligo-dT-Primern an die polyadenylierten mRNAs, wodurch eine Abtrennung von den anderen RNA-Spezies ermöglicht wird. Mit beiden Kits konnte bei Durchführung des Standard-Protokolls keine deutliche Anreicherung von mRNA erreicht werden.

C.3.4 Ethanolpräzipitation von Nukleinsäuren

DNA wurde durch Zugabe von 1/10 Volumen 3 M NaAcetat (pH 5,2) und dem 2,5-fachen Volumen an absolutem, -20°C kaltem Ethanol präzipitiert. Bei RNA wurde dafür 2 M NaAcetat pH 4,2 verwendet (Endkonzentration 0,3 M). Alternativ wurde DNA auch mit NH₄Acetat (Endkonzentration 2,5 M) und Ethanol präzipitiert, v.a. wenn salzempfindliche Reaktionen folgen sollten. Je nach Menge der DNA wurde sofort oder erst nach einer Inkubation bei -20°C (30 Minuten bis üN) bei 4°C, 12.000 U.p.m (MC 12V) mindestens 30 Minuten zentrifugiert. Die DNA wurde mit 70%igem Ethanol gewaschen und unter Vakuum oder an der Luft getrocknet. Je nach der sich anschließenden Reaktion wurde die DNA in Chromatographiewasser oder in 10 mM Tris/HCl pH 8–8,5 gelöst, andernfalls in TE. RNA wurde in RNase-freiem Wasser gelöst.

C.3.5 Bestimmung der Nukleinsäurekonzentration

C.3.5.1 Photometrische Bestimmung der Nukleinsäurekonzentration (Kaiser und Hogness, 1960)

Die Konzentration von Nukleinsäuren in wässriger Lösung wurde photometrisch durch Messung der Absorption bei einer Wellenlänge von $\lambda=260$ nm (OD_{260}) und $\lambda=280$ nm (OD_{280}) in Quarzküvetten bestimmt. Eine OD_{260} von 1 entspricht einer Konzentration von 50 $\mu\text{g/ml}$ doppelsträngiger DNA, 27–35 $\mu\text{g/ml}$ einzelsträngiger DNA und 40 $\mu\text{g/ml}$ reine einzelsträngige RNA. Der Reinheitsgrad einer Lösung kann durch den Quotienten OD_{260}/OD_{280} ermittelt werden. Reine DNA-Präparationen ergeben einen Wert von 1,8, reine RNA-Präparationen einen Wert von 2,0. Der Wert verringert sich bei Kontamination durch Proteine oder Phenol.

C.3.5.2 Bestimmung der DNA-Konzentration durch ein Tüpfelgel

Zur Bestimmung der Konzentration von geringkonzentrierten DNA-Lösungen wurden 1–3 µl der gereinigten DNA mit einer DNA bekannter Konzentration entweder auf ein Agarosegel aufgetragen und einer Elektrophorese unterzogen oder auf ein Tüpfelgel (1,5% TBE-Agarose mit etwa 330 ng/ml Ethidiumbromid) pipettiert und die Konzentration aufgrund der Fluoreszenz-Intensität nach UV-Licht-Bestrahlung abgeschätzt.

C.4 Genomisches Sequenzieren nach der Bisulfit-Methode (Frommer et al., 1992; Clark et al., 1994)

Die Technik des genomischen Sequenzierens nach der Bisulfit-Methode ermöglicht es, jedes Desoxycytidin eines DNA-Stranges auf seinen Methylierungszustand hin zu untersuchen. Nur unmethylierte Desoxycytidine werden durch Bisulfit in Desoxyuridine umgewandelt, während 5-Methyl-Desoxycytidine die Behandlung unverändert überstehen. Somit kann auch die überwiegende Mehrheit von Desoxycytidinen untersucht werden, die nicht in einer Erkennungssequenz einer 5-Methyl-Desoxycytidin-sensitiven Restriktionsendonuklease liegen.

Genomische DNA aus peripheren Blutzellen und HeLa, die bereits nach dem folgenden Protokoll mit Bisulfit umgesetzt worden war, wurde freundlicherweise von Ralph Remus zur Verfügung gestellt.

C.4.1 Behandlung der DNA mit Bisulfit (entnommen aus Remus, 2000)

5 µg genomische DNA wurde mit Chromatographiewasser auf ein Gesamtvolumen von 64 µl aufgefüllt. Nach Zugabe von 2 µl 10 M NaOH (Endkonzentration 0,3 M) wurde die DNA zuerst 15 min bei 37°C inkubiert und dann zwei Minuten im kochenden Wasserbad vollständig denaturiert. Nach sofortigem Abkühlen in Eiswasser wurden 1,2 ml Bisulfit-Lösung (8,5 g NaHSO₃ in 16 ml entgastem Chromatographiewasser gelöst, 1 ml 40 mM Hydrochinon zugesetzt und mit 0,6 ml 10 M NaOH auf pH 5 eingestellt; jeweils frisch angesetzt und mit N₂ begast), das Reaktionsgemisch mit Mineralöl überschichtet und das Reaktionsgefäß mit Parafilm versiegelt. Es folgte eine 16-stündige Inkubation bei 55°C in einem Wasserbad mit geschlossenem Deckel. Nach Entfernen des Mineralöls wurde die DNA mittels 5 µl Glasmilch (GeneClean II, Bio 101) nach Angaben des Herstellers aufgereinigt. Das getrocknete Glasmilch-Sediment wurde in 100 µl Chromatographiewasser resuspendiert und zur Freisetzung der DNA 5 min bei 55°C. Nach einminütiger Zentrifugation (12.000 x g) wurde der Überstand ohne Reste von Glasmilch in ein neues Reaktionsgefäß überführt. Zum Beenden der Bisulfit-Behandlung wurden 3 µl 10 M NaOH zugegeben (Endkonzentration 0,3 M) und 10 min bei 37°C inkubiert. Danach wurde die umgesetzte DNA mit 100 µl 6 M NH₄OAc und 400 µl eiskaltem Ethanol (absolut) präzipitiert und einige Stunden bei -70°C inkubiert. Nach einer halbstündigen Zentrifugation (13.000 x g, 4°C) wurde das Sediment mit

0,5 ml 70% Ethanol versetzt, zentrifugiert (15 min, 13.000 x g, 4°C) und bei Raumtemperatur getrocknet. Dann wurde es in 20 µl Chromatographiewasser gelöst und bei -20°C gelagert. Für eine anschließende PCR-Reaktion wurde 1 µl dieser DNA-Lösung (entsprechend 250 ng der eingesetzten DNA) verwendet. Diese teilweise recht drastischen Reaktionsbedingungen führten zu einer Degradation von schätzungsweise 99% der eingesetzten DNA.

C.4.2 PCR im Rahmen des genomischen Sequenzierens (modifiziert nach Remus, 2000)

Die mit Bisulfit behandelte DNA wurde in einer PCR eingesetzt. Aufgrund der starken Degradation der genomischen DNA nach der Behandlung mit Bisulfit wurden für die zu amplifizierenden Abschnitte je zwei Primerpaare ausgewählt, wobei das eine Paar die innere Region des anderen Paares amplifizierte (*nested PCR*). Diese Strategie erhöht die Ausbeute und Spezifität einer PCR-Reaktion. Es wurde 1 µl der mit Bisulfit behandelten DNA in der ersten PCR-Reaktion eingesetzt, die den 1x Puffer der jeweiligen Taq-Polymerase enthielt (Sigma, Qiagen). Alle Reaktionen wurden mit 1,7 mM MgCl₂ und 0,1 mM der dNTPs angesetzt. Für die Amplifikation einiger Fragmente war es nötig, den Zusatz Q-Solution der Firma Qiagen zu verwenden, der die Ausbeute der PCR deutlich erhöhte. Die Zusammensetzung der Q-Solution ist nicht bekannt. Von der ersten PCR-Reaktion wurde dann 1 µl zur Amplifikation der inneren Region unter fast identischen Bedingungen eingesetzt. Die beiden aufeinanderfolgenden PCR-Reaktionen unterschieden sich nur in der Temperatur zur Anlagerung der Primer. In der ersten PCR erfolgte dieser Schritt bei 53°C, in der zweiten PCR bei 57°C. Das PCR-Programm für alle Regionen verlief wie folgt:

Denaturierung	94°C	2 min	
35 Zyklen	94°C	15 sec	Denaturieren
	5x°C	15 sec	Anlagerung der Primer
	72°C	30 sec	Elongation
Elongation 72°C	5 min		
Abkühlen auf	4°C		

Der gesamte PCR-Ansatz wurde danach auf ein 1–1,5%iges Agarosegel aufgetragen und Produkte mit der erwarteten Größe wurden mit einer Skalpellklinge ausgeschnitten und aus dem Gel extrahiert.

Die Produkte wurden anschließend mit dem pGEM-T-Vektor ligiert (C.11.1.2) und XL1blue MRF' Zellen damit transformiert (C.11.3). Durch eine Subklonierung der PCR-Produkte war es möglich, den Methylierungszustand einzelner Moleküle zu bestimmen. Aus Einzelkolonien wurde schließlich DNA präpariert (C.3.2), deren Sequenz unter Verwendung des ABI Prism[®] BigDye[™] Terminator Cycle Sequencing Ready Reaction-Kits bestimmt wurde (C.5.8).

C.5 Enzymatische Reaktionen mit Nukleinsäuren

Soweit nicht anders angegeben wurden DNA-modifizierende Enzyme (Restriktionsenzyme, DNA-Polymerasen, Ligase, Phosphatase, Nuklease und 5'-Polynukleotidkinase) unter den bei den verschiedenen Herstellern angegebenen Bedingungen eingesetzt.

C.5.1 Restriktionsspaltung von DNA (Nathans und Smith, 1975)

DNA wurden mit der angemessenen Menge Enzym (3–5 U/ μ g für Plasmid- bzw. 5–10 U/ μ g für genomische DNA, entsprechend 1–1,5 U/ μ l) unter den von den Herstellern angegebenen Bedingungen inkubiert. Das Volumen der Ansätze entsprach mindestens dem Zehnfachen der eingesetzten Enzymlösung. Die Reaktionen wurden nach mindestens 3 Stunden (oder 16 bis 18 Stunden) bei genomischer DNA bzw. nach mindestens 30 Minuten bei Plasmid-DNA beendet. Die Reaktionsansätze mit genomischer DNA wurden durch Zugabe von 1/10 Volumen NaAcetat und 2,5-fachen Volumen Ethanol präzipitiert, andere Reaktionen wurden durch Zugabe von 1/6 des Volumens an Stoppuffer beendet.

C.5.2 Polymerase Kettenreaktion (Saiki et al., 1988)

Für die Amplifizierung von DNA mit der Polymerase Kettenreaktion (PCR; *polymerase chain reaction*) wurden ca. 200 ng genomische DNA, 25–50 pg Plasmid-DNA oder 150 pg λ -DNA mit je 50 ng der jeweiligen Primer in einem Endvolumen von 25 μ l eingesetzt. Die Primer waren zwischen 18 und 24 nt lang mit einem GC-Gehalt zwischen 42 bis 70%. Der Reaktionsansatz enthielt jeweils den 1x Puffer des jeweiligen Hersteller (Promega, Sigma, Qiagen) und 0,2 mM der dNTPs. Die optimale $MgCl_2$ -Konzentration wurde für jedes Primerpaar individuell bestimmt. Bei Regionen mit hohem GC-Gehalt wurde der Reaktion bei Bedarf 5% DMSO oder Q-Solution (Qiagen) zugesetzt.

Die Amplifikation erfolgte in 30–40 Zyklen des im folgenden beschriebenen PCR-Protokoll in der PCR-Maschine Geneamp PCR System 9600 (Perkin Elmer). Die Anlagerungstemperatur der jeweiligen Primerpaare wurde für diese optimiert. Die Elongationszeit wurde der Größe des zu amplifizierenden Fragmentes angeglichen (ca. 1 min/kb).

Denaturierung	94°C	2 min	
35 Zyklen	94°C	30 sec	Denaturieren
	5x°C-63°C	30 sec	Anlagerung der Primer
	72°C	30 sec -4 min	Elongation
Elongation	72°C	5 min	
Abkühlen auf	4°C		

C.5.3 *In-vitro*-Methylierung von DNA (Vardimon et al., 1982)

Zur *in-vitro*-Methylierung wurde Plasmid-DNA mit 2,5 U der eingesetzten Methylase pro μg DNA in dem vom Hersteller mitgeliefertem Puffer mit der benötigten Menge S-Adenosyl-Methionin (SAM) μN inkubiert. Aufgrund der geringen Halbwertszeit des SAM bei 37°C , wurde den Reaktionen nach 4 Stunden frisches SAM zugesetzt. Eine parallel angesetzte Reaktion ohne SAM diente als Negativkontrolle (*mock*-Methylierung). Anschließend wurden die Reaktionen mit Phenol/Chloroform mit einer Rückextraktion gereinigt (C.3.1.3). Die DNA wurde dann präzipitiert (C.3.4), in TE resuspendiert und bei 4°C gelagert. Der Erfolg der *in-vitro*-Methylierung wurde in Restriktionsanalysen mit methylierungssensitiven Restriktionsendonukleasen überprüft. Jeweils 500 ng der Plasmide wurden hierzu 2 bis 4 h bei 37°C mit den entsprechenden methylierungssensitiven Restriktasen inkubiert (C.5.1). Die HhaI- bzw. HpaII-methylierten Plasmide wurden mit dem jeweils entsprechenden Restriktionsenzym behandelt. Es gibt kein zu SssI äquivalentes Restriktionsenzym, da die Erkennungssequenz der Methylase nur aus dem Dinukleotid CpG besteht. SssI-methylierte Plasmide wurden hier dann mit HhaI und HpaII gespalten, um näherungsweise den Methylierungszustand zu überprüfen. Ein Teil der Ansätze wurde dann auf $1 \text{ ng}/\mu\text{l}$ verdünnt. Beide Ansätze wurden anschließend auf verschiedene 1,5%ige Agarosegele aufgetragen. In der ersten Kontrolle wurde das eine Gel mit Ethidiumbromid gefärbt und unter UV-Licht dokumentiert. Um kleinste Mengen nicht methylierten Plasmids auszuschließen, wurde das Gel mit den $1 \text{ ng}/\mu\text{l}$ Ansätzen in einer DNA-Transfer-Analyse verwendet. Die *mock*-methylierten Ansätze dienten jeweils als nicht methylierte Referenz.

C.5.4 Modifikation von DNA-Enden

Um nicht kompatible DNA-Enden klonieren zu können, war es notwendig, "stumpfe" DNA-Enden zu generieren. Durch die Modifikation von DNA-Enden konnten auch neue Restriktionsschnittstellen hergestellt werden. Alle folgenden Reaktionen wurden unter denen vom Hersteller empfohlenen Bedingungen (Puffer, $\text{U}/\mu\text{g}$ DNA, Inkubationsdauer und -temperatur) vorgenommen. Ein 5'-Überhang wurde entweder mit DNA-Polymerase I (Klenowfragment) oder der T4-DNA-Polymerase aufgefüllt. Ein 3'-Überhang wurde durch die 3'-5'-Exonuklease-Aktivität der T4-DNA-Polymerase in ein stumpfes Ende umgewandelt. Alternativ dazu wurden beide Überhänge mit der S1-Nuklease entfernt.

C.5.5 Einführen von fortlaufenden Deletionen

Exonuklease III greift bevorzugt 5'-Überhänge an DNA-Enden an, während 3'-Überhänge weitgehend unverändert bleiben. Diese Eigenschaft der Exonuklease III (ExoIII) wurde ausgenutzt, als fortlaufende Deletionen in ein DNA Molekül eingeführt werden sollten. Gemäß den Angaben der Hersteller des ExoIII/S1 Deletion Kits (Fermentas MBI) wurde das Ausgangsplasmid mit KpnI und XhoI behandelt, wobei der zu erhaltende Bereich mit dem 3'-

Überhang von KpnI geschützt wurde und der zu deletierende Bereich von dem 5'-Überhang von XhoI flankiert wurde. Nach der Linearisierung wurde das Plasmid durch eine Phenol-Chloroform Extraktion aufgereinigt (C.3.1.3) und mit Ethanol präzipitiert (C.3.4). Mit Exonuklease III im Restriktionspuffer 2 von NEB (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂ und 1 mM DTT) wurde bei 25°C, 30°C und 37°C bis zu 25 min inkubiert. In Vorversuchen waren geeignete Intervalle bestimmt worden, in denen dann alle 15 sec ein Aliquot aus der Reaktion entnommen wurde. Da die Exonuklease III nur einen Strang des doppelsträngigen DNA-Moleküls angreift, wurde der verbliebene Einzelstrang mit Hilfe der Nuklease S1 innerhalb von 30 min bei RT abgebaut. Beide Enzyme wurden schließlich 10 min bei 70°C inaktiviert. 7 µl der 10 µl Reaktionsansätze wurden einer Gelelektrophorese unterzogen. Im Vergleich zum linearisierten Ausgangsplasmid und dem linearisierten Vektor ohne Insert wurden für eine Ligation geeignete Plasmide ausgewählt. Die Ligation mit den verbliebenen 3 µl erfolgte gemäß den Angaben des Herstellers in den mitgelieferten Puffern. Die Hälfte der Ligation wurde zur Transformation in *E. coli* eingesetzt. Nach Präparation der Plasmid-DNA und Bestimmung der ungefähren Größe der eingeführten Deletionen wurden schließlich nach Sequenzierung die Plasmide mit geeigneten Deletionen für eine Transfektion in eukaryontische Zellen ausgewählt.

C.5.6 Dephosphorylierung

Um eine Religation einfach geschnittener Vektoren zu vermeiden, wurden, wenn nicht anders beschrieben, Vektoren nach dem entsprechenden Restriktionsansatz im selben Puffer mit 1 U alkalischer Phosphatase/µg DNA für 1 h bei 37°C inkubiert. In einer Ligation kann ein Insert mit seinem 5'-Phosphat immer noch mit dem dephosphorylierten Vektor verknüpft werden. Die verbleibenden nicht verknüpften Enden wurden nach erfolgreicher Transformation dann durch bakterielle Enzyme geschlossen. Die alkalische Phosphatase wurde anschließend durch 20 min Inkubation bei 65°C inaktiviert. Der Vektor wurde in einer folgenden Gelextraktion isoliert.

C.5.7 Radioaktive Markierung von DNA

C.5.7.1 Oligo-Markierung (Feinberg und Vogelstein, 1983)

Vor der radioaktiven Markierung wurden 50 ng doppelsträngige DNA in einem Volumen von 10 µl durch fünfminütiges Kochen denaturiert und direkt danach in einem Ethanol-Eisbad (-80°C) abgekühlt. Während das Reaktionsgefäß auf Eis gelagert wurde, wurden 2 µl 10x Hexanukleotidpuffer, 1 µl einer Stammlösung der drei nicht radioaktiv markierter Desoxyribonukleotidtriphosphate, 1 µl BSA Lösung (1 mg/ml), 5 µl (50 µCi) des radioaktiv markierten α-³²P-dNTP und 1 µl Klenowfragment der DNA-Polymerase I (2 U/µl) dazugegeben. Für die Markierung mit α-³²P-dCTP enthielt die Stammlösung der dNTPs jeweils 10 mM dTTP, dGTP und dATP. Der Ansatz wurde mindestens 2 h bis ∞ bei 37°C inkubiert.

Nicht eingebaute dNTPs wurden durch eine Säulenausschlußchromatographie (Sephadex G-50 in TE) von der markierten Sonde getrennt. Dafür wurde eine 1 ml Spritze mit etwas Aquariumsfilterwatte gestopft, mit gequollener, autoklavierter Sephadex G-50-Suspension gefüllt und 5 Minuten bei 3.500 U.p.m (TC 6) zentrifugiert. Das Volumen des Reaktionsansatzes wurde vor dem Auftragen auf die Säule mit TE auf 100 µl erhöht. Nach einer Zentrifugation (Wifug Doctor, 7.000 U.p.m, 5 Minuten bei Raumtemperatur) wurde das Eluat entweder in ein Eppendorf-Reaktionsgefäß überführt oder direkt in die Hybridisierungslösung pipettiert. Von 1 µl des Eluats oder 100 µl der Hybridisierungslösung wurde die Zerfallsrate im Szintillationszähler (Cerenkov-Zählung; Haberer, 1966) bestimmt. Es wurde nur mit solchen Sonden hybridisiert, die mindestens 2×10^8 Zerfälle pro Minute und µg DNA aufwiesen.

C.5.7.2 Kinasierung von DNA-Enden

Zur radioaktiven Markierung von doppelsträngigen DNA-Fragmenten für Gelretardierungs-Experimente (C.2.4) wurden ca. 500 fmol doppelsträngige DNA mit 10 U der T4-5'-Polynukleotidkinase (NEB) mit bis zu 40 µCi γ - ^{32}P -ATP in vom Hersteller gelieferten Puffer 30 min bis 1 h bei 37°C phosphoryliert. Anschließend wurde die DNA von den freien Nukleotiden durch Gelfiltration gereinigt (C.5.7.1). Die DNA wurde in ca. 100 µl TE eluiert (Endvolumen) und direkt verwendet oder bis zu 14 Tage bei -20°C gelagert.

C.5.8 Sequenzierung von DNA mit der Kettenabbruchstechnik (Sanger, 1977) mit fluoreszenzmarkierten Didesoxyribonukleotiden

Plasmid-DNA, die mittels Alkalischer Lyse oder Kochlyse präpariert wurde (C.3.2), mußte vor der Sequenzierung weiter aufgereinigt und denaturiert werden. Zum gesamten Ansatz wurde dafür 1 µl RNase A zugesetzt und 15 min bei 37°C inkubiert. Im Anschluß wurde 1 µl 10 N NaOH zugegeben und exakt 15 min bei RT inkubiert. Um die nach der RNase-A-Behandlung vorhandenen Ribonukleotide abzutrennen, wurden die Proben mit Ammoniumacetat und Ethanol gefällt und einige Stunden bei -20°C gelagert. Nach einer mindestens 20 minütigen Zentrifugation bei 13.000 U.p.m. und 4°C wurde die präzipitierte DNA mit 70% igem Ethanol gewaschen und bei Raumtemperatur getrocknet. Bei Plasmid-DNA, die mit Anionenaustauschersäulen aufgereinigt worden war, konnten diese Schritte ausgelassen werden.

Die Sequenz von DNA-Fragmenten wurde durch die Kettenabbruchstechnik (Sanger, 1977) bestimmt. Die hier angewandten, mit Fluoreszenzfarbstoffen markierten ddNTPs aus dem ABI Prism® BigDye™ Terminator Cycle Sequencing Ready Reaction-Kit wurden in einem automatisierten Applied Biosystems DNA Sequenzierer Typ 377 detektiert.

Gewöhnlich wurde, je nach Molekulargewicht der DNA, 1–3 µg λ- oder Plasmid-DNA bzw. 150 ng bis 200 ng PCR-Produkte als Matrize und 50 ng spezifisches Oligodesoxyribonukleo-

tid als Primer eingesetzt. Im Falle eines sehr hohen GC-Gehaltes wurde der Reaktion 5% DMSO zugesetzt, um Sekundärstrukturen der DNA aufzulösen. Die Reaktionen wurde auf ein Volumen von 12 µl aufgefüllt. Nach Zugabe von 4 µl des Big Dye Terminator Ready Reaction-Mixes wurde die Reaktion mit Mineralöl überschichtet und die Sequenzreaktion in einem DNA-Thermocycler 480 mit folgendem Programm gestartet:

Denaturierung:	5 min	96°C	
25 Zyklen	30 sec	96°C	Denaturieren
	15 sec	50°C-60°C	Anlagerung des Primers
	60°C	4 min	Elongation

Abkühlen auf 4°C

Danach wurden die Proben gemäß der Herstellerangaben aufgereinigt und auf dem Sequenziergerät getrennt. Die PCR-Reaktionen, das Aufreinigen und Auftragen der Proben auf ein denaturierendes Polyacrylamidgel wurden freundlicherweise von Birgit Schmitz, Gudrun Schell oder Sabrina Auerochs übernommen.

C.5.9 RT-PCR

Die RT-PCR Analyse von 100–300 ng DNaseI-behandelter Gesamt-RNA wurde in Ein-Gefäß-Reaktionen unter Verwendung von RT-PCR Kits verschiedener Hersteller (Promega, Qiagen, Life Technologies) durchgeführt. Primer wurden, wenn möglich, so ausgewählt, daß sie in verschiedenen Exons lagen. Optimale Bedingungen für die Hybridisierung der spezifischen Oligodesoxyribonukleotide an die RNA wurde für jedes Primerpaar neu bestimmt, die restlichen Bedingungen entsprachen den Herstellerangaben.

C.6 Elektrophoretische Trennung von Nukleinsäuren (McDonell et al., 1977)

C.6.1 Agarosegele

C.6.1.1 Elektrophoretische Trennung von DNA in Agarosegelen

Mit Restriktionsendonukleasen gespaltene genomische DNA (C.3.1.1) wurden 14 bis 18 Stunden bei 1,5–2 V pro cm in 0,8–1,5%igen Agarosegelen (20 x 20 x 0,75 cm) ihrer Größe nach getrennt. Als Laufpuffer wurde 1x TBE verwendet. Nach der Elektrophorese wurden die Gele in einer Ethidiumbromidlösung (2µg/ml) angefärbt und anschließend mit Wasser abgespült. Die entstandenen Ethidiumbromid-Nukleinsäure-Komplexe wurden durch UV-Licht sichtbar gemacht und mittels einer Videokamera photographiert. Zur Auftrennung von PCR-Reaktionen oder Plasmid DNA wurden Mini- und Midigele (mit 15 V oder 8 V pro cm) verwendet.

C.6.1.2 Qualitätskontrolle von RNA in Agarosegelen

Zur Kontrolle der Qualität präparierter RNA wurden 0,5–1 µg RNA mit dem doppelten Volumen an RNA-Probenpuffer versetzt und 5 min bei 65°C erhitzt. Nach Abkühlung auf Eis wurden 2 µl RNA-Ladepuffer dazugegeben und die Proben auf ein 1%iges Agaroseminigel in 0,5x TBE aufgetragen. Die Gelelektrophorese erfolgte bei 35 mA bis der Indikator Bromphenolblau etwa 2/3 der Laufstrecke erreicht hatte. Anschließend wurde das Gel mit Ethidiumbromid gefärbt und photographiert. Intakte RNA war an den Banden der 28 S und 18S rRNA zu erkennen, wobei die 28S-rRNA Bande etwa doppelt so intensiv sein mußte wie die 18S rRNA. mRNA war nur als leichter Schimmer im Hintergrund zu erkennen.

C.6.2 Elektrophoretische Trennung von DNA in Polyacrylamidgelen

Für die Auftrennung von DNA-Fragmenten, die kleiner als 100 bp waren, und für Protein/DNA-Komplexe wurde eine Elektrophorese in Polyacrylamidgelen (PAGE) vorgenommen. Polyacrylamidgele (PAA-Gele) weisen eine wesentlich höhere Ladekapazität als Agarosegele auf, was bei der Aufreinigung doppelsträngiger Oligodesoxyribonukleotide ausgenutzt wurde. Acrylamidmonomere kopolymerisierten mit dem Vernetzer N,N'-Methylenbisacrylamid, nachdem diese Kettenreaktion durch Ammoniumpersulfat und TEMED gestartet wurde.

Lösung für die Herstellung von nativen Polyacrylamidgelen:

Acrylamid/Bisacrylamid-Lösung (19:1), 40%	3,5 - 5 ml (12,5 ml = 10%)
TBE 10x	2,5 ml
Wasser	auf 50 ml
APS (10% (w/v))	0,32 ml
TEMED	0,05 ml

Die so vorbereitete Acrylamidlösung wurde nach Zugabe von TEMED zwischen zwei sorgfältig gereinigte Glasplatten gegossen, die an drei Seiten durch Abstandshalter mit einer Dicke von 0,75 bis 2 mm voneinander getrennt waren. Ein geeigneter Kamm wurde dann an der vierten Seite zwischen die Glasplatten geschoben und erst nach der vollständigen Polymerisierung des Gels entfernt. Die Gele wurden entweder am gleichen Tag (frühestens 3 Stunden nach dem Gießen) verwendet oder bis zu 3 Tage im Kaltraum gelagert. Die Glasplatten wurden in die entsprechende Vorrichtung eingespannt, die Geltaschen mit Elektrophoresepuffer gespült und vorhandene Luftblasen zwischen den Glasplatten mittels einer Spritze entfernt. Vor dem Auftragen der Proben wurde das Gel mindestens 30 min in 0,5x TBE bei 200-300 V eingefahren. Nach der Gelelektrophorese wurden die Glasplatten entfernt und entweder die gewünschten Fragmente ausgeschnitten und eluiert, oder das Gel auf Whatman-Papier aufgezogen.

C.7 DNA-Transfer-Hybridisierung „Southern Blot“ (Southern, 1975; Chomczynski, 1992; Koetsier et al., 1993)

Nach einer Elektrophorese in Agarosegelen wurden DNA-Fragmente nach der Methode von Southern (1975), variiert durch Chomczynski (1992) und Koetsier et al. (1993) auf Nylonmembranen übertragen. Die Gele wurden im Anschluß an das Photographieren in 0,25 M HCl bis zum Farbumschlag des Bromphenolblau markers von blau nach gelb inkubiert und dabei depuriniert (max. 20 min). Nach der folgenden Neutralisierung mit 0,4 M NaOH wies der Bromphenolblau marker seine ursprüngliche Farbe auf. Zwei auf Gelgröße zugeschnittene Filterstücke (GB003) wurden auf einen Stapel Papierhandtücher gelegt. Ein in 0,4 M NaOH getränktes Filterstück und die mit 0,4 M NaOH angefeuchtete Nylonmembran wurden ebenfalls auf dem Papierstapel plaziert. Die Geltaschen wurden entfernt, bevor das Gel auf die Membran gelegt und mit zwei weiteren getränkten Filterstücken bedeckt wurde. Den Abschluß bildeten zwei seitlich längere Filterstücke, die die Verbindung zur Transferlösung (0,4 M NaOH) herstellten. Der Aufbau wurde mit Plastikfolie bedeckt, um ein Austrocknen zu verhindern. Nach dem Transfer (mindestens 2,5 Stunden) wurde die Membran kurz in 2x SSC neutralisiert und anschließend erhitzt (Membranen von Qiagen 2 Stunden bei 80°C, von Boehringer Mannheim 30 Minuten bei 120°C). Danach konnte die Membran prähybridisiert werden.

C.8 RNA-Transfer-Hybridisierung „Northern Blot“

Zunächst wurde RNA in einem 1%igen Agarosegel in 1x MOPS-Puffer aufgetrennt. Dazu wurde die Agarose in MOPS-Puffer aufgeköcht, und nach Abkühlen auf 55°C wurde 37%iges Formaldehyd hinzugefügt (6,6% Endkonzentration). Die RNA-Proben (bis zu 20 µg Gesamt-RNA bzw. 3 µg für mRNA angereicherte RNA-Präparationen) wurden mit dem doppelten Volumen an RNA-Probenpuffer versetzt, auf 65°C erwärmt und auf Eis abgekühlt. Die Proben wurden mit 2 µl RNA-Ladepuffer versetzt und nach einem zehnmütigen Vorlauf des Gels in 1x MOPS Puffer bei 4-5V/cm aufgetrennt. Die Gelelektrophorese wurde frühestens gestoppt, wenn der Indikator Bromphenolblau 10 cm in das Gel eingelaufen war. Ethidiumbromid verschlechtert den Transfer von RNA auf eine Nylonmembran, weshalb eine Dokumentation des zu transferierenden Geles nicht möglich war. Um trotzdem eine Aussage über die Größe der RNAs treffen zu können, wurden 2 µg eines RNA-Größenstandards (Promega) zusammen mit einer doppelt aufgetragenen Gesamt-RNA-Probe am Rand des Geles aufgetragen, nach der Gelelektrophorese abgetrennt und mit Ethidiumbromid gefärbt. Um einzelne Banden des Größenmarkers nach der intensiven Anfärbung des Formaldehydgels sichtbar zu machen mußte das Gel in Wasser entfärbt werden, bevor es unter UV-Licht photographiert wurde. Der Rest des Gels wurde direkt nach der Gelelektrophorese zunächst gründlich mit Wasser gespült, um das Formaldehyd zu entfernen. Anschließend wurde es 20 min in 0,05 N NaOH inkubiert und danach mit reinem Wasser abgespült. Vor dem Transfer, der analog zum DNA-Transfer erfolgte (C.7), wurde das Gel 40 min in 20x SSC belassen. Als

Transferlösung wurde 20x SSC verwendet und immer $\bar{u}N$ auf die Membran übertragen. Die Membran wurde danach 5 min lang in 5x SSC inkubiert, dann 5 min in Whatman-Papier gehüllt bei RT getrocknet und abschließend 2 Stunden bei 80°C gebacken. Die Membran wurde bis zur Hybridisierung trocken bei RT gelagert.

C.9 Hybridisierung radioaktiv markierter DNA-Sonden an membran-gebundene Nukleinsäuren

C.9.1 Hybridisierung der membrangebundenen DNA-Fragmente

Zum Nachweis spezifischer Nukleinsäuren auf Nylonmembranen durch Hybridisierung mit einer radioaktiv markierten DNA-Sonde wurde die Membran zunächst 1–4 Stunden bei 68°C präinkubiert. Dazu wurde zunächst die Prähybridisierungslösung 10 Minuten auf 100°C erhitzt, kurz auf Eis gekühlt und auf die Membran gegeben. Vor der mindestens 14-stündigen Hybridisierung wurde die Prähybridisierungslösung entfernt. Die Hybridisierungslösung mit der radioaktiv markierten Sonde wurde ebenfalls aufgeköcht und vor der Hybridisierung bei Rotation in einem Ofen und kurz gekühlt.

Um unspezifisch gebundene Sonde zu entfernen, wurde die Hybridisierungsröhre einmal mit Waschlösung I ausgespült, nachdem die Hybridisierungslösung in ein 50 ml Blaukappröhrchen abgossen wurde. Sie konnte je nach spezifischer Aktivität wiederverwendet werden. Anschließend ließ man Waschlösung I zweimal 10 Minuten bei 68°C einwirken, Waschlösung II entsprechend der Stärke des zu erwartenden Signals ein- bis zweimal 20 Minuten. Die hybridisierten Membranen wurden in Plastikfolie eingeschweißt und in Filmkassetten mit zwei Verstärkerfolien auf Röntgenfilmen (Kodak X-OMAT AR5) autoradiographiert.

C.9.2 Entfernen radioaktiver Sonden von hybridisierten Membranen

Vor einer Hybridisierung mit einer weiteren Sonde wurde eine Membran 20 bis 30 Minuten in Waschlösung II gekocht. Um sicherzugehen, daß die Sonde vollständig entfernt worden war, wurde die Membran über Nacht autoradiographiert.

C.10 Herstellung einer subgenomischen Bibliothek

C.10.1 Größenselektion durch Zonenzentrifugation im Saccharosegradienten

Genomische DNA, die vollständig mit SacI fragmentiert worden ist, besteht aus einer Population von DNA-Molekülen zwischen einigen 100 Basenpaaren (bp) bis über 100 Kilobasenpaaren (kb) Länge. Um den Anteil der Größenklasse, in der das gesuchte Gen lag, vor der Herstellung einer subgenomischen Bibliothek anzureichern, wurden die DNA-Fragmente durch Zonensedimentation in einem Saccharosedichtegradienten fraktioniert. Ein

linearer 10% bis 40%iger Saccharosegradient wurde in SW-41 Zentrifugenröhrchen (Ultra-Clear, Beckman) mittels eines Gradientenmischers hergestellt. Die Vollständigkeit der endonukleolytischen Reaktion mit SacI wurde zunächst durch eine DNA-Transfer-Analyse nachgewiesen. Erst dann wurde das Enzym durch Erhitzen inaktiviert (68°C für 20 Minuten). Vor dem Auftragen des Restriktionsansatzes auf den Gradienten (200 µg als maximale Menge an genomischer DNA) wurden möglicherweise entstandene DNA-Aggregate durch fünfminütiges Erwärmen auf 65°C aufgelöst. Der beladene Gradient wurde 20,5 Stunden bei 27.000 U.p.m und 20°C (L7 55) zentrifugiert. Anschließend wurde ein Loch in den Gefäßboden eingeführt, durch das der Gradient austropfen konnte. Die Lösung wurde in einer Mikrotiterplatte aufgefangen (etwa 500 µl pro Vertiefung). Um die Fraktionen zu identifizieren, die das gewünschte 14 kbp Fragment enthielten, wurde jeweils 1/16 des Volumens auf einem 0,8%igem TBE-Agarosegel elektrophoretisch getrennt und, eine Membran übertragen, die schließlich in einer DNA-Transfer-Analyse eingesetzt wurde (C.7). Die positiven Fraktionen wurden anschließend dreimal mindestens 4 Stunden bei 4°C gegen TE dialysiert, mit Ethanol präzipitiert (C.3.4) und in 30 µl 1x *shrimp-alkaline*-Phosphatase-Puffer (SAP) gelöst.

C.10.2 Präparation der zu klonierenden DNA-Fragmente

Um den Einbau von mehreren genomischen DNA-Fragmenten in ein Vektormolekül zu verhindern, wurden die DNA-Fragmente gemäß den Angaben des Herstellers dephosphoryliert und die Phosphatase durch Erhitzen inaktiviert (68°C, 20 Minuten). Die mit Natriumacetat und Ethanol präzipitierte DNA (C.3.4) wurde in 3,5 µl 10 mM Tris-HCl pH 8,5 resuspendiert.

C.10.3 Präparation des Vektors λ-DASH II (Stratagene)

Bei dem Vektor λ-DASH II handelt es sich um einen Substitutions-Vektor, bei dem auf zwei Arten auf rekombinierte Phagen selektioniert wird: zum einen durch die Größe der zu verpackenden λ-DNA (Größe der Fremd-DNA 9-23 kbp), zum anderen durch den Verlust des Spi⁺-Phänotyps (sensitive to P2 interference). Das Fragment zwischen den beiden Klonierungsstellen enthält die red und gam Gene, aufgrund derer λ-Phagen keine *E. coli*-Bakterien infizieren können, die lysogen für den Phagen P2 sind. Nach Ersetzen dieses Fragmentes gegen fremde DNA kann der rekombinierte Phage dann auf solchen *E. coli*-Stämmen wachsen.

Der Vektor λ-DASH II-DNA (5 µg) wurde gemäß den Empfehlungen des Herstellers 3 Stunden mit 15 U SacI und XhoI inkubiert. Durch den Einsatz von zwei in der multiplen Klonierungsstelle schneidenden Enzymen wird der Anteil an religierten Phagen verringert. Nach Hitzeinaktivierung der Enzyme wurde die λ-DNA durch Zugabe eines halben Volumen

8 M Ammoniumacetat und der 2,5-fachen Menge an absolutem Ethanol (Raumtemperatur) präzipitiert (C.3.4) und in 5 μ l TE pH 7,5 resuspendiert.

C.10.4 Ligation und Verpackung (Stratagene)

Ein Teil der λ -DASH II-DNA (1 μ g) wurde mit 2,5 μ g DNA aus der 14 kbp Fraktion des Saccharosegradienten (C.10.1) in einem Gesamtvolumen von 5 μ l mit 0,5 μ l 10x Ligationspuffer (ATP enthaltend) und 2 Weiss Einheiten T4 Ligase 16 Stunden bei 4°C ligasiert.

Für die Verpackung der DNA in Phagenpartikel mit Stratagenes Gigapack II Plus Verpackungsextrakt wurde gemäß den Angaben die Hälfte des Ligationsansatzes eingesetzt. Der Titer der subgenomischen Bibliothek sollte mindestens 1×10^6 pfu/ml, nach der Amplifikation 1×10^9 pfu/ml betragen (*plaque forming units*). War dies der Fall, wurde mit der Durchmusterung der genomischen Bank begonnen. ($6,7 \times 10^6$ vor und $3,8 \times 10^9$ nach der Amplifikation)

C.10.5 Analyse der subgenomischen Bank – Plaquehybridisierung

Zum Durchsuchen der Genbibliothek (*screening*) wurde das Protokoll von Stratagene verwendet. Plaques wurden dabei durch Auflegen von ungeladener Nylonmembran auf diese übertragen und schafften dort ein präzises Abbild des Plaquemusters auf der Platte. Phagenproteine wurden alkalisch denaturiert und Phagen-DNA nach einer Neutralisierung durch zweistündiges Erhitzen der Membranen auf 80°C an der Filteroberfläche gebunden. Anschließend wurden die Filter mit einer Lösung inkubiert, die eine radioaktiv markierte Sonde enthielt. Signale konnten nach einer Autoradiographie aufgrund ihrer Position auf dem Film bzw. der Platte einem Plaque zugeordnet werden. Aus diesem Plaque wurden Phagen isoliert, deren Plaques erneut mit einer Plaquehybridisierung analysiert wurden. In dreimaligen Isolierungen von einzelnen Plaques (C.10.6.2) wurde die Reinheit der Phagenpopulation mittels Plaquehybridisierung überprüft. Dabei wurde kontrolliert, ob auch jeder Plaque dieser Population nach der Autoradiographie zu einem Signal auf dem Röntgenfilm führte.

Die einzige Abweichung zum Protokoll von Stratagene bestand in der Verwendung der auch bei DNA-Transfer-Analysen eingesetzten Prähybridisierungs- und Hybridisierungslösung.

C.10.6 Arbeiten mit λ -DASH II

C.10.6.1 Präparation der Wirtsbakterien XL1 blue MRA (P2) (Stratagene)

Eine Einzelkolonie XL1 blue MRA (P2) wuchs in 2,5 ml LB Medium mit 0,2% Maltose und 10 mM $MgSO_4$ 12 bis 14 Stunden im Schüttelinkubator bei 30°C. 200 μ l dieser Übernachtskultur wurden in eine 25 ml Tageskultur überimpft, die zu einer optischen Dichte

OD₆₀₀ von maximal 0,9 herangezüchtet wurde. Da Phagen sich auch an tote Bakterien anheften, war es wichtig, eine Kultur in der exponentiellen Wachstumsphase einzusetzen. Die Zellen wurden durch eine zehninütige Zentrifugation bei 2.500 U.p.m. (TC 6) sedimentiert und mit sterilem 10 mM MgSO₄ auf eine OD₆₀₀ von 0,5 eingestellt. Die Zellen konnten maximal 2 bis 4 Tage verwendet werden, wenn sie bei 4°C gelagert worden waren.

C.10.6.2 Isolierung einzelner Plaques

Aus einem mit einer Pasteurpipette aus der Agaroseplatte ausgestochenen Plaque wurden Phagen 16 Stunden in 0,5 ml SM-Puffer mit 0,3% Chloroform eluiert. Vor einem erneuten Ausplattieren wurde der Titer nach dem Protokoll von Stratagene bestimmt. Dazu wurden 600 µl *E. coli* XL1 blue MRA (P2) mit einer Verdünnungsreihe der Phagensuspension versetzt. Die Verdünnungsschritte waren 1:10, 1:100 und 1:1.000 in maximal 300 µl SM-Puffer. Die Bakterien wurden zur Erstinfektion 15 Minuten lang bei 37°C inkubiert. Vor dem Ausplattieren auf NZY-Agarplatten wurden 7,5 ml Topagarose (48°C) unter leichtem Schütteln zu den infizierten Bakterien gegeben. Die Platten wurden 16 Stunden bei 37°C inkubiert. Am nächsten Tag konnte der Titer der Ausgangslösung durch Auszählen der Plaques bei den verschiedenen Verdünnungsschritten bestimmt werden.

C.10.6.3 Präparation von λ-DNA

Für die Präparation von λ-DNA wurden 35 NZY-Agarose Platten (150 mm Durchmesser) eingesetzt. Mit 4×10^5 infektiösen Einheiten λ-DASH II Phagen infizierte XL1 blue MR (P2) Bakterien wurden mit 8,5 ml NZY-Top-Agarose ausplattiert. Nach einer 16-stündigen Inkubation bei 37°C war der Bakterienrasen vollständig lysiert. Die Platten wurden hiernach mit 12,5 ml λ-Verdünnungspuffer überschichtet, mit dem die Phagen 16 Stunden bei 4°C auf einem Schüttelgerät eluiert wurden. Das Eluat wurde mit der Waschlösung (5 ml λ-Verdünnungspuffer pro Platte) vereinigt. Dann wurde diese Suspension mit Chloroform (5%) versetzt. Die Lösung wurde 20 Minuten auf einem Rotator bewegt und schließlich wurden die Bakterientrümmere durch eine 15-minütige Zentrifugation bei 10.000 U.p.m. (Sorvall RC5B) und 4°C abgetrennt. Die Phagen wurden durch eine 16-stündige Zentrifugation bei 10.000 U.p.m. und 4°C sedimentiert. Nach Abgießen des Überstandes wurde das Phagensediment in 4 ml 4 M CsCl-Lösung resuspendiert und in Ultra-Clear-Zentrifugenröhrchen überführt. Anschließend wurde bis zum Gleichgewicht 22 Stunden lang in einem SW-60 Rotor bei 30.000 U.p.m. (Beckman L7 55) bei 20°C zentrifugiert. Mit einer 1 ml Spritze (Kanüle 20 G x 1 1/2") wurden 0,6–0,7 ml der gut als Bande sichtbaren Phagen abgenommen und jeweils 0,4 ml davon auf den Gefäßboden eines neuen Ultra-Clear-Zentrifugenröhrchens plaziert. Zur Entfernung von Bakterien-DNA und RNA folgte nun ein von oben nach unten dichter werdender Blockgradient: Über die 0,4 ml Phagen in Schwebbedichte-CsCl Lösung gab man zunächst 0,4 ml gesättigte 7,2 M CsCl Lösung, die wiederum mit 2,4 ml 5 M CsCl überschichtet wurde. Den Abschluß bildeten 0,8 ml 3 M CsCl. Der Gradient wurde 1 Stunde

bei 30.000 U.p.m. (Beckman L7 55) und 20°C ohne Bremse in einem SW-60 Rotor zentrifugiert. Die Phagen (0,42 ml) wurden wie oben beschrieben aus dem Gradienten entnommen und 1:1 mit λ -Verdünnungspuffer versetzt. Um Proteine abzutrennen, wurde die Richtung des Gradienten umgekehrt, weshalb 0,8 ml 5 M CsCl auf dem Gefäßboden plaziert und mit 2,4 ml 3 M CsCl Lösung überschichtet wurden. Diesmal bildete die Phagensuspension den Abschluß des Gradienten, der wiederum bei 30.000 U.p.m 1 Stunde bei Raumtemperatur zentrifugiert wurde. Die isolierten Phagen wurden schließlich dreimal mindestens 4 Stunden gegen λ -Verdünnungspuffer bei 4°C dialysiert. Dann wurde die DNA über eine Phenol/Chloroform Behandlung (C.3.1.3) gereinigt, wobei die Lösungen 30 Minuten lang auf dem Rotator vermischt wurden.

C.11 Klonierungstechniken

C.11.1 Ligation (Dugaiczky et al., 1975)

C.11.1.1 Präparation und Ligation der Vektor- und Fremd-DNA

Die Vektor-DNA wurde mit dem ausgewählten Enzym gemäß den Angaben der Hersteller gespalten, mit alkalischer Phosphatase dephosphoryliert und elektrophoretisch getrennt. Die Bande wurde aus dem Gel ausgeschnitten und die DNA mit einem Gelextraktionskit (Qiagen) aus dem Gel extrahiert. Die zu klonierende DNA wurde wie der Vektor mit Restriktionsendonukleasen behandelt, aber soweit nicht anders angegeben, nicht dephosphoryliert. Nach einer Gelelektrophorese wurde die gewünschte Bande anhand eines DNA-Größenstandards identifiziert und wie oben gereinigt.

Zur Ligation wurden Vektor (mindestens 20 ng) und zu klonierende DNA in einem Gesamtvolumen von 8 μ l meist in einem molaren Verhältnis von 1:3 vermischt. Es wurde 1 μ l 10x Ligationpuffer (ATP enthaltend) und 1 μ l T4-DNA Ligase (5 Weiss Einheiten) beigemischt und der Ansatz gemäß den Herstellerangaben oder bei 4°C ∞ inkubiert.

C.11.1.2 Ligation von PCR-Produkten in den pGEM-T-Vektor

Die mit der Taq-DNA-Polymerase synthetisierten PCR-Produkte wurden unter Verwendung des pGEM-T-Vektor Kits der Firma Promega den Herstellerangaben folgend kloniert. Die Taq-Polymerase besitzt eine terminale Transferaseaktivität, mit der sie jedes PCR-Produkt am 3'-Ende mit einem zusätzlichen Desoxyadenin versieht. Dies ermöglichte die Ligation in einen T-Überhangs-Vektor wie dem pGEM-T. In rekombinanten Molekülen ist das offene Leseraster der β -Galactosidase unterbrochen, so daß Blau-Weiß-Selektion (C.11.4) nach der Transformation möglich war.

C.11.2 Herstellung elektrokompetenter Bakterien

Für die Herstellung elektrokompetenter Bakterien wurden 5 ml SOB-Medium mit einer Einzelkolonie des gewünschten Bakterienstammes inokuliert und 12 bis 14 Stunden bei 37°C im Schüttelinkubator herangezogen. Ein Teil dieser Übernacht-Kultur (0,5 ml) wurde in 500 ml SOB überimpft und bis zu einer OD₆₀₀ von 0,7 bis 0,8 unter kräftigem Schütteln bei 37°C wachsen gelassen (2 bis 5 Stunden). Anschließend wurde die Bakteriensuspension 10 bis 20 Minuten auf Eis gekühlt. Die Zellen wurden durch eine zehnminütige Zentrifugation bei 4.500 U.p.m. (JA-10 Rotor) und 4°C pelletiert. Der Überstand wurde so schnell wie möglich mit einer Saugpumpe entfernt. Das Zellsediment wurde in 50 ml WB-Lösung durch kräftiges, kreisförmiges Schwenken der Zentrifugenbecher resuspendiert, bevor das Volumen auf 500 ml mit WB aufgefüllt wurde. Dieser Zentrifugations- und Resuspensionsschritt zum Entsalzen der Bakteriensuspension wurde einmal wiederholt. Nach der letzten Zentrifugation wurden die Zellen in insgesamt 2 ml WB aufgenommen (entspricht 3×10^{10} Zellen/ml). Fraktionen von 40 µl wurden in einem Ethanol-Trockeneisbad tiefgefroren und konnten bei -80°C bis zu sechs Monate gelagert werden. Die Transformationseffizienz wurde mit 50 pg superspiralierter Plasmid-DNA getestet und in Kolonien pro µg DNA bestimmt.

C.11.3 Transformation von Bakterien

C.11.3.1 Elektroporation

Ein Ligationsansatz wurde vor der Elektroporation mit Ammoniumacetat und Ethanol gefällt, wobei 20 µg tRNA (Transfer-RNA) als Träger-Nukleinsäure hinzugefügt wurden. Der Ansatz wurde vor der halbstündigen Zentrifugation 10 Minuten bei -80°C gelagert. Nach dem Waschen mit 70%igem Ethanol wurde die DNA für die Elektroporation in 10 µl deionisiertem Wasser oder einem salzarmen Puffer gelöst. Pro Ligationsansatz wurden 40 µl einer elektrokompetenten Bakteriensuspension (-80°C) bei Raumtemperatur aufgetaut, mit 1 bis 3 µl der Ligationsreaktion gründlich vermischt und 1 Minute auf Eis gelagert. Die DNA/Zellsuspension wurde in auf 4°C gekühlte, sterile Elektroporationsküvetten überführt und durch Schütteln luftblasenfrei auf den Küvettenboden gebracht. Die außen trockenen Küvetten wurden in den -20°C kalten Halter überführt. Die Bakterien wurden bei 25 µF, 200 Ω, 1,5 bis 1,8 kV und einer Pulsdauer von etwa 4,5 bis 5 Millisekunden elektroporiert. Die Zellsuspension wurde nun so rasch wie möglich in 1 ml SOC Medium aufgenommen und maximal 1 Stunde bei 37°C ohne Antibiotikum in einem Inkubator geschüttelt. Anschließend wurden 20 µl, 200 µl und der Rest der Bakterienkultur auf Selektionsmedium ausplattiert.

C.11.3.2 Transformation durch Hitzeschock (Hanahan, 1983)

Alternativ zum oben genannten Protokoll wurden kompetente Bakterien bei der Firma Stratagene bezogen. Für die Transformation der „subcloning grade competent XL1 blue

MRF““ wurde etwa die Hälfte eines Ligationsansatzes gemäß den Herstellerangaben transformiert. Abweichend vom Protokoll wurden im Rahmen des genomischen Sequenzierens die transformierten Zellen vor dem Ausplattieren nicht bei 37°C wachsen gelassen, da durch das Wachstum der Bakterien eine Verschiebung der Häufigkeiten bestimmter Methylierungsmuster, die durch Bakterienklone repräsentiert wurden, hätte auftreten können.

C.11.4 Identifizierung rekombinierter Bakterienkolonien „Blau-Weiß-Selektion“

Bei Verwendung geeigneter Bakterienstämme und Vektoren konnten rekombinante Bakterienkolonien aufgrund ihrer Farbe von nicht rekombinanten unterschieden werden. Die Bakterien mußten dafür das lacZ-Gen ohne das α -Peptid enthalten, während das Plasmid ein offenes Leseraster für das α -Peptid der β -Galaktosidase beinhaltete. So waren die Bakterien durch die sogenannte α -Komplementation in der Lage, eine funktionsfähige β -Galaktosidase zu bilden. Wurde das offene Leseraster auf dem Plasmid durch die Aufnahme eines DNA-Fragmentes bei der Klonierung unterbrochen, konnte keine β -Galaktosidase gebildet werden. Die Agarplatten enthielten hierfür 0,5 mM IPTG (zur Induktion der β -Galaktosidase) und 40 μ g/ml X-Gal. X-Gal wurde durch das Enzym zu einem blauen Reaktionsprodukt umgesetzt. Da rekombinierte Bakterien X-Gal nicht umsetzen konnten, blieben die Kolonien weiß.

C.11.5 Identifizierung rekombinanter Bakterienkolonien durch Kolonie-Hybridisierung

Bakterienkolonien wurden auf ungeladene Nylonmembranen übertragen, indem die Membran auf die bewachsene Platten gelegt wurde, wobei die Orientierung der Membran gekennzeichnet wurde. Nach Entfernung der Membran wurden die Platten mindestens 3 Stunden im Inkubator belassen. Zur Lyse der Bakterien wurde die Membranen mit den Kolonien auf der Oberseite 5 min auf mit 2x SSC und 5% SDS getränkte Blotting-Papiere gelegt. Anschließend wurden die Filter in einer Mikrowelle bei etwa 750 W für 2 min erhitzt und danach auf RT abgekühlt. Nach dem 30-minütigen Prähybridisieren der Membranen in 20 ml 1% Sarcosyl und 5x SSC bei 65°C, wurde die Hybridisierungslösung mit 100 μ g Heringspermien-DNA und der entsprechende Sonde versetzt und 10 min aufgekocht. Nach einer Hybridisierung μ N wurden die Filter nach Entfernung der Hybridisierungslösung zunächst 15 min bei 65°C in 2x SSC und 0,5% Sarcosyl unter Rotation gewaschen. Anschließend wurden die Filter in eine Schale mit 0,5x SSC und 0,5% Sarcosyl überführt und dort 15 bis 30 min bei 65°C inkubiert. Nach der Exposition der Filter auf Röntgenfilmen konnten positive Bakterienkolonien anhand ihrer Position identifiziert werden.

C.12 Sequenzvergleiche in Datenbanken mit dem BLAST Programm

Das Programm BLAST (Altschul et al., 1997) findet innerhalb einer ausgewählten Datenbank Bereiche, die zu Regionen der Suchsequenz ähnlich sind. Es arbeitet hierbei mit einem

statistischen Ansatz, bei dem der Ähnlichkeitswert (*score*) oberhalb der statistischen Erwartung für den Vergleich zweier nicht verwandter Sequenzbereiche liegen muß. Dies kann man am E-Wert erkennen. Dieser Wert gibt die Wahrscheinlichkeit an, mit der ein paarweiser Sequenzvergleich mit seinem *score* bei der gegebenen Größe der Datenbank und der gegebenen Länge der Sequenzhomologie (Anzahl der Basen bzw. Aminosäuren) per Zufall gefunden werden könnte. Ein E-Wert von 1 bedeutet also, daß dieses Paar von miteinander verglichenen Sequenzen auch per Zufall entstehen könnte. Hier wurde eine Sequenzhomologie als statistisch signifikant angesehen, wenn sie einen E-Wert von $< 8e^{-5}$ aufwies. Bei der Analyse von Proteindatenbanken mit dem Programm BLASTX kann die Sequenzähnlichkeit jedoch trotz eines großen E-Wertes Bedeutung haben, da es hier häufig auf die Position von bestimmten Aminosäuren ankommt. BLASTX übersetzt eine Nukleotidsequenz in allen sechs Leserastern und vergleicht die Aminosäuresequenz mit einer Proteindatenbank. TBLASTN übersetzt eine Nukleotiddatenbank in allen Leserastern, so daß man sie mit einer Proteinsequenz durchsuchen kann. Meistens wurde das Programm BLAST online am *National Center for Biotechnology Information* genutzt.

D Ergebnisse

D.1 Vorarbeiten

D.1.1 Der genomische Klon des CGGBP1-Gens des Menschen

Nach der biochemischen Reinigung und partiellen Sequenzierung des CGGBP1 war der EST-Klon 269133 (Zugriffsnummer AJ000258, siehe I.1) durch eine Datenbankrecherche anhand der abgeleiteten Nukleotidsequenz identifiziert worden (Deissler et al., 1997). Diese partielle cDNA wurde für die Klonierung des menschlichen CGGBP1-Gens verwendet. Dazu wurde eine subgenomische Bank in λ -Phagen aus HeLa-DNA angelegt. Der isolierte und anschließend vollständig sequenzierte λ -Klon war 19.839 nt lang (Zugriffsnummer AF094481) und enthielt im zentralen Bereich das Gen für CGGBP1 (Abb. 2). Das offene Leseraster (ORF) des CGGBP1-Gens war im Gegensatz zum 5'-UTR nicht durch Introns unterbrochen. Die genaue Exon/Intron-Struktur des CGGBP1-Gens des Menschen, der Maus und der Ratte ist inzwischen geklärt (D.4.3).

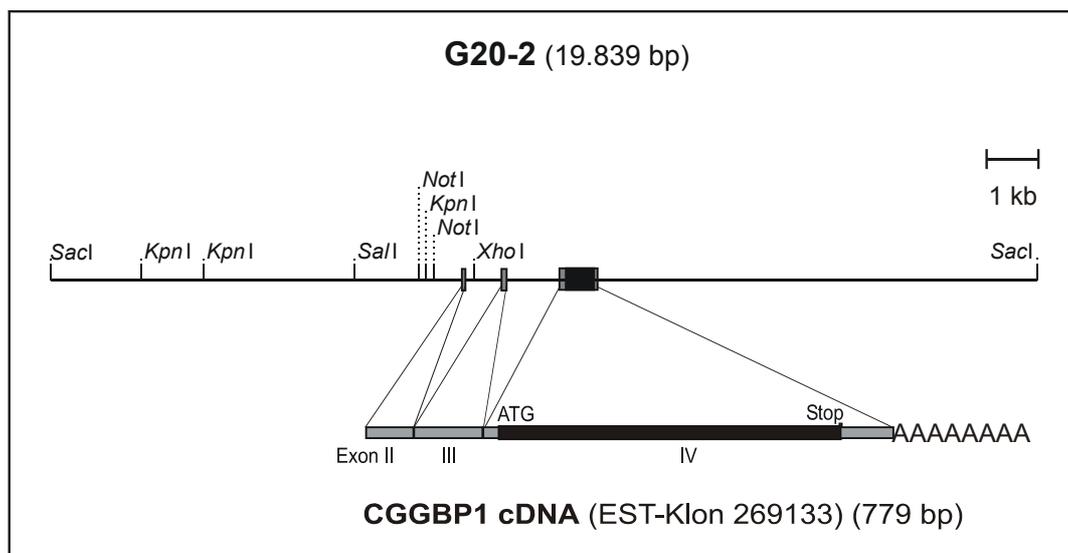


Abbildung 2 Übersichtskarte des genomischen Klons **g20-2** des **CGGBP1-Gens** aus **HeLa-DNA**.

Die Position der für die Klonierung verwendete Sonde, der EST-Klon 269133 wurde in den genomische Klon eingetragen. Der EST-Klon enthält neben dem gesamten ORF des CGGBP1-Gens zwei Exons der 5'-UTR.

D.1.2 Das CGGBP1-Gen in anderen Spezies

Genomische DNA von verschiedenen Spezies wurde unter stringenten Bedingungen mit der partiellen cDNA des CGGBP1-Gens des Menschen hybridisiert (Abb. 3). Es zeigte sich, daß das CGGBP1-Gen bei verschiedenen Säugetieren und beim Huhn hoch konserviert war. Erst nach dem Absenken der Hybridisierungstemperatur um 10°C wurde ein Signal in genomischer DNA aus Fischen detektiert.

Es blieb ungeklärt, ob es sich dabei um ein für das CGGBP1-Gen spezifisches Signal handelte. Bei Datenbankrecherchen bis März 2003 in den weitgehend sequenzierten Genomen von Fischen (http://www.sanger.ac.uk/Projects/D_rerio/; <http://www.fugu-sg.org/>) wurde keine CGGBP1-ähnliche Nukleotidsequenz gefunden. Möglicherweise ist die entsprechende chromosomale Region bisher nicht sequenziert.

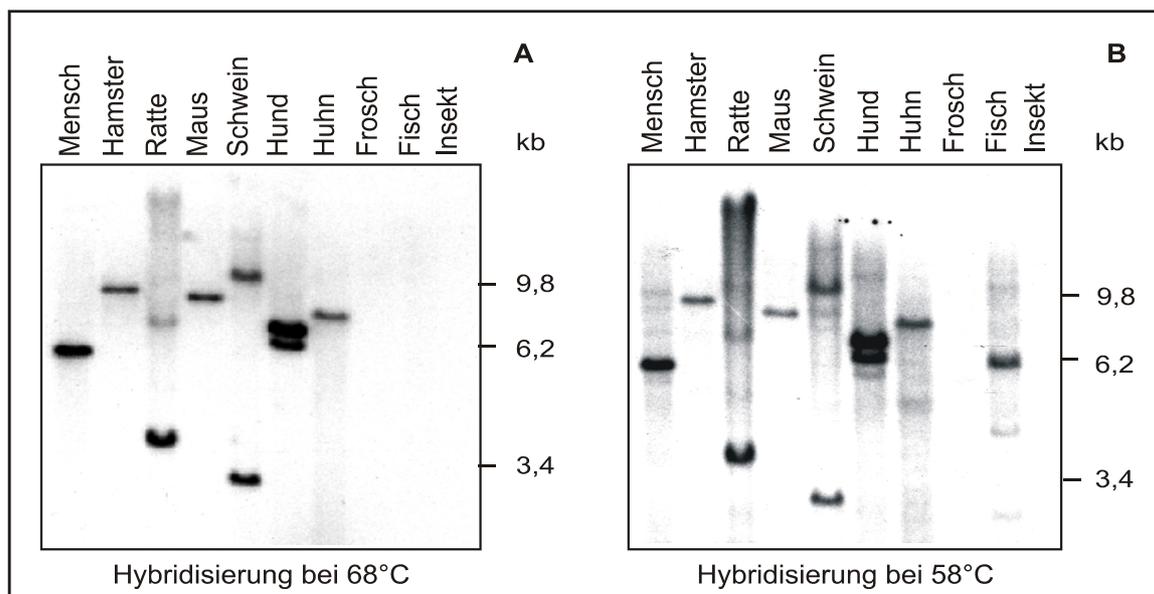


Abbildung 3 Hybridisierung von genomischer DNA verschiedener Spezies mit der Sonde für die cDNA des menschlichen CGGBP1-Gens.

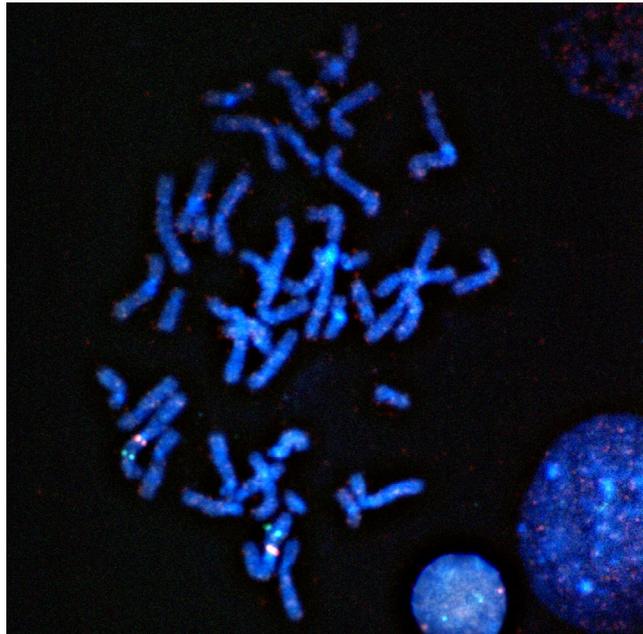
Jeweils 15 µg genomische DNA wurde mit HindIII gespalten, auf einem 0,8 %igen Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Die verwendeten DNAs wurde entweder aus Zelllinien (Mensch HeLa, Hamster BHK21, Hund MDCK, Fisch FHM) oder aus Geweben (Froschleber, Rattenniere, Fliegenembryonen) präpariert. Die DNA wurde bei 68°C oder 58°C mit der ³²P-markierten partiellen cDNA des CGGBP1-Gens des Menschen hybridisiert und stringent gewaschen.

D.1.3 Chromosomale Lokalisation des CGGBP1-Gens

Schließlich wurde das CGGBP1-Gen mittels der Fluoreszenz-*in-situ*-Hybridisierung auf dem menschlichen Chromosom 3p in der Nähe des Zentromers lokalisiert (Abb. 4). Als Sonden dienten hierbei der genomische Klon des CGGBP1-Gens und eine kommerziell erhältliche Sonde für die cytogenetische Bande 3p21. In der aktuellen Version der Sequenz des Genoms des Menschen (http://www.ensembl.org/Homo_sapiens/ Version 11.31.1 vom 27.2.2003; Hubbard et al., 2002) liegt das CGGBP1-Gen auf dem langen Arm von Chromosom 3.

Abbildung 4 Chromosomale Lokalisation des CGGBP1-Gens des Menschen auf dem kurzen Arm von Chromosom 3(cen).

Die Chromosomen aus einer Metaphase eines Lymphocyten (46, XX) wurden mit dem biotinylierten genomischen Klon des CGGBP1-Gens des Menschen hybridisiert und mit DAPI (blau) gegengefärbt. Biotin wurde durch ein Avidin-Cy3-Konjugat (rot) detektiert. Für die chromosomale Kartierung des CGGBP1-Gens wurde eine für Chromosom 3p21 spezifische Sonde digoxigeniert (Oncor) und mittels FITC-markierten Antikörpern (grün) nachgewiesen



Ein Großteil der in Kapitel 1 vorgestellten Ergebnisse hatte ich im Rahmen der Diplomarbeit (Naumann et al., 1998; Deissler et al., 1997; Müller-Hartmann et al., 2000) erarbeitet.

D.2 Klonierung des *cggbp1*-Gens des Mausstamms C57BL/6

Das CGGBP1-Gen ist bei Säugetieren hochkonserviert (D.1.2). Deshalb wäre es möglich, daß mit einem Geninaktivierungs-Experiment in der Maus (*knock out*) auch Erkenntnisse über die Funktion des CGGBP1-Gens des Menschen erbracht werden könnten. Mit der Klonierung des Gens der Maus wurde die Voraussetzung für ein solches Geninaktivierungs-Experiment geschaffen. Die Sequenz des genomischen Klons der Maus sollte außerdem mit der des Menschen und anderen, in Datenbanken zugänglichen Sequenzen weiterer Spezies verglichen werden, um bei hoher Sequenzähnlichkeit Hinweise auf biologisch wichtige Regionen zu

erhalten. Im folgenden wird der genomische Klon des Menschen als g20-2, der der Maus als m20 bezeichnet. Zur leichteren Unterscheidung vom menschlichen Gen CGGBP1 wird das Gen der Maus klein geschrieben (cggbp1).

D.2.1 Herstellung und Durchsuchen einer genomischen Bank

Vor der Erstellung der genomischen Bank mußte eine passende Restriktionsstrategie für die Klonierung des cggbp1-Gens gefunden werden. Dafür wurde genomische DNA des Mausstamms C57BL/6 mit verschiedenen Restriktionsendonukleasen gespalten und mittels einer Southern-Transfer-Hybridisierung analysiert (Abb. 5).

Aus früheren eigenen Studien mit heterologen Hybridisierungen war bekannt, daß die menschliche cDNA auch mit DNA aus Mäusen zu einem spezifischen Signal führte (Abb. 3). Für die Klonierung wurde der λ -Substitutionsvektor λ -DASH®-II gewählt, der Fragmente zwischen 9 bis max. 23 kb aufnehmen kann. Daher eignete sich das Enzym SacI, bei dem das cggbp1-Gen auf einem etwa 14 kb großen Fragment lag. In der subgenomischen Bank wurde die Größenklasse der 14 kb-Fragmente angereichert (C.10.1).

Insgesamt wurden 1×10^6 Plaques analysiert und positive Plaques durch drei Isolierungsschritte gereinigt (C.10.6.). Die Reinheit einer Phagenpopulation wurde durch Filterhybridisierung kontrolliert. Nur wenn jeder der auf dem Bakterienrasen sichtbaren Plaques nach der Hybridisierung mit der CGGBP1-cDNA ein Signal zeigte, wurde DNA aus den Phagen präpariert und mit verschiedenen Restriktionsendonukleasen gespalten. Da vier unabhängig voneinander isolierte Phagen das gleiche

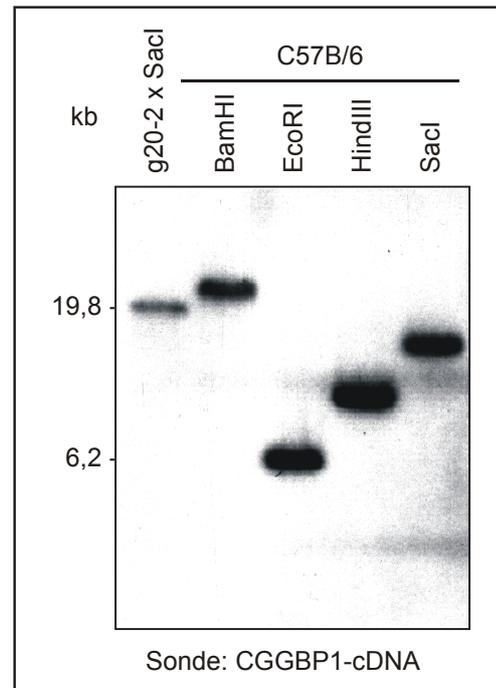


Abbildung 5 Restriktionsanalyse von C57BL/6 DNA

Jeweils 15 μ g genomische DNA des Mausstamms C57BL/6 wurden mit den angegebenen Enzymen gespalten. Die Fragmente wurden in einem 0,8 %igen TBE-Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Zum Größenvergleich wurde auch der genomische Klon g20-2 des Menschen (19.839 bp) aufgetragen. Als Hybridisierungssonde in der DNA-Transfer-Analyse diente die partielle cDNA des menschlichen CGGBP1-Gens.

Restriktionsmuster aufwiesen (Daten nicht gezeigt), konnten große Umlagerungen ausgeschlossen werden, so daß mit einem Klon (m20-1.1) weitergearbeitet wurde. Das *SacI*-Fragment des λ -Klons m20-1.1 wurde in den Vektor pBSII subkloniert. In Abhängigkeit von der Orientierung des *cggbp1*-Gens im Plasmid wurde die Subklone als m20-9 bzw. m20-10 bezeichnet (siehe Plasmidkarten I.4).

D.2.2 PCR-Klonierung der 5'-flankierenden Sequenz des *cggbp1*-Gens

Das erste Exon des *cggbp1*-Gens war nur unvollständig im genomischen Klon enthalten, und es fehlte 5'-flankierende Sequenz (D.2.4). Im Rahmen des Projektes zur Bestimmung der Genomsequenz der Maus war vor der eigentlichen Veröffentlichung im Jahr 2002 (Mouse Genome Sequencing Consortium, 2002) bereits im November 2000 eine Sequenz auf der Datenbank GenBank (Benson, 1998) zugänglich, die als vorläufig bezeichnet worden war (*contig* c078902648). Da die zu EST-Klonen des *cggbp1*-Gens ähnliche Sequenz das Ende dieses *contigs* darstellte, und Fehler bei der Zusammensetzung von Sequenzen häufig am Ende passieren (Mouse Genome Sequencing Consortium, 2002), mußte die Colinearität des veröffentlichten *contigs* mit dem hier klonierten Gen nachgewiesen werden. Eine PCR auf genomischer DNA des Mausstammes C57BL/6 zeigte, daß die publizierte Sequenz tatsächlich der 5'-flankierenden Sequenz entsprach, da ein spezifisches Produkt von 902 bp gebildet wurde (Abb. 6). Die anschließende Sequenzierung des PCR-Produkts bestätigte zudem die publizierte Sequenz.

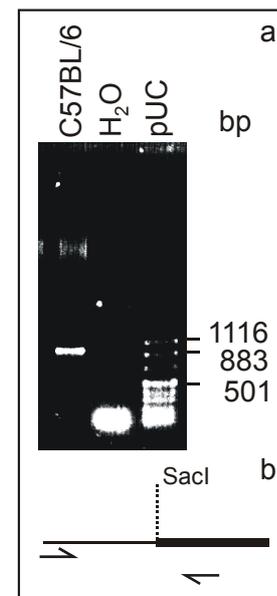


Abbildung 6 Kontrolle der 5'-flankierenden Sequenz

In einer PCR mit genomischer Maus-DNA wurde ein spezifisches Produkt gebildet (a). Die Lage der Primer ist in dem Schema eingetragen (b). Die publizierte Sequenz wird durch eine dünne Linie gekennzeichnet, die hier klonierte Sequenz mit einer dicken Linie.

D.2.3 Kontrolle des isolierten cggbp1-Gens aus C57BL/6-DNA

Um sicherzustellen, daß es sich bei dem identifizierten λ -Klon auch tatsächlich um das cggbp1-Gen handelte, wurde das gesamte SacI-Fragment von m20-1.1 als Sonde in einer DNA-Transfer-Hybridisierung eingesetzt (Abb. 7). Die Hybridisierung der menschlichen cDNA und PCR-amplifizierter Exons der Maus mit dem isolierten λ -Klon erbrachte spezifische Signale. Die Primer für die Amplifizierung der Exons aus genomischer Maus-DNA waren zuvor aus EST-Klonen der Maus abgeleitet worden.

Bei der DNA-Transfer-Analyse von genomischer DNA mit dem λ -Klon m20-1.1 wurde die gesamte Spur abgebildet, d.h. der Klon enthielt repetitive Sequenzen. Die klonierten DNA-Fragmente und genomische DNA wurden mit den gleichen Restriktionsendonukleasen gespalten. Die resultierenden Fragmente wurden nebeneinander auf ein Gel aufgetragen und mit einer DNA-Transfer-Hybridisierung analysiert, um Umlagerungen der genomischen Sequenz bei den vorgenommenen Klonierungsschritten auszuschließen. Nach der Hybridisierung mit der partiellen menschlichen cDNA stimmte das Bandenmuster der Klone mit dem der genomischen DNA überein (Abb. 8). Somit waren Umlagerungen, zumindest im Bereich des ORF des cggbp1-Gens, unwahrscheinlich. Nach diesen Kontrollen des isolierten genomischen Klons m20 konnte mit der Sequenzierung begonnen werden (D.2.4).

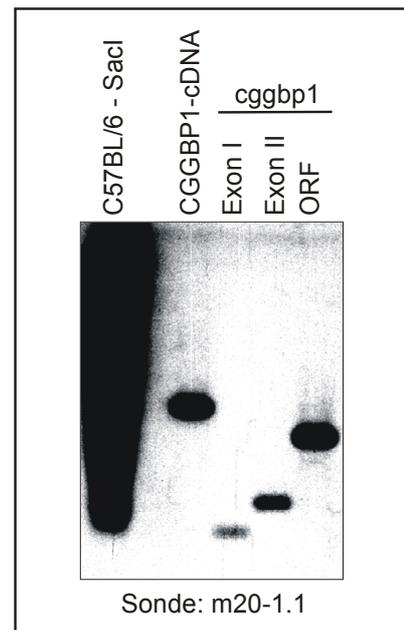


Abbildung 7 Rückhybridisierung
DNA-Transfer-Analyse von je etwa 200 pg der partiellen cDNA des CGGBP1-Gens und PCR-amplifizierten Exons des cggbp1-Gens. Als Hybridisierungssonde diente der λ -Klon des cggbp1-Gens m20-1.1.

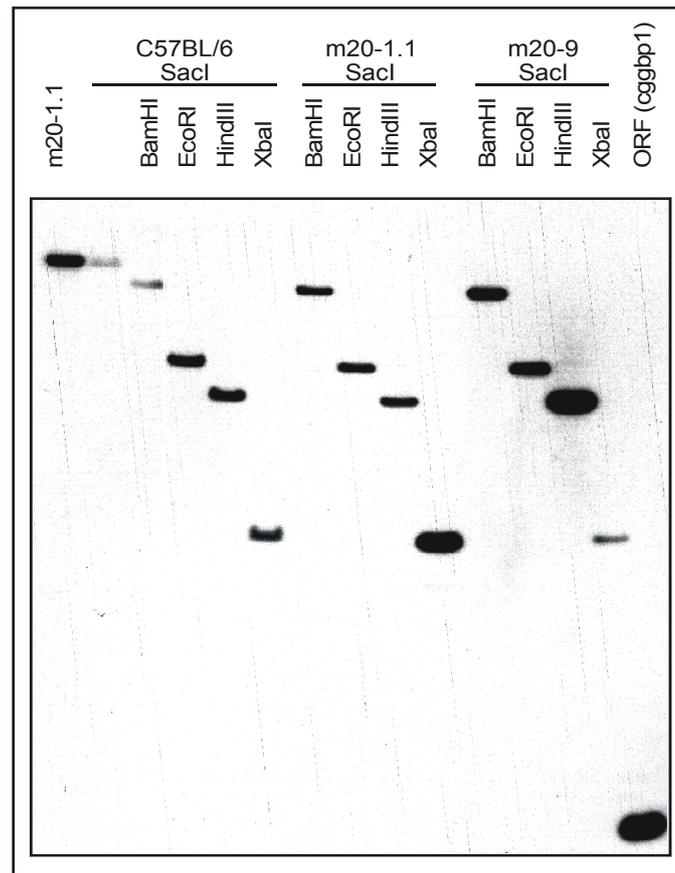


Abbildung 8 DNA-Transfer-Analyse von genomischer Maus-DNA und den klonierten Fragmenten m20-1.1 und m20-9.

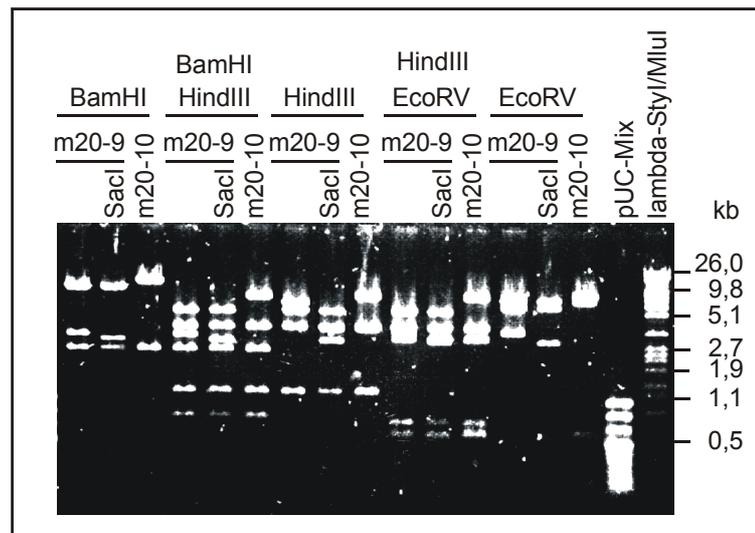
Jeweils 300 pg des λ -Klons m20-1.1 und des Plasmids m20-9 bzw. 10 μ g genomische Maus-DNA wurden mit den angegebenen Restriktionsendonukleasen gespalten, die Fragmente wurden in einem 0,8%igen TBE-Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Als Hybridisierungssonde im DNA-Transfer-Experiment diente die menschliche cDNA des CGGBP1-Gens.

D.2.4 Restriktionskartierung des genomischen Klons

Die Klone m20-9 und m20-10 wurden mit verschiedenen Restriktionsendonukleasen gespalten (Abb. 9). Es erleichterte die Restriktionskartierung, daß die genomischen Fragmente in unterschiedlicher Orientierung in das Plasmid pBSII kloniert worden waren (siehe Abb.42; I.4). Bis zur vollständigen Sequenzierung des genomischen Klons beruhte die Strategie zur Herstellung des Geninaktivierungs-Vektors auf der so erstellten Restriktionskarte (D.5).

Abbildung 9
Restriktionskartierung des
genomischen Klons des cggbp1-
Gens.

Jeweils 750 ng der Plasmide m20-9 und m20-10 wurden mit den angegebenen Restriktionsendonukleasen gespalten. Die genaue Größe der Fragmente, die am 5'- oder am 3'-Ende des Klons lagen, wurde durch die zusätzliche Spaltung des m20-9-Klons mit SacI ermittelt.



D.2.5 Sequenzierung des isolierten cggbp1-Gens

Die Sequenz des Klons m20-9 des cggbp1-Gens wurde mit der Kettenabbruchstechnik (Sanger et al., 1977) nach dem Prinzip des *primer walking* bestimmt. Dabei wurde je nach Qualität der im vorhergehenden Schritt bestimmten Sequenz ein neuer Primer im Abstand von 400 bis 600 Nukleotiden ausgewählt (I.1). Die ersten Oligodesoxyribonukleotide waren die zum Plasmid pBSII komplementären T3 und T7 Primer. Außerdem wurden Oligodesoxyribonukleotide auf der Basis von EST-Klonen entworfen. Beide Stränge wurden komplett sequenziert, womit Fehler bei der Sequenzierung weitgehend ausgeschlossen wurden. Die Sequenzierung bestätigte die Restriktionskartierung des genomischen Klons und die Anwesenheit von repetitiven Elementen (Abb. 7 und 10). Die mit Stern versehenen EcoRI-/BamHI Fragmente (Abb. 10) wurden in das Plasmid pBSII subkloniert, um die repetitive Region stromabwärts des cggbp1-Gens vollständig sequenzieren zu können. Im Plasmid m20-9 lagerten Primer aus dieser Region mehrmals an verschiedenen Stellen an, so daß die daraus resultierenden Sequenzreaktionen nicht lesbar waren.

Die gesamte Sequenz von m20-9 wurde zusammen mit der der PCR-klonierten 5'-flankierenden Sequenz (14.530 bp; siehe I.3) bei der Datenbank GenBank (Benson et al., 1998) unter der Zugriffsnummer AY189895 veröffentlicht.

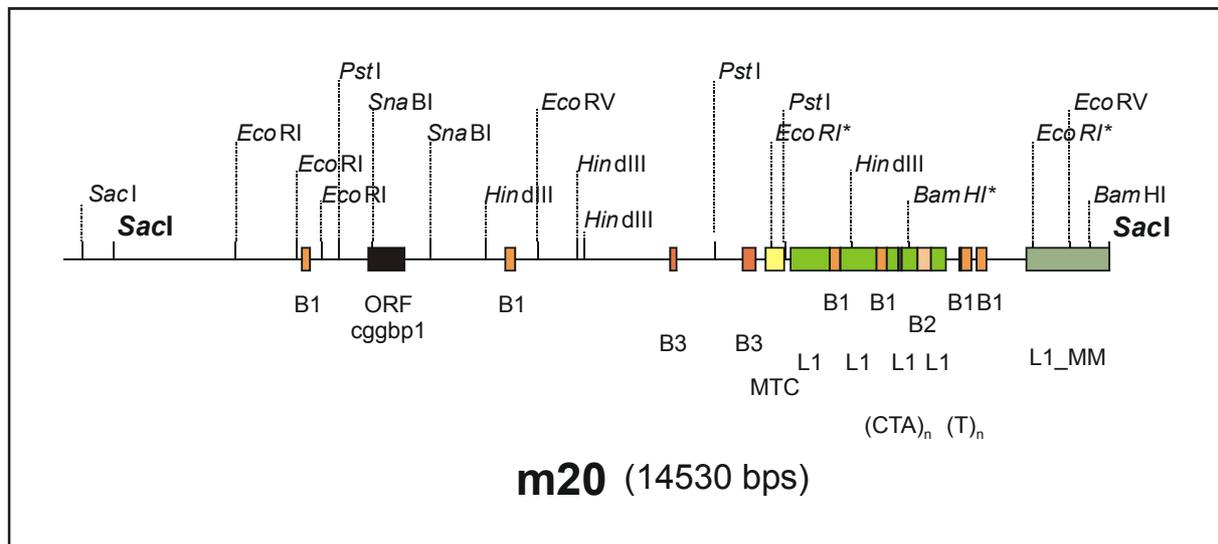


Abbildung 10 Restriktionskarte der genomischen Sequenz

In der schematischen Übersicht sind Schnittstellen von Restriktionsendonukleasen und verschiedene Klassen von repetitiven Elementen eingetragen. Das offene Leseraster des cggbp1-Gens wurde mit einem schwarzen Rechteck gekennzeichnet. Die fett gedruckten SacI-Schnittstellen entsprechen den Klonierungsstellen des Klons m20-1.1. Die kursiv gedruckten und mit Sternchen versehenen EcoRI und BamHI-Schnittstellen kennzeichnen die subklonierten Fragmente. Die repetitiven Elemente wurden mit Hilfe des Programms *Repeatmasker* (Smit und Green) identifiziert und sind je nach Familienzugehörigkeit unterschiedlich farbig markiert: die SINE-Elemente B1–B3 (Orange) (*short interspersed elements*), die LINE-Elemente L1 und die Maus-spezifische L1-Familie L1-MM (grün) (*long interspersed elements*) und das zur LTR-Familie gehörige MTC-Element (gelb) (*long terminal repeats*) (zur Übersicht Jurka, 1998).

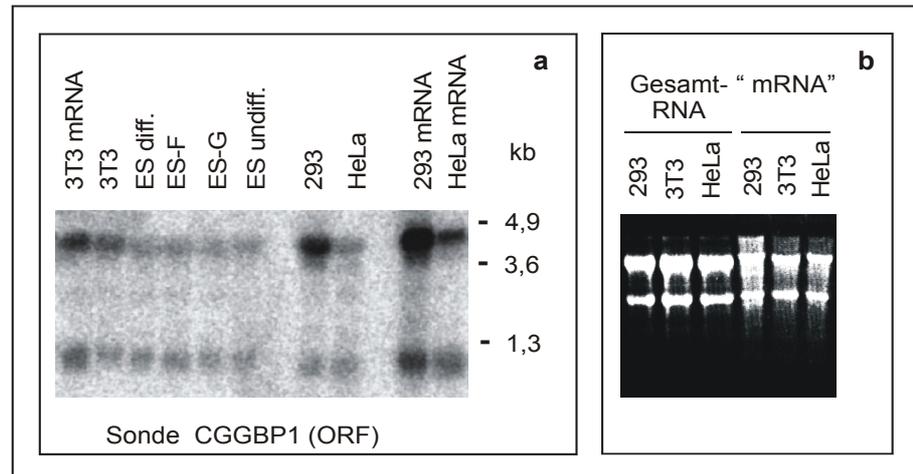
D.3 Expressionsanalysen von CGGBP1 und cggbp1

Das Expressionsmuster eines Gens in einem multizellären Organismus ist eine grundlegende Eigenschaft und kann zur Aufklärung der Funktion beitragen. Deshalb wurde RNA aus verschiedenen menschlichen Zellen und Geweben mit dem ^{32}P -markierten offenen Leseraster (ORF) des CGGBP1-Gens hybridisiert. Es wurden in allen getesteten Geweben und Zelllinien mindestens zwei spezifische Signale nachgewiesen: ein starkes Signal von 4,5 kb und eine meist schwächere, alternativ polyadenylierte mRNA von 1,2 kb. Die Hinweise auf die alternative Polyadenylierung werden im Kapitel über die Genstruktur (D.4.3.2.2) besprochen. Im folgenden wird das Auftreten der beiden Transkripte im Zusammenhang der untersuchten Gewebe beschrieben.

D.3.1 Expression von CGGBP1 und *cggbp1* in Zelllinien

Bei einer Transfer-Hybridisierung von Gesamt-RNA, die aus den menschlichen Zellen in Kultur HeLa, HEK293 und KB isoliert worden war, wurden mit einer Sonde für das ORF des CGGBP1-Gens zwei Transkripte identifiziert. Das gleiche galt auch für RNA aus embryonalen Stammzellen der Maus und für RNA aus der transformierten Mauszelllinie 3T3 (Abb. 11a).

Abbildung 11
Northern-Transfer-
Hybridisierung von
Gesamt-RNA und
Poly-A-
vorselektionierter
mRNA
verschiedener
Zelllinien des
Menschen und der
Maus.



a) Hybridisierung von jeweils 15 μ g Gesamt-RNA bzw. 2 μ g mRNA der angegebenen Zelllinien mit dem 32 P-markierten ORF des CGGBP1-Gens. RNA aus embryonalen Stammzellen (ES) der Maus wurde aus Zellen, die unter verschiedenen Bedingungen kultiviert worden waren, isoliert. So waren ES D und ES G bereits differenziert. ES-F ging ebenfalls auf differenzierte ES-Zellen zurück, die außerdem mit EF-Zellen verunreinigt waren. ES (undiff.) waren undifferenzierte ES-Zellen, die vorher von EF-Zellen getrennt worden waren. (b) Kontrolle der mRNA-Anreicherung

Das 1,2 kb große Transkript war bereits in einer anderen RNA-Transfer-Hybridisierung nachgewiesen worden. In jenem Experiment war jedoch das Signal bei 4,5 kb nicht aufgetreten (Deissler et al., 1997). Die Auflösung bei der elektrophoretischen Trennung von RNA im Agarosegel reichte hier nicht aus, die 28S rRNA-Gene des Menschen (5 kb) und der Maus (4,7 kb) von den beobachteten Signalen bei etwa 4,5 kb zu unterscheiden. Deshalb mußte überprüft werden, ob es sich bei dem Signal um eine Kreuzhybridisierung mit der 28S rRNA handelte. Dazu wurde Gesamt-RNA von Maus und Mensch neben für mRNA angereicherte Proben aufgetragen und die Stärke der Signale in den verschiedenen RNA-Präparationen verglichen (C.3.3.2). Das Signals bei 4,5 kb wurde in den Ansätzen intensiver, die für mRNA angereichert worden waren. Bei der Kontrolle der mRNA-Anreicherung auf einem Agarosegel hatte die 28S rRNA-Bande in Relation zur mRNA-Fraktion jedoch abgenommen (Abb. 11b). Vom CGGBP1-Gen des Menschen und der Maus wurden in Zelllinien somit zwei Transkripte von etwa 4,5 kb und 1,2 kb abgelesen.

D.3.2 Expression von CGGBP1 in Geweben

D.3.2.1 Expression von CGGBP1 in adulten Geweben

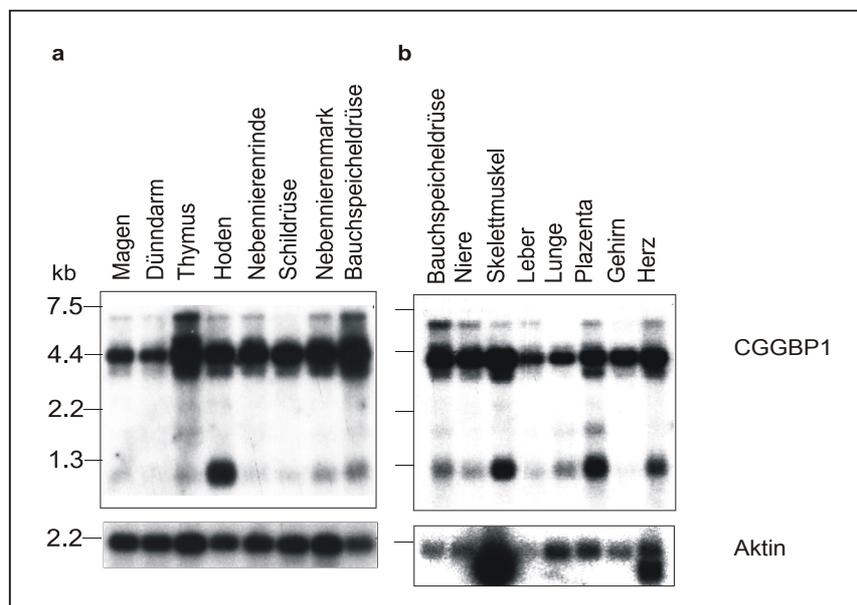
Die Transkription des menschlichen CGGBP1-Gens in verschiedenen adulten Geweben wurde unter Verwendung von kommerziell erhältlichen RNA-Filtern (Clontech) untersucht. Die RNA wurde dafür aus verschiedenen, gesunden Geweben von einem bis 110 Individuen, die zwischen 8 und 78 Jahre alt waren, extrahiert und vereinigt (siehe Quellinformationen der Firma Clontech I.6). In der Regel wurde RNA als Gemisch von RNA aus mehr als 10 Spendern verwendet. Bei einer größeren Anzahl von Spendern war ein Einfluß von individueller Variation unwahrscheinlich. Es war nicht bekannt, ob mit einem Filter RNA aus verschiedenen Geweben eines Spenders analysiert wurde.

Beide Transkripte des CGGBP1-Gens wurden in allen untersuchten Geweben nachgewiesen, wobei das Signal der mRNA von etwa 4,5 kb immer stärker war als das der 1,2 kb großen mRNA (Abb. 12). In einigen Geweben war zusätzlich ein Signal bei etwa 7 kb sichtbar. Dabei könnte es sich um noch ungespleißte Primärtranskripte handeln, die die 2,6 kb des ersten und zweiten Introns enthielten.

Abbildung 12

Hybridisierung von RNA aus adulten Geweben des Menschen

Hybridisierung von je 3 µg mRNA aus verschiedenen adulten Geweben mit dem offenen Leserahintergrund des CGGBP1-Gens (oben) bzw. mit β -Aktin (unten). Der Unterschied in der Intensität der Signale in a und b beruhte wahrscheinlich darauf, daß der Filter des RNA-Transfer-Experimentes in



(a) nur einmal vorher verwendet worden war, und die RNA auf dem Filter in (b) bereits zum dritten Mal mit einer Sonde hybridisiert wurde. Im Falle der β -Aktin-Sonde unterschieden sich zudem die Expositionzeiten. Das CGGBP1-spezifische Signal in RNA aus Leber (b) war sehr schwach. Da dies auch für das β -Aktin-spezifische Signal galt, war es möglich, daß diese RNA degradiert war.

Das Signal der 4,5 kb großen mRNA war nach der RNA-Transfer-Hybridisierung immer stärker als das Signal der 1,2 kb großen mRNA. In einigen Geweben korrelierte die Intensität des Signals bei 4,5 kb mit der des 1,2 kb großen Transkriptes, so z.B. in der Bauchspeicheldrüse, im Nebennierenmark und im Skelettmuskel. Ein starkes Signal bei 1,2 kb bedingte jedoch nicht immer ein starkes Signal bei 4,5 kb. So entsprach das Signal bei 1,2 kb in der Spur mit mRNA aus Hoden dem stärksten bisher beobachteten. Die Menge des 4,5 kb großen Transkriptes war in Hoden jedoch mit der in der Nachbarspur aufgetragenen RNA aus der Nebennierenrinde vergleichbar. Betrachtet man das Verhältnis der Signale der 4,5 kb-mRNA zu dem der 1,2 kb-mRNA, so ist es in Plazenta und Hoden am kleinsten.

D.3.2.2 Expression von CGGBP1 in verschiedenen Gehirnregionen

In einer weiteren RNA-Transfer-Hybridisierung wurde die Expression des CGGBP1-Gens in einzelnen Gehirnregionen des Menschen analysiert (Abb. 13). Die 1,2 kb große mRNA war in allen hier aufgetragenen Hirnregionen vorhanden. In der Großhirnrinde, in der das Signal der 4,5 kb großen mRNA deutlich intensiver war als in den anderen Gehirnregionen, war die 1,2 kb große mRNA im Vergleich zu den anderen Geweben jedoch quantitativ schwächer ausgeprägt. Die RNA aus der Großhirnrinde ging jedoch nur auf eine Spenderin zurück, so daß es sich hier um die individuelle Eigenschaft einer Person handeln könnte. Die RNA-Probe aus Rückenmark diente als weiteres Beispiel für die Unabhängigkeit des Vorkommens der 1,2 kb großen mRNA von der starken Transkription der 4,5 kb großen mRNA. Die Menge der

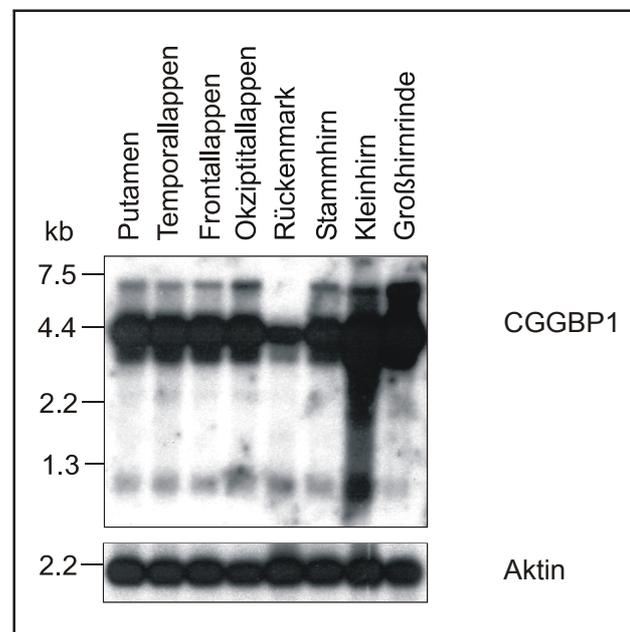


Abbildung 13 Hybridisierung von RNA aus verschiedenen Gehirnregionen des Menschen

Je 3 μ g mRNA aus verschiedenen Gehirnregionen wurde mit dem ORF des CGGBP1-Gens (oben) bzw. mit β -Aktin (unten) hybridisiert. Die RNA-Probe aus Kleinhirn zeigt das stärkste Signal aller bisher untersuchten Gewebe. Die Hintergrundsignale in der gesamten Spur deuten jedoch auf eine Degradation der RNA-Probe hin, deshalb konnte hier die Stärke der 1,2 kb mRNA nicht ausgewertet werden.

4,5kb großen mRNA war im Rückenmark im Vergleich zu den verschiedenen Gehirnregionen reduziert, während die geringe Intensität des Signals der 1,2 kb großen mRNA mit denen anderer Gehirnregionen vergleichbar war.

D.3.2.3 Expression von CGGBP1 in fötalen Geweben des Menschen

Die Hybridisierung von mRNA aus fötalen Geweben des Menschen mit einer Sonde für das offene Leseraster des CGGBP1-Gens zeigte, daß das CGGBP1-Gen in fötaler Niere und fötalem Gehirn im Vergleich zu fötaler Lunge und Leber stärker abgelesen wurde (Abb. 14). Im fötalen Gehirn war die Transkription der 1,2 kb großen mRNA in Relation zur 4,5 kb großen mRNA verringert. Neben gewebespezifischen Eigenschaften könnte es sich hier auch um Variationen aufgrund unterschiedlicher Entwicklungsstadien handeln. Die mRNAs aus Niere und Gehirns stammten aus älteren Embryonen als die mRNAs aus Lunge und Leber (Siehe Quellinformation I.6). Zudem war ein Einfluß von individuellen Eigenschaften nicht auszuschließen.

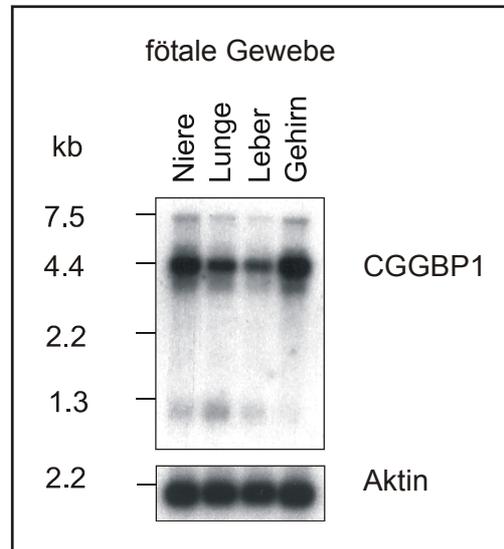


Abbildung 14 Hybridisierung von RNA aus fötalen menschlichen Geweben

Je 3 µg RNA aus den angegebenen Geweben wurde mit einer Sonde für das ORF des CGGBP1-Gens (oben) oder für β-Aktin (unten) hybridisiert.

D.3.3 Alternatives Spleißen von CGGBP1-Transkripten

Mit CGGBP1-spezifischen Primern wurden zwei Fragmente amplifiziert, wenn Gesamt-RNA aus peripheren weißen Blutzellen (PWBZ) und verschiedenen Zelllinien in RT-PCR-Experimenten eingesetzt worden war (Abb. 15a). Die Fragmente wurden aus dem Gel extrahiert, subkloniert und sequenziert. Das PCR-Produkt von 612 bp Größe enthielt im Unterschied zu dem von 502 bp das Exon III des CGGBP1-Gens. Die 5'-UTR des CGGBP1-Gens wurde also alternativ gespleißt. Beide Spleißvarianten kamen in allen getesteten Geweben gleichzeitig vor. In RNA-Transfer-Experimenten konnten Transkripte, die das Exon III enthielten, nicht nachgewiesen werden (Daten nicht gezeigt). Offenbar war das Exon III in den meisten Fällen nicht im CGGBP1-Transkript enthalten.

Um zu analysieren, ob das *cggbp1*-Gen der Maus ebenfalls alternativ gespleißt wurde, wurde Gesamt-RNA aus primären embryonalen Zellen und aus der transformierten Zelllinie 3T3 in einer RT-PCR eingesetzt (Abb. 15b). Dabei wurde nur ein Fragment von 770 bp amplifiziert. Das *cggbp1*-Gen wurde demnach u. a. in embryonalen Stammzellen der Maus transkribiert, aber nicht alternativ gespleißt. Bei Sequenzanalysen wurde zudem keine dem Exon III ähnliche Sequenz gefunden.

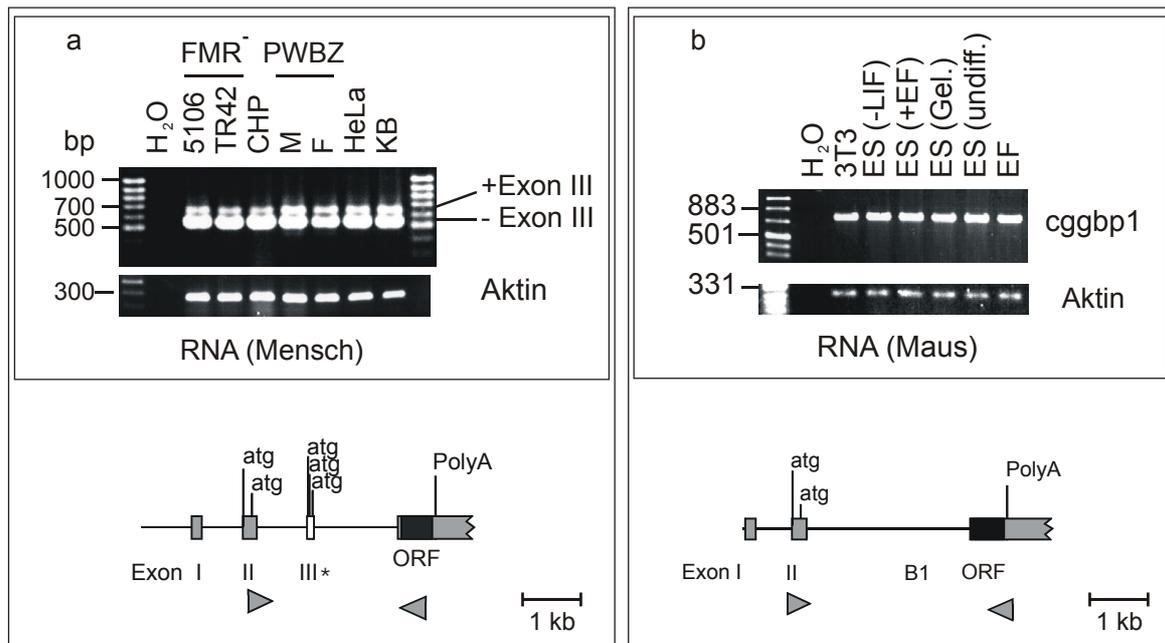


Abbildung 15 Alternatives Spleißen in der 5'-UTR des CGGBP1-Gens.

RT-PCR mit 300 ng cytoplasmatischer oder Gesamt-RNA mit CGGBP1- bzw. *cggbp1*-spezifischen Primern. Die Lage der Primer sind aus den schematischen Karten ersichtlich. a) In transformierten Zelllinien von Patienten mit dem Fragilen-X-Syndrom (5106 und TR42), in peripheren weißen Blutzellen (PWBZ) eines männlichen Probanden (M) und einer weiblichen Probandin (F) und in den menschlichen Zelllinien HeLa, CHP und KB wurden zwei Transkriptvarianten amplifiziert. b) In der murinen Zelllinie 3T3, in embryonalen Stammzellen (ES) und in embryonalen Fibroblasten (EF) der Maus wurde ein Signal von 770 bp nachgewiesen. Das *cggbp1*-Gen wurde unter allen getesteten Bedingungen transkribiert. So wuchsen ES (-LIF) in Abwesenheit des LIF-Proteins (C.1.3.1), und ES (Gel.) auf Gelatine ohne eine Schicht embryonaler Fibroblasten. Die ES-Zellen differenzierten unter beiden Kulturbedingungen. Die undifferenzierten ES (+EF) Zellen waren noch mit EF-Zellen verunreinigt, bevor RNA präpariert wurde, während ES (undiff.) vorher von EF-Zellen getrennt worden waren. In EF-Zellen wurde das *cggbp1*-Gen ebenfalls transkribiert.

D.3.4 Transkription des *cggbp1*-Gens während der Embryonalentwicklung

Das *cggbp1*-Gen wurde in embryonalen Stammzellen transkribiert (D.3.1; Abb. 11). Bei RNA-Transfer-Experimenten mit RNA aus Mausembryonen, wurden *cggbp1*-spezifische Signale in allen vier analysierten Entwicklungsstadien nachgewiesen (Abb. 16). In der Maus wurden – wie beim Menschen auch – zwei Transkripte von 1,2 kb und 4,5 kb gebildet. Es wurden weder Veränderungen der Signalintensitäten der beiden Transkripte in den verschiedenen Entwicklungsstadien, noch im Verhältnis der 4,5 kb zur 1,2 kb mRNA beobachtet. Aufgrund der großen Anzahl von Mausembryonen (siehe Quellinformation I.6) war hier auszuschließen, daß eine Abweichung in einem einzelnen Individuum dieses Ergebnis verfälschte.

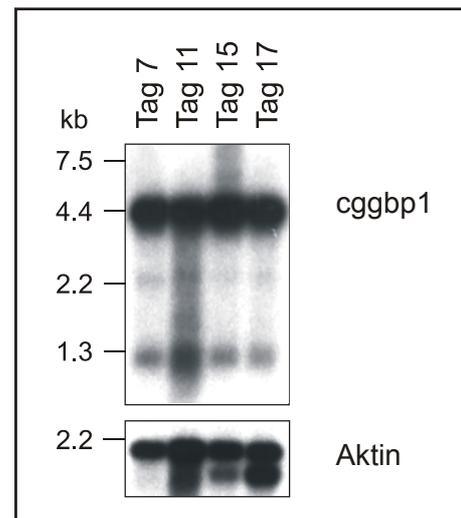


Abbildung 16 Hybridisierung von RNA aus Mausembryonen unterschiedlicher Entwicklungsstadien

Je 3 µg RNA aus Mausembryonen wurde mit einer ^{32}P -markierten Sonde für das ORF des CGGBP1-Gens (oben) oder für β -Aktin (unten) in einem RNA-Transfer-Experiment hybridisiert.

D.4 Struktur und Konservierung des CGGBP1-Gens

Die Genstrukturen der hier klonierten CGGBP1-Gene des Menschen und der Maus wurden aus RNA-Transfer- und RT-PCR-Analysen abgeleitet (Abb. 11-16). Die aus 19 EST-Klonen abgeleitete Genstruktur für das CGGBP1-Gen (Diplomarbeit Naumann, 1998) konnte ergänzt werden. Die genomische Sequenz des CGGBP1-Gens der Ratte (*r-cggbp1*) wurde hier in der vom Rattengenomprojekt (I.7) bisher veröffentlichten Sequenz über Datenbankrecherchen identifiziert und mit den Sequenzen des CGGBP1- und des *cggbp1*-Gens verglichen. Zur Aufklärung der genomischen Organisation der drei Gene wurden auch EST-Klone hinzugezogen (Boguski et al., 1993; Schuler et al., 1997). EST-Klone anderer Spezies, die eine hohe Sequenzidentität im Vergleich zum CGGBP1-Gen aufwiesen, wurden ebenfalls analysiert. Die Zugriffsnummern der verwendeten EST-Klone sind im Anhang aufgeführt (I.5).

D.4.1 Identifizierung von CGGBP1-Genen in anderen Spezies

Als das Humane Genomprojekt entworfen wurde, war die Zielsetzung, neben der Sequenz des menschlichen Genoms, auch die Genomsequenz anderer Spezies, sogenannter Modellorganismen, zu bestimmen (Watson und Cook-Deegan, 1991; International Human Genome Sequencing Consortium, 2001). So wurden u. a. bisher die Genome der Hefe *Saccharomyces cerevisiae* (Goffeau et al., 1996); des Nematoden *Caenorhabditis elegans* (The *C.elegans* Sequencing Consortium, 1998) und der Maus *Mus musculus* C57BL/6J (Mouse Genome Sequencing Consortium, 2002) weitgehend sequenziert. Außerdem wurden im Rahmen verschiedener EST-Klon-Projekte (Adams et al., 1991; Okubo et al., 1992; Aaronson et al., 1996; Hillier et al., 1996; Schuler et al., 1997; Strausberg et al., 1999) zufällig ausgewählte cDNAs sequenziert, um eine Übersicht über den transkribierten Teil des Genoms zu gewinnen. Bei der Analyse von EST-Klonen ist jedoch Vorsicht geboten, da sie von einzelnen Molekülen abgeleitet werden, deren biologische Relevanz unbekannt ist. Außerdem werden die meisten EST-Klonsequenzen nicht editiert, so daß mögliche Fehler v.a. bei längeren Sequenzen oft nicht zu beurteilen sind.

Da die genomischen Sequenzen der Modellorganismen sowie die EST-Klone in den öffentlichen Datenbanken zugänglich waren, und der Vergleich der Gensequenzen oder Genstrukturen zur Aufklärung biologisch wichtiger Regionen im CGGBP1-Gen hätte beitragen können, wurden alle am *National Center for Biotechnology Information* (NCBI) (I.7) verfügbaren Datenbanken wie im Methodenteil beschrieben nach CGGBP1-ähnlichen Sequenzen durchsucht. In den anderen eukaryontischen Genomsequenzen, die bis März 2003

am NCBI zugänglich waren (z. B. Fische, Nematoden, Hefen) (*genomic BLAST pages* @ <http://www.ncbi.nlm.nih.gov/BLAST/>), wurden keine dem CGGBP1-Gen ähnlichen Sequenzen entdeckt.

D.4.1.1 Identifizierung des CGGBP1-Gen der Ratte – r-cggbp1

Eine dem ORF des CGGBP1-Gens ähnliche Sequenz wurde unter Verwendung des BLAST-Programmes (Altschul et al., 1997) in einer genomischen Sequenz der Ratte identifiziert. Unter der Zugangsnummer AC128161 waren in der Datenbank am NCBI insgesamt 205.274 bp aus dem Rattengenom abgelegt (Rat Genome Sequencing Consortium). Die Sequenz war nach den Angaben im Datenbankeintrag lückenhaft, was durch eine Folge des Buchstabens N gekennzeichnet worden war. Zudem war die Reihenfolge einzelner durchgängig sequenzierter Fragmente noch nicht bekannt war. Deshalb wurde aus der insgesamt 205.274 bp langen Sequenz ab Nukleotid 111970 ein 12084 bp langer Abschnitt entnommen und mit den genomischen Sequenzen des CGGBP1-Gens des Menschen und der Maus verglichen. In den ausgewählten 12.084 bp war eine Folge von 50 Ns enthalten, die auf fehlende Sequenz hinwies (siehe Unterbrechung der durchgezogenen Linie in Abb. 17 c). Der Vergleich der genomischen Sequenzen des Menschen, der Maus und der Ratte zeigte, daß diese Lücke in der Sequenz der Ratte von Sequenzen flankiert war, die v. a. dem Gen der Maus sehr ähnlich waren (Abb. 18). Anhand von EST-Klonen und dem Vergleich zu den CGGBP1-Genen des Menschen und der Maus wurde eine Genstruktur des r-cggbp1 postuliert. Die Lücke in der genomischen Sequenz lag somit im ersten Intron. Es war nicht auszuschließen, daß das Gen der Ratte dort zusätzliche Sequenz enthielt.

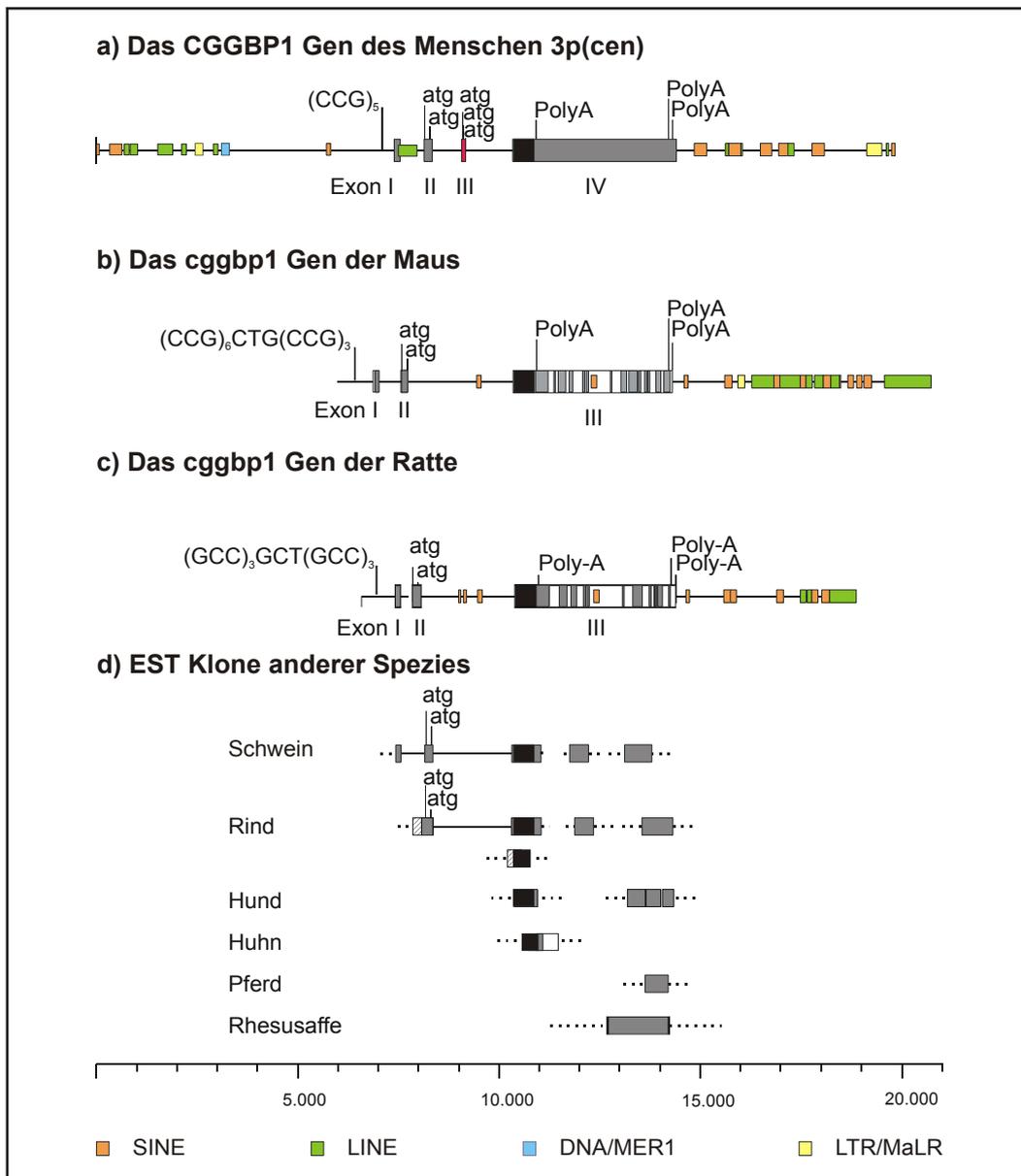


Abbildung 17 Genomische Organisation des CGGBP1-Gens des Menschen (a) und der Maus (b).

In der schematischen Darstellung sind Exons des CGGBP1-Gens als große und repetitive Elemente als kleine Rechtecke gekennzeichnet. Aufgrund von Ähnlichkeiten der Sequenz und EST-Analysen wurde die gezeigte Genstruktur der Ratte (c) postuliert. Zusätzliche, wahrscheinlich nicht funktionelle, Startcodons in der 5'-UTR wurden durch „atg“ aufgeführt. Graue Bereiche in den UTRs der Gene der Maus und Ratte zeigen Sequenzidentitäten zwischen 78 und 95 % zu der entsprechenden Region des menschlichen Gens auf. Das ORF der Gene ist als schwarzes Rechteck dargestellt und das alternativ gespleißte Exon III ist rot markiert. (d) In der Datenbank dbEST (Boguski et al., 1993) abgelegte EST-Klone anderer Spezies wurden gemäß ihrer Ähnlichkeit zum CGGBP1-Gen eingetragen. Dabei bedeutete eine durchgezogene Linie zwischen zwei Rechtecken, daß die Sequenz zwischen diesen Rechtecken durchgehend war, was wahrscheinlich auf gespleißte Transkripte zurückging. Bei gepunkteten Linien fehlten weitere Sequenzinformationen. Eine Schraffierung in (d) besagt, daß eine hohe Sequenzidentität zur genomischen Sequenz der Introns bestand. Ein weißes Rechteck in (d) zeigt an, daß eine EST-Sequenz bekannt, jedoch zum CGGBP1-Gen nicht ähnlich war.

Um die Bedeutung von hohen Sequenzidentitäten innerhalb von Exons beurteilen zu können, sollten – wenn möglich – zu dem Gen benachbarte Sequenzen untersucht werden, die wahrscheinlich keinem oder einem geringen Selektionsdruck unterliegen. Damit erhält man einen Eindruck über die Substitutionsraten, die in dieser chromosomalen Region herrschen. Zwei EST-Sequenzen aus Rindern waren zu den Sequenzen des menschlichen Introns I und III sehr ähnlich (Abb. 17d). Bisher – und möglicherweise voreilig – geht man davon aus, daß Introns wenig funktionelle Elemente enthalten und deshalb Nukleotidaustausche geduldet werden. Allein aus der Konservierung der Exons des CGGBP1-Gens in Sequenzen von Rinder-ESTs könnte somit nicht auf einen negativen Selektionsdruck auf die translatierte Region geschlossen werden. Um einen Überblick über die Verteilung konservierter Regionen in den untersuchten genomischen Sequenzen zu erhalten, wurden die genomischen Sequenzen des Menschen, der Maus und der Ratte untereinander mit Hilfe der Programme *Compare* und *Dotplot* aus dem GCG-Programmpaket untersucht (Abb.18).

Die genomischen Sequenzen war zwischen Mensch und Maus (Abb. 18a) und Mensch und Ratte (Abb. 18b) jeweils in den Exons, in den die Exons direkt flankierenden Intronsequenzen und in den Promotorregionen hoch konserviert, während die genomischen Sequenzen zwischen Maus und Ratte insgesamt sehr ähnlich waren (Abb. 18c). Die Abweichungen im 3'-Bereich der genomischen Sequenz von Maus und Ratte beruhten entweder auf Fehler oder möglicherweise darauf, daß die Sequenzierung der chromosomalen Region der Ratte noch nicht abgeschlossen war und der Datenbankeintrag als vorläufige Sequenz zu betrachten war. Außerdem handelte es sich bei den zwischen Ratte und Maus nicht übereinstimmenden Bereichen um repetitive Elemente. Die Aufklärung der Reihenfolge hochrepetitiver DNA-Abschnitte gestaltet sich generell schwierig (International Human Genome Consortium, 2001).

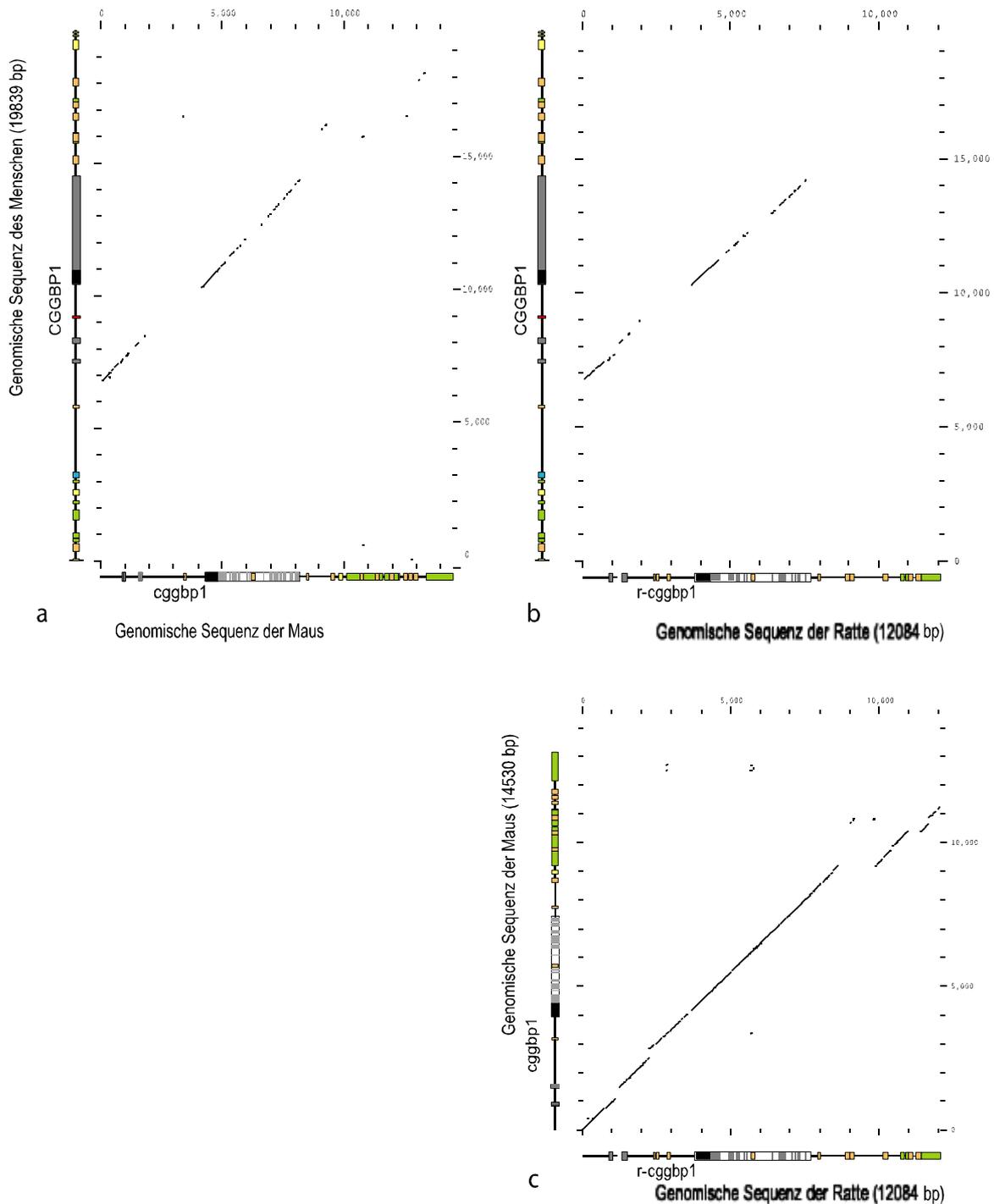


Abbildung 18 Ähnlichkeiten in den genomischen Sequenzen.

Die genomischen Sequenzen des Klons für das CGGBP1-Gen des Menschen (AF094481), der Maus (AY189895) und der Ratte (entnommen aus AC128161) wurden untereinander mit Hilfe des Programms Compare verglichen und mit dem Programm Dotplot dargestellt. Die window size wurde auf 40 und die stringency auf 30 eingestellt, d.h. daß die Sequenz mit einem 40 bp-Raster abgesucht wurde, in dem 30 bp übereinstimmen mußten, damit ein Punkt in einem Diagramm durch das Programm Dotplot gezeichnet wurde. Viele Punkte, also Übereinstimmungen, ergeben eine Linie. Zur Orientierung wurde neben den Diagrammen die genomische Organisation der Gene analog der Abb. 17 aufgeführt.

D.4.1.2 EST-Klone des CGGBP1-Gens anderer Spezies

Mit Hilfe des BLAST-Programms (Altschul et al., 1997) in der Datenbank für EST-Klone (Boguski et al., 1993) wurden auch EST-Klone von anderen Säugetieren (Ratte, Schwein, Rind, Hund, Pferd, Rhesusaffe, Schimpanse) und von Hühnern gefunden, die der Nukleotidsequenz des offenen Leserasters des CGGBP1-Gens sehr ähnlich waren. Die Datenbankzugriffsnummern der EST-Klone sind im Anhang I.5 aufgeführt. Die Suche nach zur Proteinsequenz ähnlichen Sequenzen erbrachte keine zusätzliche Information. Da die EST-Klone häufig miteinander überlappten, wurden Consensussequenzen aus den vorhandenen Sequenzen gebildet, die die sonst in EST-Klonen vorhandene Fehlerquote von etwa 3% verringerten (Hillier et al., 1996).

D.4.2 Identifizierung von Domänen in der Aminosäuresequenz des CGGBP1

Wurde der nicht redundante Teil von Genbank (Benson et al., 1998) mit dem BLAST Programm (Altschul et al., 1997) durchsucht, wurde mit dem genomischen Klon keine hohe Ähnlichkeit zu einer bereits charakterisierten Sequenz gefunden. Mit der Proteinsequenz des CGGBP1 wurde bei der Durchsuchung verschiedener Gen- und Proteindatenbanken ebenfalls keine überzeugende Ähnlichkeit zu anderen Proteinen gefunden. Eventuell vorhandene Proteindomänen oder Sequenzmotive sollten hier mit Hilfe der Datenbank Interpro (Mulder et al., 2003) identifiziert werden. Die Kernlokalisationssequenz, deren Funktionalität experimentell nachgewiesen worden war (Müller-Hartmann et al., 2000), wurde bei dieser Sequenzbasierten Recherche erkannt.

Besonders auffällig war eine Ähnlichkeit mit einer Unterklasse von Zink-Finger-Proteinen, den U1-ähnlichen Zinkfingerproteinen (Nelissen et al., 1991; Matsushima et al., 1997; Bedford et al., 1998). Einige der Proteine, aus denen dieses Motiv abgeleitet worden war, binden RNA und scheinen eine Rolle beim Spleißen zu spielen. Diese Unterklasse der Zink-Finger-Proteine weist abweichend zu DNA-bindenden Zink-Finger-Proteinen einen anderen Abstand zwischen den das Zink-Ion komplexierenden Cystein- und Histidinresten auf (http://www.expasy.org/cgi-bin/ft_aligner?psa=PS50171). Außerdem wird es N-terminal von aromatischen Aminosäuren und C-terminal von positiv geladenen Aminosäuren flankiert. Im CGGP1 finden sich die charakteristischen Abstände sowie die flankierenden Aminosäuren wieder (Abb. 19).

```

          *           20           *
CIZ1_HUMAN : PFCCTVGNRYFKTPRKFVE--HVKSQGHKDKAKE
CIZ1_HUMAN : YICRICHKFYHSNS-GAQLSHCKSLGHFENLQK
MAT3_HUMAN : FYCKLCSLFYTNEE-VAKNTHCSSLPHYQKLLK
MAT3_RAT/7 : FYCKLCSLFYTNEE-VAKNTHCSSLPHYQKLLK
RU1C_HUMAN : FYCDYCDTYLTHDSPSVRKTTHCSGRKHKENVKD
RU1C_MOUSE : FYCDYCDTYLTHDSPSVRKTTHCSGRKHKENVKD
RU1C_XENLA : FYCDYCDTYLTHDSPSVRKTTHCSGRKHKENVKD
S3A2_HUMAN : YECKLCLTLHNNEGSYL--AHTQGKKHQTNLAR
S3A2_MOUSE : YECKLCLTLHNNEGSYL--AHTQGKKHQTNLAR
S3A3_HUMAN : YNCEICGNYTYRGPKAFQR-HFAEWRHAHGMRG
S3A3_MOUSE : YNCEICGNYTYRGPKAFQR-HFAEWRHAHGMRG
PR09_YEAST : IYCFECSRWFKTSS-VFE-SHLVGKIHKKNESK
PR09_YEAST : YRCEICSNKVYNGRRTFER-HFNEERHLYHLRC
CGGBP1     : LFCCTSNVVLNHRKSAISDHLKSKTHTKRKA
          C  C           H  H

```

Abbildung 19 Sequenzvergleich von CGGBP1 (Aminosäure 41-91) mit dem C2H2-Zinkfingermotiv der U1-ähnlichen Zinkfingerproteine (Matrin-Typ).

Der Sequenzvergleich der 14 Proteine aus Mensch, Maus, Ratte, Frosch und Hefe, die das U1-ähnliche Zinkfingermotiv aufweisen, wurde der Adresse http://www.expasy.org/cgi-bin/ft_aligner?psa=PS50171 entnommen und durch die Sequenz von CGGBP1 erweitert. Die Zugriffsnummern der Proteine sind im Anhang I.8 aufgelistet. Die das Zink-Ion komplexierenden Reste sind auf schwarzem Hintergrund gedruckt. Eine graue Hinterlegung weist auf Sequenzidentitäten zwischen CGGBP1 und den aufgeführten Proteinen hin. Unter Berücksichtigung von Aminosäureeigenschaften wäre die Ähnlichkeit zwischen der Sequenz aus CGGBP1 und der die Cystein- und Histidinreste flankierenden Sequenz noch größer.

D.4.3 Allgemeine Struktur der CGGBP1- und cggbp1-Gene

Die Strukturen der CGGBP1-Gene des Menschen, der Maus und der Ratte waren einander sehr ähnlich (Abb. 17). Die Ergebnisse von RNA-Transfer- und RT-PCR-Experimenten wurden durch Sequenzanalysen bestätigt. Die hier entworfenen Genstrukturen wurden durch die nachgewiesenen Spleiß- und Polyadenylierungssignale belegt (D.4.3.2.2), die mit den hochkonservierten Consensussequenzen übereinstimmten (Dibb, 1993; International Human Genome Sequencing Consortium, 2001). So begannen alle Introns der CGGBP1-Gene immer mit GT und endeten mit AG.

D.4.3.1 Das offene Leseraster der CGGBP1-Gene

Das offene Leseraster (ORF) des menschlichen CGGBP1-Gens war durch die partielle Sequenzierung des 5'-(CGG)_n-3'-bindenden Proteins 1 festgelegt worden (Deissler et al., 1997). So war es direkt möglich, die transkribierte Region in translatierte und in nicht-translatierte Regionen (UTR) zu unterteilen. Die Nukleotidsequenz des ORF war zwischen dem menschlichen Gen und den Genen von Maus und Ratte mit 91% hoch konserviert. Diese

Sequenzähnlichkeit bestätigte frühere Beobachtungen aus heterologen Hybridisierungs-Experimenten (Abb. 3). Das ORF war nicht durch Introns unterbrochen, wenn man die genomischen Sequenzen des Menschen, der Maus und der Ratte analysierte. Nur drei Sequenzabweichungen in der kodierenden Region führten zu Austausch von Aminosäuren im 167 Aminosäuren langen Protein der Maus: Mensch → Maus Threonin 95 →Prolin, Valin 101→Alanin, Asparagin 163→Serin (Abb. 20). Die Proteinsequenz des CGGBP1-Gens der anderen hier analysierten Spezies wurde meist aus Consensussequenzen von EST-Klonen abgeleitet. Das CGGBP1 war in Schweinen und Rindern im Vergleich zum menschlichen Protein identisch.

	*	20	*	40	*				
CGGBP1	:	MERFVV	TAPPARNRSKTALYV	TPLDRVTE	FGGELHEDGGKLF	CTSCNVV	LNHVVRKS : 56		
Schwein	:	MERFVV	TAPPARNRSKTALYV	TPLDRVTE	FGGELHEDGGKLF	CTSCNVV	LNHVVRKS : 56		
Ratte	:	MERFVV	TAPPARNRSKTALYV	TPLDRVTE	FGGELHEDGGKLF	CTSCNVV	LNHVVRKS : 56		
Rind	:	MERFVV	TAPPARNRSKTALYV	TPLDRVTE	FGGELHEDGGKLF	CTSCNVV	LNHVVRKS : 56		
Maus	:	MERFVV	TAPPARNRSKTALYV	TPLDRVTE	FGGELHEDGGKLF	CTSCNVV	LNHVVRKS : 56		
Hund*	:	-----	-----	-----	-----	-----	----- : -		
Huhn*	:	-----	-----	-----	-----	-----	----- : -		
		60	*	80	*	100	*		
CGGBP1	:	AI	SDHLKSKTHT	TKRKA	EFEEQ	NVRKKQ	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 112	
Schwein	:	AI	SDHLKSKTHT	TKRKA	EFEEQ	NVRKKQ	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 112	
Ratte	:	AI	SDHLKSKTHT	TKRKA	EFEEQ	NVRKKQ	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 112	
Rind	:	AI	SDHLKSKTHT	TKRKA	EFEEQ	NVRKKQ	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 112	
Maus	:	AI	SDHLKSKTHT	TKRKA	EFEEQ	NVRKKQ	RPLTASIQCN	STAQTEKVS	VIQDFVKMCL : 112
Hund*	:	-----	-----	-----	-----	NKQR	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 32	
Huhn*	:	-----	-----	-----	-----	NVRKKQ	RPLTASIQCNSTAQTEKVS	VIQDFVKMCL : 35	
		120	*	140	*	160			
CGGBP1	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	QLLNSQDC : 167		
Schwein	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	QLLNSQDC : 167		
Ratte	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	QLLNSQDC : 167		
Rind	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	QLLNSQDC : 167		
Maus	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	QLLNSQDC : 167		
Hund*	:	G	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSDQLRRAYL	PDGYENEN	KQKNSQDC : 87		
Huhn*	:	E	ANIPLEKADHFAVRAFLSR	HVKNGGSI	PKSEQLRRAYL	PDGYENEN	QLINTEER : 90		

Abbildung 20 Aminosäuresequenz des CGGBP1-Gens in verschiedenen Spezies.

Die Proteinsequenzen des Menschen, der Maus und der Ratte wurden aus genomischen Sequenzen abgeleitet. Vor der Ableitung der Proteinsequenz des CGGBP1-Gens bei Schweinen und Rindern wurde eine Consensussequenz - soweit möglich - aus mindestens 4 EST-Klonen gebildet. Ein 60 bp langer Abschnitt am N-Terminus der Sequenz aus Schweinen beruht jedoch nur auf einem EST-Klon. Die mit Stern gekennzeichneten Proteinsequenzen beruhen z. T. auf nur einem EST-Klon und enthalten möglicherweise deswegen mehr Sequenzabweichungen als die anderen Sequenzen. Bei Übereinstimmungen mit der menschlichen Proteinsequenz (erste Zeile) sind die Aminosäuren auf schwarzem Grund abgebildet. Abweichende Aminosäuren, die den chemischen Charakter des Proteins wahrscheinlich nicht wesentlich verändern, sind auf grauem Grund gedruckt. Die Kernlokalisationssequenz ist mit grüner, das potentielle Zinkfinger-Motiv mit gelber Farbe hinterlegt.

Am Verhältnis von Mutationen, die zu einem Austausch einer Aminosäure führen (nicht synonyme Mutation), zu Mutationen, die die kodierte Proteinsequenz nicht verändern (stille Mutation), kann die Richtung und die Stärke des an diesem Genlocus herrschenden Selektionsdruckes abgelesen werden. Mit dem dN-Wert wird die Mutationsrate in nicht synonymen Positionen berechnet (Mutationen/Anzahl der nicht-synonymen Positionen) (Yang und Nielson, 2000). dS gibt die Mutationsrate in synonymen Codons an. Im allgemeinen geht man davon aus, daß Aminosäureaustausche negativer Selektion unterliegen, während stille Mutationen eher als neutral anzusehen sind (siehe aber auch Akashi, 1995; Duan et al., 2003). Der dN/dS-Wert spiegelt das Verhältnis dieser Mutationsraten wider und liegt bei orthologen Paaren von Genen meist unter 0,05. Die dN/dS-Werte für das CGGBP1 des Menschen und vier Säugern wurden unter Verwendung des Programms PAML (Yang, 1997) nach der Methode von Yang und Nielson (2000) berechnet (Tab. 2).

dN/dS	Mensch	Maus	Ratte	Rind	Schwein
Mensch	-				
Maus	0,0181	-			
Ratte	0,0010	0,0326	-		
Rind	0,0010	0,0212	0,0010	-	
Schwein	0,0010	0,0169	0,0010	0,0010	-

Tabelle 2 dN/dS-Werte.

Die dN/dS-Werte wurden nach der Methode von Yang und Nielson, 2000 aus genomischer Sequenz (Mensch, Maus, Ratte) oder aus Consensus-Sequenzen aus EST-Klonen (siehe Legende Abb. 20) berechnet. Diese Sequenzen aus Hunden und aus Hühnern wurden in der Berechnung der dN/dS-Werte nicht berücksichtigt, da sie unvollständig waren und hauptsächlich auf nur einem EST-Klon beruhten.

D.4.3.2 Die nicht-translatierten Regionen der CGGBP1-Gene

D.4.3.2.1 Die 5'-nicht-translatierte Region (5'-UTR)

Diese Region der CGGBP1-Gene wurde durch EST-Klon-Analysen definiert und beim Menschen und der Maus durch RT-PCR-Experimente bestätigt (Abb. 15). Die Sequenzen der 5'-UTR waren zwischen Mensch, Maus und den zur Verfügung stehenden Sequenzen aus Ratte, Rind und Schwein mit Sequenzidentitäten von mindestens 78 % (Mensch-Ratte) bis zu 91 % (Mensch-Rind) hoch konserviert (Abb. 21). Die Exons des Maus- und Rattengens waren etwas länger, als die des menschlichen Gens. Die zusätzliche Sequenz stimmte im Falle des 3'-Endes des ersten Exons zu 70% mit der menschlichen Intronsequenzen überein. Für die Existenz eines dritten Exons in den verfügbaren genomischen oder EST-Sequenzen der anderen Säugetiere gab es keine Hinweise.

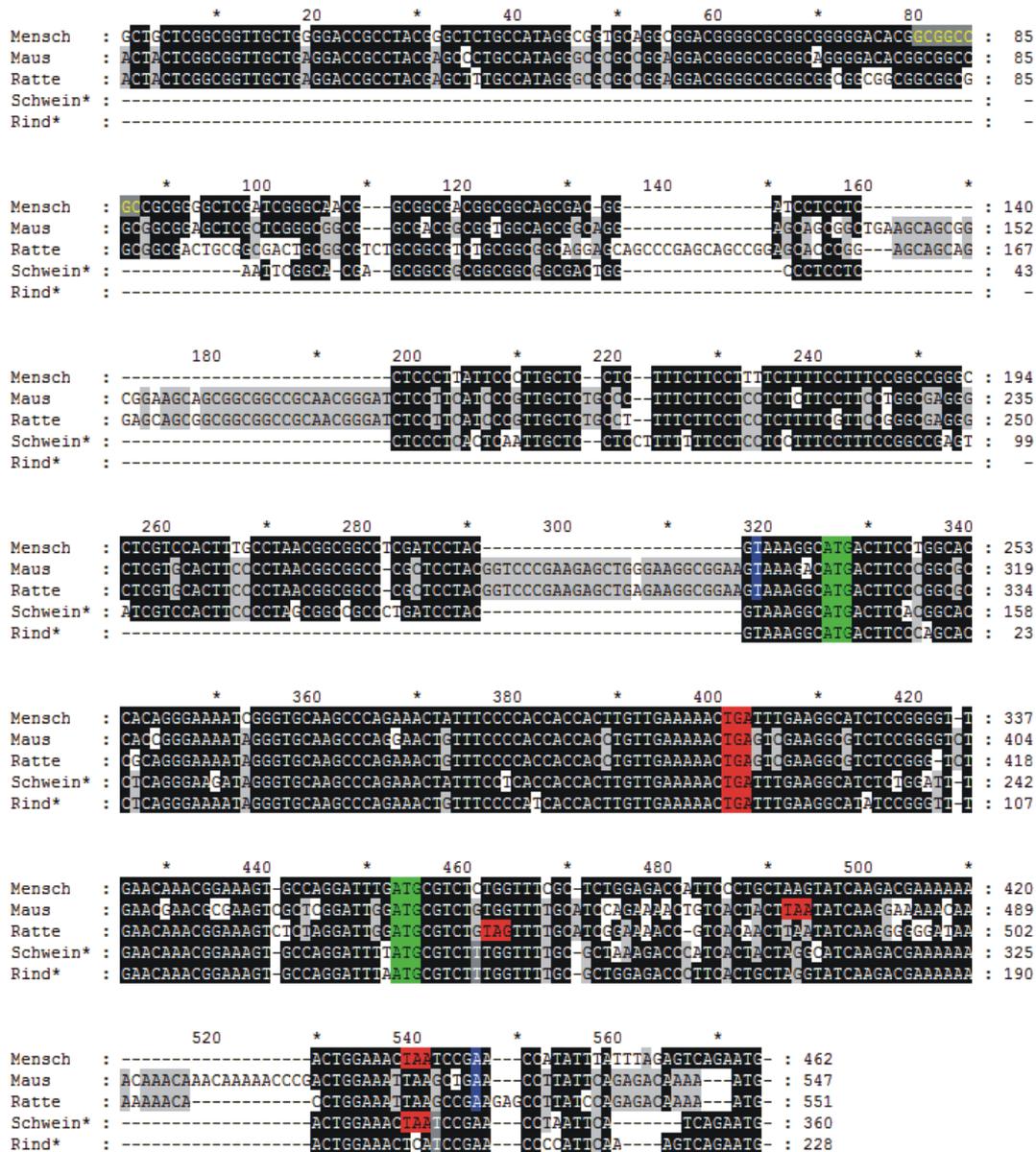


Abbildung 21 Die 5'-UTR der CGGBP1-Gene Sequenzvergleiche der 5'-UTR des menschlichen CGGBP1-Gens mit derjenigen der Maus, Ratte, Schwein und Rind. Die Sequenz aus HeLa-Zellen (Mensch) ist in der obersten Zeile angegeben. Als 5'-Ende der 5'-UTR des menschlichen Gens wurde hier das 5'-Ende des am weitesten 5'-gelegenen EST-Klons gewählt (BI601953). Nukleotide der anderen Sequenzen, die mit der Sequenz des Menschen übereinstimmen, sind auf schwarzem Grund dargestellt. Abweichungen von der menschlichen Sequenz (weißer Hintergrund), die in mehr als einer Sequenz zu finden waren, wurden durch eine graue Schattierung gekennzeichnet. Die mit Stern versehenen Sequenzen aus Schwein und Rind stammten von EST-Klonen ab, so daß Exon/Exon Grenzen (blaue Farbe) nicht angegeben werden konnten. Sie waren am 5'-Ende zudem nicht vollständig. Die zusätzlichen ATGs in der 5'-UTR sind grün, die Stopcodons der dazugehörigen Minicistrons rot unterlegt. Um diese Abbildung mit der Abb. 24 vergleichen zu können, wurde zur Orientierung die NotI-Schnittstelle im ersten Exon mit gelber Schrift auf grauem Grund eingetragen. Die Zahlen beziehen sich auf die Länge der gespleißten 5'-UTR-Sequenzen.

Die 5'-UTR der CGGBP1-Gene wiesen Merkmale auf, die bei der posttranskriptionellen Regulation eine Rolle spielen könnten. Die 5'-UTR der CGGBP1-Gene des Menschen und der Nagetiere waren mit ~450 bp etwa doppelt so lang, wie durchschnittliche 5'-UTRs (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). Es wurde postuliert, daß Sekundärstrukturen in der 5'-UTR die Translation verringern könnten (Kozak, 1991). Das erste Exon des CGGBP1-Gens war GC-reich, so daß sich Sekundärstrukturen hoher Stabilität bilden könnten. Im zweiten Exon der 5'-UTR begannen in allen untersuchten Sequenzen zwei sogenannte uORFs (*upstream ORFs*) (Kozak, 1996). Es handelt sich dabei um stromaufwärts vom eigentlichen Startcodon des produzierten Proteins gelegene kurze ORFs. Diese Minicistrons werden ebenfalls im Rahmen von posttranskriptionaler Regulation der Genexpression diskutiert (zur Übersicht Meijer und Thomas, 2002). Alle zusätzlichen ATGs in der 5'-UTR waren nicht im Leseraster des CGGBP1. Die Sequenzumgebung dieser ATGs wurde mit der Consensussequenz (Kozak, 1987) für ein funktionelles Startcodon verglichen (Tab. 3). Die Sequenzumgebung der ATGs wichen z. T. stark von der Kozak-Sequenz ab. Das dritte Exon des menschlichen CGGBP1-Gens enthielt drei weitere solcher Minicistrons (Tab. 3). Über Sequenzvergleiche mit der spezialisierten Datenbank für UTR-Sequenzen (Pesole et al., 2002) wurden keine signifikanten Ähnlichkeiten zu den UTR anderer Gene gefunden.

	Abstand [nt] vom ATG (+1) in der 5'-UTR						ORF (Codons)						
		-9	-6	-3	+1	+4							
	Consensus	G	C	C	G	C	<u>R</u>	C	C	ATG	<u>G</u>		
Mensch	-323 / -221	a	g	t	a	a	G	g	C	ATG	c	25 (+/- Exon III)	
	-198 / -96	c	a	g	G	a	t	t	t	g	ATG	c	22 (+/- Exon III)
	-113 /	t	t	C	t	g	t	G	g	a	ATG	t	7
	- 93	t	C	t	G	g	a	t	C	C	ATG	a	3
	- 44	a	g	a	G	a	g	t	t	C	ATG	t	19 *
	+ 1	a	g	a	G	t	C	A	g	a	ATG	G	167
Maus	-242	a	g	t	a	a	G	a	C	ATG	a	25	
	-115	t	C	g	G	a	t	t	g	g	ATG	a	13
	+ 1	a	g	a	G	a	C	A	a	a	ATG	G	167
Ratte	-228	a	g	t	a	a	G	g	C	ATG	a	25	
	-102	t	a	g	G	a	t	t	g	g	ATG	c	3
	+ 1	a	g	a	G	a	C	A	a	a	ATG	G	167
Schwein	-214	c	g	t	a	a	G	g	C	ATG	a	25	
	-89	c	a	g	G	a	t	t	t	t	ATG	c	22
	+ 1	t	t	C	a	t	C	A	g	a	ATG	G	167
Rind	-227	a	g	t	a	a	G	g	C	ATG	a	25	
	-102	c	a	g	G	a	t	t	t	a	ATG	c	35 *
	+ 1	a	a	a	G	t	C	A	g	a	ATG	G	167

Tabelle 3 Sequenzumgebung der Startcodons in den CGGBP1-Genen in fünf Säugetierspezies.

Vergleich der Consensussequenz eines funktionellen Startcodons („Kozak-Sequenz“) mit allen auf den CGGBP1-Transkripten verschiedener Spezies vorkommenden möglichen Startcodons. Positionen, die innerhalb der Consensussequenz für die Stärke des Initiationscodons entscheidend sind, sind unterstrichen. Beim Purin (R) in Position -3 handelt es sich bei funktionellen Startcodons häufiger um Adenin als um Guanosin. Die Sequenzen des Menschen und der Maus wurden den in dieser Arbeit klonierten DNA-Sequenzen entnommen (g20 und m20), während der Sequenz der Ratte der Datenbankeintrag AC128161 aus dem Rattengenomprojekt zu Grunde lag. Die Startcodons von Schwein und Rind wurden aus einer Consensussequenz, die aus mehreren EST-Klonen bestand, abgeleitet. Die Positionen der ATGs werden relativ zum eigentlichen ATG (+1) in Nukleotiden angegeben. Mit Kleinbuchstaben werden Abweichungen von der Consensussequenz dargestellt, während Großbuchstaben auf Übereinstimmungen hinweisen. Mit Sternchen versehene ORFs überlappten mit dem eigentlichen ORF des CGGBP1-Gens (Fettdruck).

Übersetzt man die Nukleotidsequenz der Minicistrons in eine Aminosäuresequenz, weist das zweite Minicistron nur in Schweinen die gleiche Länge wie beim Menschen auf (Tab. 3). Das erste Minicistron war jedoch in allen Spezies gleich lang. Die Peptidsequenz war, wie die ihr zugrundeliegende DNA- bzw. RNA-Sequenz, im ersten Minicistron in allen untersuchten Sequenzen sehr ähnlich (Abb. 22). Nur die Hälfte aller Mutationen auf DNA-Ebene führten in diesem uORF zu Aminosäureaustauschen.

Mensch	:	M	H	S	W	H	H	E	D	N	R	V	C	A	Q	K	L	F	E	H	H	H	H	I	L	K	N	*	:	25
Maus	:	M	I	S	F	H	H	E	D	N	E	V	C	A	Q	E	L	F	E	H	H	H	H	I	L	K	N	*	:	25
Ratte	:	M	I	S	F	H	R	R	D	N	R	V	C	A	Q	K	L	F	E	H	H	H	H	I	L	K	N	*	:	25
Schwein	:	M	I	S	F	H	I	R	D	D	R	V	C	A	Q	K	L	F	E	H	H	H	H	I	L	K	N	*	:	25
Rind	:	M	I	S	Q	H	I	E	E	N	E	V	C	A	Q	K	L	F	E	H	H	H	H	I	L	K	N	*	:	25
Mensch-nt	:	ATGACTTCC	GGCACCACAGGGAAAAT	CGGGTGCAAGCCCAGAACTATT	TCCCACCACCACCT	TGTTGAAAAACTGA	:	78																						
Maus-nt	:	ATGACTTCC	GGGCGCCAC	GGGAAAATAGGGTGCAAGCCCAG	AACTGTTTCCCACCACCACCT	TGTTGAAAAACTGA	:	78																						
Ratte-nt	:	ATGACTTCC	GGGCGCG	CAGGGAAAATAGGGTGCAAGCCCAGAACTG	TTTCCCACCACCACCT	TGTTGAAAAACTGA	:	78																						
Schwein-nt	:	ATGACTTCC	GGCACCTCAGGGAA	CATAGGGTGCAAGCCCAGAACTATT	TCCACACCACCCT	TGTTGAAAAACTGA	:	78																						
Rind-nt	:	ATGACTTCC	AGCACCTCAGGGAAAAT	AGGGTGCAAGCCCAGAACTG	TTTCCCAC	TACCACCCT	TGTTGAAAAACTGA	:	78																					

Abbildung 22 Sequenzvergleich der Peptid- und Nukleotidsequenzen im ersten Minicistron der CGGBP1-Gene.

Die Peptidsequenz des ersten uORFs des CGGBP1-Gens des Menschen ist in der ersten Zeile angegeben. Aminosäuren der Peptidsequenzen der vier Säugetierspezies (Zeile 2-5), die mit der Sequenz des Menschen übereinstimmt, wurden auf schwarzem Grund dargestellt. Aminosäuren, die in mehreren Spezies von der Peptidsequenz des Menschen abwichen, wurden grau schattiert. Analog dazu wurde die uORF-Nukleotidsequenz des Menschen mit der der anderen Spezies in den Zeilen 6-10 verglichen.

D.4.3.2.2 Die 3'-nicht-translatierte Region – alternative Polyadenylierung der CGGBP1-Gene

In RNA-Transfer-Experimenten wurden Transkripte der CGGBP1- und *cggbp1*-Gene von etwa 4,5 kb detektiert (Abb. 11-14; 16). Bei einem ORF von 504 bp, einer 5'-UTR von etwa 450 bp und einem Poly-A-Fortsatz von etwa 100-200 bp (Sheets et al., 1990) war die 3'-UTR somit bis zu 3,4 kb lang.

Die Sequenz stromabwärts des Stopcodons des offenen Leserasters wurde nach möglichen Polyadenylierungssignalen durchsucht. Auf ein Polyadenylierungssignal mit der Consensussequenz 5'-AAUAAA-3' folgen stromabwärts meist eine GU, bzw. U-reiche Region (Sheets et al., 1990; Chen et al., 1995), deren Sequenz nicht genauer definiert werden kann. Die von der Lage und der Sequenz wahrscheinlichsten Polyadenylierungssignale sind in der Tabelle (Tab. 4) aufgeführt.

In der EST-Datenbank waren für alle drei Polyadenylierungssignale EST-Klone verfügbar, deren Sequenz einen Poly-A-Fortsatz aufwies. In den EST-Sequenzen waren dann auch die Polyadenylierungssignale enthalten, so daß der Abstand zwischen Signal und Poly-A-Fortsatz bestimmt werden konnte. Der Abstand entsprach mit 18-23 Nukleotiden Beobachtungen bei 75 % der Vertebraten RNAs (Sheets et al., 1990).

Das Polyadenylierungssignal ab Nukleotid 14.315 der menschlichen Sequenz, bzw. ab Nukleotid 8.140 der Sequenz der Maus, entsprach der Consensussequenz 5'-AAUAAA-3' (Sheets et al., 1990). Eine Polyadenylierung an dieser Stelle führte rein rechnerisch zu einem

Transkript von 4,4 kb, wenn man das alternativ gespleißte dritte Exon des Menschen ausließ (D.3.3). Das stärkste Signal nach RNA-Transfer-Hybridisierungen lag bei etwa 4,5 kb, was einem Poly-A-Fortsatz von etwa 100 Basen gleichkam.

Die anderen Signale entsprachen einer der häufigsten, noch funktionalen Abweichung des Polyadenylierungssignals 5'-AUUAAA-3' (Sheets et al., 1990). Wenn man einen Poly-A-Fortsatz von ungefähr 100 Basen annahm, ergab sich nach dem ersten Polyadenylierungssignal mit einer 3'-UTR von nur 43 Basen ein Transkript von 1,2 kb, was ebenfalls in einer RNA-Transfer-Analyse beobachtet wurde. Da die zwei Polyadenylierungssignale, die zu der langen Form der 3'-UTR führten, nur ca. 110 Basen voneinander entfernt waren, war es nicht möglich, in RNA-Transfer-Experimenten zwischen diesen Transkripten zu unterscheiden. In der genomischen Sequenz des *cggbp1*-Gens der Ratte wurden in vergleichbaren Abständen zum ORF dieselben Polyadenylierungssignale gefunden (Tab. 4).

Während die 450 bp der 5'-UTR insgesamt Sequenzidentitäten von mindestens 78% zwischen Mensch und den anderen Spezies aufwies, war die 3'-UTR zwischen Mensch und Nagetieren nur in isolierten Regionen hoch konserviert (graue Schattierung in Abb. 17). Die kurze 3'-UTR von 43 Basen vor dem ersten Polyadenylierungssignal in der menschlichen Sequenz war bei allen Spezies, selbst bei Hühnern, zu 89 % identisch zu der entsprechenden Region des menschlichen *CGGBP1*-Gens. In der Sequenz von Hühnern konnte jedoch dort kein Polyadenylierungssignal identifiziert werden.

Polyadenylierungssignale in EST-Klonsequenzen der CGGBP1-Gene		
Position	EST-Klon	Sequenz
1. Im genomischen Klon des Menschen (AF094481)		
10928	AJ00258	AUUAAA GUUAUGUGUUGA UUGUGUGGUUCAUUUUUGUAUUUAUUUCAUUUAAA
14206	BQ030241	AUUAAAA UUUUACUAUUUCA UUUUUGUAUUCUGUGCUUAUUUUUUUUUGCUCAC
14301	BQ006755	AAUAAA UUUCAGAAAAAUUGUUU CAAAAGAUUUUGAAAGCACAUUUUUGUUUUG
2. Im genomischen Klon der Maus (AY189895)		
4789	BE135158	AUUAAA GUUAAUGUGUUGAU UGUGUGGUUCAUUUUUAUAUUUAUUUUAAAUCA
8028	BM055548	AUUAAA AUUUUUAUUUGUU AAUUUUCGUACUCCUGUGCUUCUUUUUUACUUUU
8140	BG072415	AAUAAA UUUCAAAAAAGUUUC UAAAGUUUUUGAAAUCAUAUUUUGUUUUGUGA
3. In der genomischen Sequenz der Ratte (AC128161)		
116277	BQ195795	AUUAAA GUUAAUGUGUUGA UUGUGUGGUUCAUUUUUAUAUUUAUUUCAUUUUUA
119511	AA956676	AUUAAA AUUUACUAUGUUA AUUUUCGUACUCCUGUGCUUAUUUUUUUUUUUU
119623	CA504656	AAUAAA UUGCAAAAAAAAUUCA AAAGUUUUUGAAAUCCUAUUUUUGUUUUUCA

Tabelle 4 Mögliche Polyadenylierungssignale in CGGBP1 des Menschen, der Maus und der Ratte.

In der rechten Spalte sind die genomischen Sequenzen um die möglichen Polyadenylierungssignale in RNA übersetzt aufgeführt. Neben der Position in der genannten genomischen Sequenz, wurde jeweils auch ein EST-Klon aufgeführt, dessen publizierte Sequenz eine Poly-A-Folge enthielt. Sie begann dabei an der Position der Leerstelle in der im rechten Teil aufgeführten Sequenz

D.4.4 Analyse des CGGBP1-Promotors in Transfektions-Experimenten

Die 5'-flankierende Sequenz des menschlichen CGGBP1-Gens wurde in den Promotortest-Vektor pGL2basic (Abb. 23a) kloniert, um den Promotor, also die für die Regulation der Expression verantwortlichen *cis*-aktiven DNA-Sequenzen zu lokalisieren. Die hier verwendeten Konstrukte enthielten zwischen 50 bp und 7,4 kb 5'-flankierende DNA. Die hoch konservierte, alternativ gespaltene 5'-UTR wurde ebenfalls in das Testplasmid kloniert.

Das pGL2basic-Plasmid enthält keinen eukaryontischen Promotor, so daß das Reporterogen Luciferase aus *Photinus pyralis* nur dann exprimiert wird, wenn die in das Plasmid eingebrachte Sequenz eukaryontische Promotoreigenschaften aufweist. Der Promotortest-Vektor wurde nach der Methode der Nukleofektion in HeLa-Zellen transfiziert. Das Luciferase-Protein katalysiert eine Biolumineszenz-Reaktion, die nach der Transfektion in den Zellysaten nachgewiesen und quantifiziert wurde. Über die Luciferase-Aktivität wurde indirekt die Transkriptionsrate des Gens bestimmt.

D.4.4.1 Herstellung der Reportergenkonstrukte

D.4.4.1.1 Ableitung des Transkriptionsstartpunktes

Reportergenkonstrukte sollten den Transkriptionsstart enthalten. Hier wurde als 3'-Ende der zu analysierenden genomischen Sequenz die NotI-Schnittstelle im ersten Exon des CGGBP1-Gens gewählt, die in mehr als 20 EST-Klonen vorhanden war und somit stromabwärts des Transkriptionsstartpunkts liegen mußte. Da EST-Klone von mit Poly-dT-Primern gestarteten cDNAs abstammen, sind sie aufgrund der mangelnden Prozessivität der meisten Reversen Transkriptasen am 5'-Ende häufig unvollständig. Außerdem wird bei der Produktion von EST-Klonen häufig in NotI-Schnittstellen kloniert, so daß eventuell vollständige cDNAs, die NotI-Schnittstellen enthalten, durch die Restriktion vor der Klonierung verkürzt werden. Da die meisten EST-Klone, die die NotI-Schnittstelle enthielten, auf korrekt gespleißte Transkripte zurückgingen, konnten sie nicht auf einer Kontamination durch genomische DNA beruhen.

D.4.4.1.2 Klonierung

Für die Klonierung der Promotortestplasmide wurden geeignete Restriktionsschnittstellen im genomischen Klon des CGGBP1-Gens ausgewählt und die DNA-Fragmente nach der nötigen Modifizierung der DNA-Enden in den Promotortest-Vektor pGL2bas kloniert. Die hoch konservierte 5'-UTR wurde durch RT-PCR und PCR amplifiziert, um ihren Einfluß auf die Expression des Reportergens zu testen. Die Restriktionsstrategie wird im folgenden genau beschrieben (Abb. 23).

Das Promotortestplasmid pGL2bas (Abb. 23a) wurde mit HindIII gespalten und die überhängenden Enden durch das Auffüllen mit dem Klenowfragment der DNA-Polymerase I in stumpfe Enden überführt. Um das Konstrukt c1 herzustellen, wurde das SacI-NotI-Fragment des genomischen Klons (Abb. 2) nach der Modifizierung der Enden mit dem linearisierten pGL2bas Plasmid ligiert. Das KpnI-Fragment aus dem g20-143 Plasmid (Abb. 23b) wurde in die KpnI-Schnittstelle des pGL2bas eingefügt und der Klon als KpnI-KpnI-pGL2 bezeichnet. Das NotI-KpnI-Fragment entsprach dem Exon 1, das ausgeschnitten wurde, um Konstrukt c2 zu bilden. Um das Konstrukt c6 herzustellen, wurde aus c2 das SmaI-Fragment entfernt. Für die Klonierung der zwei Spleißformen der 5'-UTR wurde der 5'-UTR mit den Primern 5'-UTR-T und 5'-UTR-B durch eine PCR amplifiziert. Geextrahierte RT-PCR-Produkte, von denen eines das alternativ gespleißte Exon III enthielt (D.3.3), dienen hierbei als Matrize. Der Primer 5'-UTR-T lag im ersten Exon des CGGBP1-Gens und

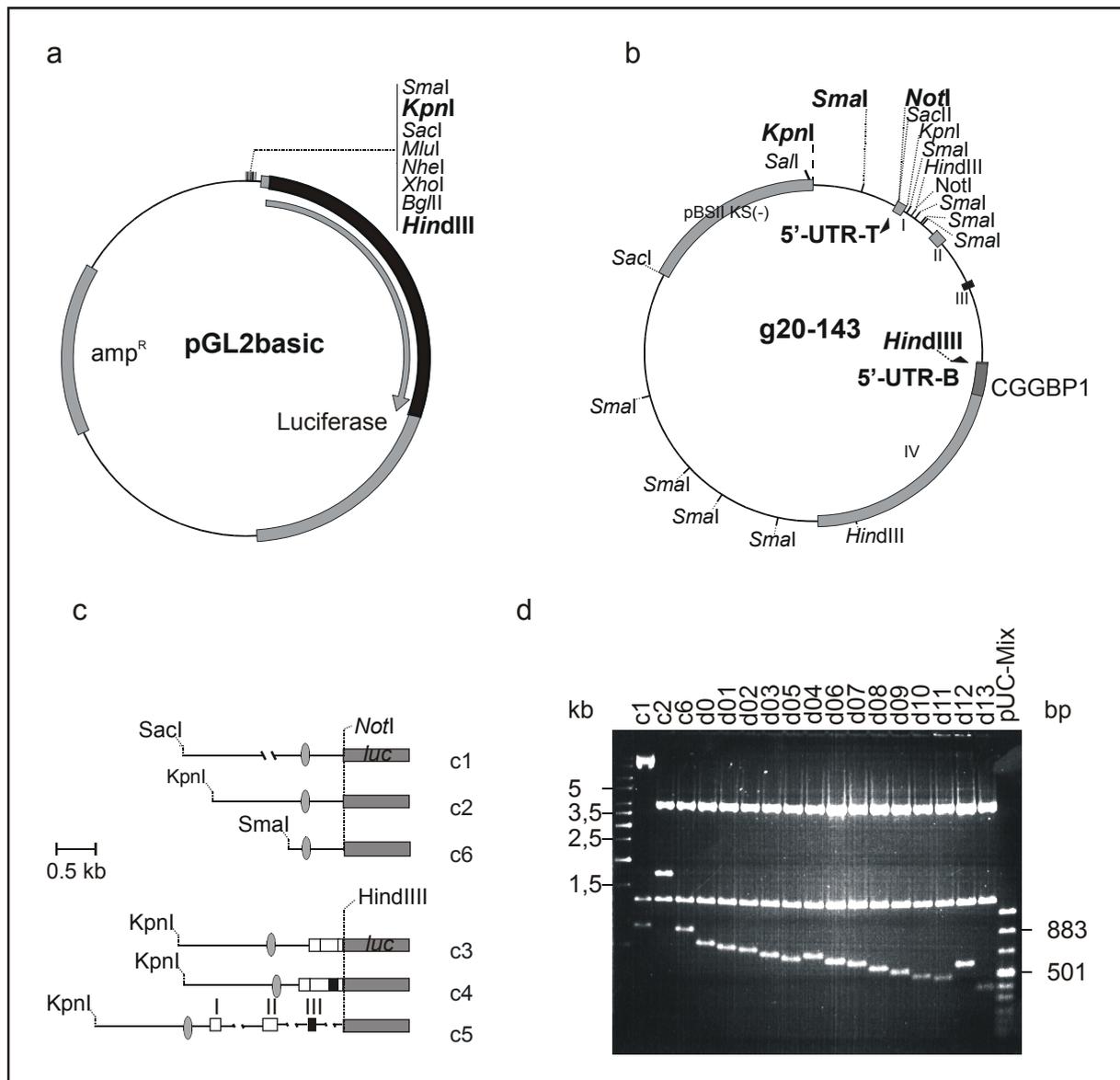


Abbildung 23 Klonierung der Reporter-genkonstrukte

Das Promotortestplasmid pGL2basic (a) wurde mit HindIII oder KpnI gespalten. In diese Schnittstellen wurden dann die im Text beschriebenen Fragmente eingefügt, die meist aus dem Plasmid g20-143 (b) stammten. Die Primer, die für die Klonierung der 5'-UTR verwendet wurden, sind durch schwarze Dreiecke in (b) gekennzeichnet. In (c) sind einige Promotortestkonstrukte mit den verwendeten Restriktionsendonukleasen aufgeführt. Das ORF des Luciferasegens (*luc*) (graues Rechteck) sowie die (GCC)₅-Wiederholung (Ellipse) aus dem CGGBP1-Promotor sind schematisch eingetragen. Die weißen Rechtecke in (c) stehen für die Exons I und II der 5'-UTR des CGGBP1-Gens, während das alternativ gespleißte Exon III schwarz dargestellt ist. In das Konstrukt c2 wurden mit der Exonuklease III fortlaufende Deletionen eingeführt. In (d) ist ein exemplarisches Agarosegel der verwendeten Konstrukte abgebildet. Die Plasmide wurden mit der Restriktionsendonuklease BglII gespalten und die DNA-Fragmente auf einem 1%iges Agarosegel elektrophoretisch getrennt. Das Konstrukt d12 hatte bei der Ligation ein zusätzliches DNA-Fragment integriert und wurde in den nachfolgenden Analysen nicht verwendet.

beinhaltete die dort vorkommende SacII-Schnittstelle (Abb. 23b). Der Primer 5'-UTR-B lag im nicht-translatierten Bereich von Exon IV und endete unmittelbar vor dem Startcodon des CGGBP1. Für Klonierungszwecke wurde in diesen Primer eine HindIII-Schnittstelle eingefügt (Abb. 23b). Die PCR-Produkte wurden zunächst in den pGEM-T-Vektor kloniert, der durch Blau-Weiß-Selektion eine Anreicherung an positiven Klonen erlaubte. Durch Sequenzierung wurden Klone identifiziert, die keine Taq-induzierten Mutationen enthielten. Anschließend wurde das SacII-HindIII Fragment ausgeschnitten und das KpnI-KpnI-Plasmid ligiert, aus dem das erste Exon des CGGBP1-Gens vorher entfernt worden war. Das Konstrukt, das die 5'-UTR ohne Exon III enthielt, wurde als c3, das mit Exon III als c4 bezeichnet. Das genomische Äquivalent der 5'-UTR wurde mittels PCR aus dem λ -Klon g20-2 (Naumann, 1998) amplifiziert. Das 3,5 kb große PCR-Produkt wurde wieder zunächst im den pGEM-T-Vektor kloniert und durch Sequenzieren auf Mutationen in den amplifizierten Exons hin untersucht. Da das genomische Fragment eine HindIII-Schnittstelle im ersten Intron des CGGBP1-Gens enthielt, erfolgte die Klonierung in zwei Schritten. Zunächst wurde das HindIII-Fragment des genomischen 5'-UTR in das mit HindIII linearisierte Kpn-KpnI-pGL2 Plasmid eingefügt. Das NotI-NotI-Fragment des resultierenden Klons wurde dann mit dem NotI-NotI-Fragment aus g20-143 ausgetauscht, so daß die im Genom vorkommende Sequenz hergestellt wurde (c5) (Abb. 23c).

Zur Feinkartierung der *cis*-aktiven DNA-Sequenzen in den 600 bp stromaufwärts des ersten Exons wurden fortlaufende Deletionen durch die Exonuklease III in das Konstrukt c2 (Abb. 23d) eingeführt, das das KpnI-NotI-Fragment aus g20-143 (Abb. 23 b) enthielt. Durch Sequenzierung wurde die genaue Position der eingeführten Deletion bestimmt und sichergestellt, daß während der Klonierung keine Mutationen eingeführt worden waren. Die in den Konstrukten d01 bis d13 enthaltenen Sequenzen sind in Abbildung 24 eingetragen.

D.4.4.2 Eigenschaften der Reporter genkonstrukte

D.4.4.2.1 Eigenschaften der Reporter genkonstrukte mit der 5'-flankierenden Sequenz des CGGBP1-Gens

Mit dem Konstrukt c1 wurden 7,4 kb der stromaufwärts des CGGBP1-Gens gelegenen genomischen Sequenz auf Promotor- oder Enhancer-Aktivität untersucht. Die Konstrukte c2 bis c5 enthielten jeweils 1,5 kb der 5'-flankierenden Sequenz des CGGBP1-Gens, im Konstrukt c6 waren 0,6 kb davon enthalten. Die 5'-flankierenden Sequenzen der CGGBP1-Gene des Menschen, der Maus und der Ratte war über 600 bp zu über 81% identisch (Abb. 24). Mit der Exonuklease III wurde die hochkonservierte Region in den Promotortestplasmiden d01 bis

d13 schrittweise verkürzt. Das erste Nukleotid stromabwärts der Deletionen wurde in dem Sequenzvergleich zwischen der Promotorregion des Menschen mit der der Maus und der Ratte eingetragen (Abb. 24).

Eine 5'-d(GCC)₅-3'-Wiederholung etwa 150 bp stromaufwärts der NotI-Schnittstelle in der menschlichen Sequenz (Abb. 24) war auch in den Sequenzen aus Maus und Ratte enthalten und im Vergleich zur menschlichen Sequenz verlängert. Die 5'-d(GCC)₉-3' der Maus und die 5'-d(GCC)₆-3' der Ratte waren jedoch jeweils einmal vom Trinukleotid 5'-d(GCT)-3' unterbrochen. Der Einfluß der Trinukleotid-Folge auf die Promotor-Aktivität wurde mit den Konstrukten d03 bis d06 genauer untersucht. Das Konstrukt d03 enthielt die Trinukleotid-Folge mit 33 nt der 5'-flankierenden Sequenz, während sie im Konstrukt d04 das 5'-Ende der untersuchten genomischen Sequenz darstellte. Im Konstrukt d05 waren nur noch drei 5'-d(CCG)-3'-Einheiten enthalten, die durch die Exonuklease III im Konstrukt d06 vollständig entfernt worden waren.

In den weiteren Konstrukten d07 bis d13 wurde der Promotor schrittweise durch die Exonuklease III jeweils um 16 bis 88 Nukleotide verkürzt. Der menschliche Promotor enthielt eine mögliche TATA-Box (Bucher et al., 1990), die in den Nagetiersequenzen jedoch nicht konserviert war. Die TATA-Box-ähnliche Sequenz 5'-GATAAA-3' befand sich 26 bp stromaufwärts von dem am weitesten stromaufwärts gelegenen EST-Klon (Zugriffsnummer BI601953) (Abb. 24). Die Positionen von Bindungsstellen von Transkriptionsfaktoren, die durch das Programm PromoterInspector (<http://www.genomatix.de/>) detektiert worden waren, sind aus Abb. 24 ersichtlich. Das Programm vergleicht die eingegebene Sequenz mit der Datenbank TransFac (Matys et al., 2003), in der experimentell bestimmte Transkriptionsfaktorbindungsstellen eingetragen sind. Alle hier eingetragenen Bindungsstellen wurden von dem Programm als signifikante Treffer bewertet. Die Vorhersage einer biologischen Funktion der identifizierten Bindungsstellen ist aufgrund der Primärsequenz eines Promotors jedoch noch nicht möglich.

Sequenzvergleich der Promotorregionen der CGGBP1-Gene

6791	CCAGGAGACGCAGAGAGGCCAGGACCCGGTGACAAGTCGGGGCGGCCA	6840	Mensch
71	CCAGGAGCCACGCAGAGGCCCGGGAGCCAGTGACAAGTCGGAGCGGCCG	120	Maus
112039	CCAGGAGCCACGCAGAGGCCAGGAGCCCGGTGACAAGTCGGAGCGGCCG	112088	Ratte
6841	ACGGACGCCTCACCTGACAGAAGCGTCGGCAGTAATCCCGACTGCTGATT	6890	
121	ACGGGCGCCTCACCTGACAGAAGCGTCGGCAGTAATCCCGGCTGCTGATC	170	
112089	ACGGGCGCCTCACCTGACAGAAGCGTCGGCAGTAATCCCGGCTGCTGATC	112138	
6891	CCGACGCGCCCGCCGCGAGTTGCCGCTGCCAGCGCCGCTCTGCCGTCGCC	6940	
171	CCGACGCGCCCGCCACAGCCGCGCCGCGCCGCCCCCTCTGCCGTCGCC	220	
112139	CCGACGCGCCCGCCAGAGCCGCTGCCGCTGGCTCCCCCTCTGCCGTCGCC	112188	
6941	CGCAGCTCTAAGCCCCACAGGGCCAGCGGGGACTCCACAATGGCCACAA	6990	
221	CGCAGCTAGGAGCCCCACTGGGCCCGGGGGGACTCCACAATGGCCACGA	270	
112189	GGCAGCTCTGAGCCCCACTGGGCCCGGGGGGACTCCACAATGGCCACGA	112238	
6991	GCTCTTCTCCGAGGCGTTCGGCCTCTTGGTCGCTCTCTTCTCCGCCATG	7040	
271	GCTCTTCTCCGAGGCGTTCGGCCTCTTGGTCGCTCTCTTCTCCGCCATG	320	
112239	GCTCTTCTCCGAGGCGTTCGGCCTCTTGGTCGCTCTCTTCTCCGCCATG	112288	
7041	AGGCTCTTCGCGCCCAgqgcGTA.CCAGCCCTgqgcCGCCCGCCCGCC	7083	
321	AGGCTCTCGCGCCCAgCGCGTACCCAGCCCTGCGCCGCGCCGCGCC	370	
112289	AGGCTCTCGCGCCCAgCGCGTACCCAGCCCTGCGCCGCGCCGCGCTGCC	112338	
7084	G.....TAGATGCCTAGGCTAACCTCCTCTCCT.CTCTAGTT	7126	
371	GCCGCTGCCCGCCGCTGCCTCAGCTAACCTCCTCCTCCTCTAGTT	420	
112339	GCCGCC.....GCTGCCTCAGCTAACCTCCTCCTCCTCCTCTAGTT	112390	
7127	CCCTCCCC.TTCCCTCCTCCCTAGccggACAGGGGCCTCTCGGCTGGGC	7175	
421	CCCTCCCCCTTTCCCTCCGCCCCAGCCGGACA.GGGCTACTCGCCTGGGC	469	
112380	CCCTCCTCCTTTCCCTCCGCCCCAGCCGGACA.GGGCCACTCCACTGGGC	112428	
7176	TCTACTTccggCTCCGCcACCCACGCCCCAGTTTTTATTGGAAGACGC	7225	
470	TCTACTTCCGGCTCCGCCACCCACGCCCCAGTTTTCACTGGAGGAAGG	519	
112429	TCTACTTCCGGCTCCGCCACCCACGCCCCAGTTTTCACTGGAAGAAGG	112478	
7226	AAGCTGCCAATCATGTTcATGCCCGCCTTCCACTACTATTGGTGAAATC	7275	
520	ACACCGCAATCATAGGCATGCCCGCCCTACC.TTCCTATTGGTTACTACT	568	
112479	AAACCGCAATCATAGGCATGCCCGCC.CTACTACTTATTGGTTAAACT	112527	
7276	GACAGTCTGTTTGATAAAccggCGTCTGTTGGTCAATGCTGCTCGGCGG	7325	
569	AGCCGCTGTTTGGCAGGCTGCCATCTGTTGGTCCACACTACTCGGCGG	618	
112528	AACCGTCTGTTTACAGCCCTGCCATCTGTTGGTCAACTACTCGGCGG	112577	

```

          .          d13          .
7326 TTGCTGGGGACCGCCTACGGGCTCTGCCATAGGCGGTGCAGGCGGACGGG 7375   Mensch
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619  TTGCTGAGGACCGCCTACGAGCCCTGCCATAGGGCGCGCCGGAGGACGGG 668   Maus
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112578 TTGCTGAGGACCGCCTACGAGCTTTGCCATAGGGCGCGCCGGAGGACGGG 112627 Ratte
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
          .          NotI          .
7376 gcgcGGCGGGGACACGGCGGCGGCGGCGGGCTCGATCGGGCAACGGCG 7425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
669  GCGCGGCAGGGGACACGGCGGCCGCGCGGAGCTCGCTCGGGCGGCGGCG 718
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112628 GCGCGGCAGGGGACACGGCGGCCGCGAGCGGGTTTCGCTCGGGCGGCGGCG 112677

7426 GCGACGGCGGCAGCG 7440
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
719  ACGGCGGTGGCAGCG 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112678 GCGGCGGCGGCGGCG 112692

```

Abbildung 24 Sequenzvergleich der Promotorregion des CGGBP1-Gens

Die 5'-flankierende Sequenz des CGGBP1-Gens des Menschen (obere Zeile) wurde mit der der Maus (mittlere Zeile) und der der Ratte (untere Zeile) verglichen. Die Zahlen beziehen sich auf die jeweilige Position in den genomischen Sequenzen, die unter den Zugriffsnummern AF094481 (Mensch), AY189895 (Maus) und AC128161 (Ratte) zugänglich sind. Ein senkrechter Strich als Zeichen für eine Übereinstimmung zwischen zwei Sequenzen bezieht sich hier immer auf eine Übereinstimmung mit der menschlichen Sequenz. Sequenzabweichungen zwischen Maus und Ratte erkennt man an einem unterschiedlichen Strichmuster. Einige Transkriptionsfaktorbindungsstellen, die durch einen Sequenzvergleich mit der TRANSFAC-Datenbank (Matys et al., 2003) durch das Programm PromoterInspector (<http://www.genomatix.de>) identifiziert worden waren, sind auf grauem Hintergrund gedruckt. Eine mögliche TATA-Box (ab nt 7289) wurde mit grauer Schrift dargestellt und fett unterstrichen. Die für die Klonierung verwendete NotI-Schnittstelle ist kursiv geschrieben und das letzte im Konstrukt enthaltene Nukleotid wird mit einer schwarzen Box gekennzeichnet. Das 5'-Ende des am meisten 5'-gelegenen EST-Klons (D.4.4.1.1) des CGGBP1-Gens ist durch ein Sternchen markiert. Die Enden der durch die Exonuklease III verkürzten Konstrukte, die in Transfektions-Experimenten verwendet wurden, sind invertiert auf dunkelgrauem Hintergrund eingetragen und mit d01, d02 etc. benannt. Alle CpG-Dinukleotide, die *in vitro* methyliert wurden (D.4.5), sind fett gedruckt. HpaII (5'-CCGG-3') und HhaI-Schnittstellen (5'-GCGC-3') wurden klein gedruckt und unterstrichen.

D.4.4.2.2 Eigenschaften der Reportergenkonstrukte mit der 5'-UTR des CGGBP1-Gens

Die Konstrukte c3 bis c5 dienen der Analyse des Einflusses der hoch konservierten 5'-UTR auf die Translation des Reportergens. In RT-PCR-Experimenten wurde festgestellt, daß die 5'-UTR des CGGBP1-Gens in HeLa-Zellen alternativ gespleißt wurde (Abb. 15). Sequenzanalysen legten außerdem nahe, daß bei der Expression des CGGBP1-Gens posttranskriptionale Regulation eine Rolle spielen könnte (E.4). Deshalb wurde die 5'-UTR des CGGBP1-Gens in drei Varianten vor das offene Leseraster der Luciferase kloniert (Abb. 23b). Im Konstrukt c3 waren die beiden ersten Exons der 5'-UTR enthalten, während das Konstrukt c4

alle drei Exons der 5'-UTR aufwies. Mit den beiden genannten Konstrukten sollte der Einfluß der alternativ gespleißten 5'-UTRs isoliert betrachtet werden können. Da in HeLa-Zellen normalerweise beide Spleißvarianten vorkamen (D.3.3; Abb. 15), wurde auch das genomische Äquivalent der 5'-UTR kloniert, aus dem die 5'-UTR dann in dem natürlich vorkommenden Verhältnis der Varianten gespleißt werden sollte.

D.4.4.3 Nukleofektion der Promotortestplasmide in HeLa-Zellen

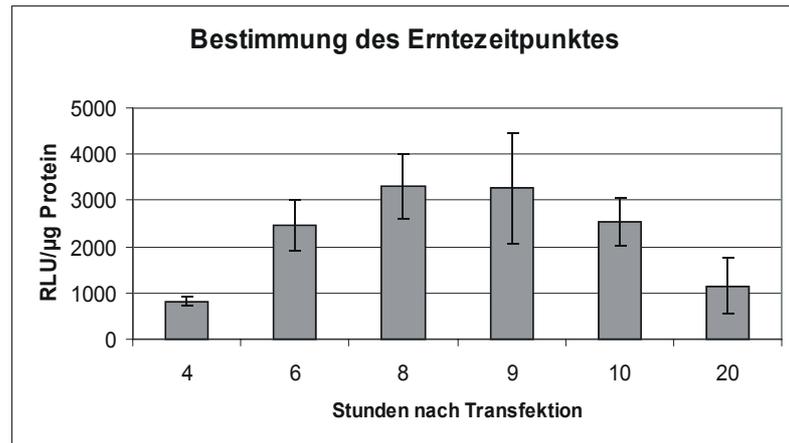
Die Plasmide wurden durch die von der Firma amaxa entwickelte Methode der Nukleofektion in HeLa-Zellen eingebracht. Dabei werden Zellen in einem für sie entwickelten Puffer aufgenommen und – ähnlich wie bei der Elektroporation (Potter et al., 1984) – kurzen elektrischen Impulsen hoher Feldstärke ausgesetzt. Die Zusammensetzung der Puffer ist nicht bekannt. Der Vorteil dieser Methode lag vor allem in einer reproduzierbar hohen Transfektionseffizienz und der schnell einsetzenden Expression der transfizierten Eigenschaft.

D.4.4.3.1 Etablierung der Transfektionsmethode der Nukleofektion

Die Transfektionseffizienz wurde hier zunächst durch eine Transfektion mit einem für das grün-fluoreszierende Protein (GFP) kodierenden Plasmid (pEGFP-C1) kontrolliert. Sie war abhängig von der Menge transfizierter DNA. Da große DNA-Mengen toxisch auf Zellen wirkten, wurden schließlich als Kompromiß zwischen Transfektionseffizienz und Toxizität 5 µg DNA pro Transfektion eingesetzt. Bei Bestrahlung der transfizierten Zellen mit Licht einer Wellenlänge von 488 nm leuchteten 24 Stunden nach der Transfektion immer zwischen 70 % und 95 % der Zellen grün. Die Kinetik der Expression des GFP war mit der der Luciferase jedoch nicht vergleichbar. Während das GFP noch bis zu 3 Tage nach der Transfektion nachweisbar war, wurde Luciferase nur wenige Stunden lang exprimiert. Um den Zeitpunkt der höchsten Aktivität zu bestimmen, wurden Zellen bis zu 24 Stunden nach der Transfektion geerntet und die Luciferase-Aktivität dann in den Zellysaten gemessen (Abb. 25). Als Zeitpunkt der Ernte eignete sich 7 Stunden nach der Transfektion, da die Luciferase zu diesem Zeitpunkt bereits stark exprimiert war.

Abbildung 25 Bestimmung des Erntezeitpunktes

Exemplarischer Verlauf der Luciferase-Aktivität nach der Transfektion. Die Aktivität der Luciferase (Relative Lichteinheiten RLU) wurde in den Lysaten von vier parallel transfizierten Ansätzen nach den angegebenen Stunden nach der Transfektion gemessen und gemittelt.



D.4.4.4 Der minimale Promotor des CGGBP1-Gens

Der minimale Promotor des CGGBP1-Gens wurde durch Transfektions-Experimente mit den oben beschriebenen Plasmiden definiert. Das Reporter-gen Luciferase war dort unter der Kontrolle von unterschiedlich langen Fragmenten der 5'-flankierenden Sequenz des CGGBP1-Gens. Für die Lokalisierung des CGGBP1-Promotors wurden 1×10^6 HeLa-Zellen mittels der Nukleofektion mit 5 µg Plasmid transfiziert. Nach 7 Stunden wurden die Zellen geerntet, lysiert und die Luciferase-Aktivität in den Lysaten gemessen (Abb. 26).

Der Einfluß der 5'-UTR auf die Reporter-gen-Aktivität wurde mit den Konstrukten c3 bis c5 untersucht. Die hoch konservierte 5'-UTR mit oder ohne Exon III wurde in die Konstrukte c4 bzw. c3 eingefügt. Im Konstrukt c5 waren außerdem die insgesamt 3,5 kb der Introns der alternativ gespleißten 5'-UTR enthalten. Die zwischen Mensch und Nagetieren konservierte Region wurde in allen weiteren Konstrukte durch die Exonuklease III schrittweise verkürzt.

Der CGGBP1-Promotor mit 1,5 kb bis 235 bp genomischer Sequenz war etwa 10-fach schwächer als der zum Vergleich transfizierte virale Promotor (SV40), der in dem Kontrollplasmid pGL2con vorhanden war (Abb. 26). Die Luciferase-Aktivität unterschied sich nur gering, wenn Plasmide mit 1,5 kb bis 235 bp humaner Sequenz in HeLa-Zellen transfiziert worden waren (c2–d08). Die Trinukleotidfolge beeinflusste die Stärke der Expression des Reporter-gens nicht. Den stärksten Einfluß auf die Reporter-gen-Aktivität übten die 30 Nukleotide aus, die dem Konstrukt d10 im Vergleich zum Konstrukt d09 fehlten. Gemäß der Analyse der Promotorsequenz enthielt diese Region zwei CCAAT-Boxen, die als Bindungsstellen für eine Reihe verschiedener Transkriptionsfaktor dienen könnten.

Die Promotor-Aktivität von kleineren Fragmenten nahm stark ab, bis sie bei einem 54 nt langen Fragment (d13) nur gering über der Hintergrund-Aktivität des pGL2bas ohne Promotor lag. Unter Berücksichtigung der unter D.4.4.1.1 beschriebenen EST-Klone enthielt dieses Konstrukt jedoch wahrscheinlich nicht den Transkriptionsstart des CGGBP1-Gens.

Die schwächere Luciferase-Aktivität bei dem Konstrukt c1 (Abb. 26) könnte auf einem Effekt der Kopienzahl oder der höheren Toxizität des größeren DNA-Moleküls beruhen. Außerdem sinkt mit größeren DNA-Molekülen die Transfektionseffizienz. Das Konstrukt c1 war mehr als doppelt so groß wie das Konstrukt c6, zudem starben mehr Zellen nach einer Transfektion des größeren Plasmids als bei den anderen Transfektionen.

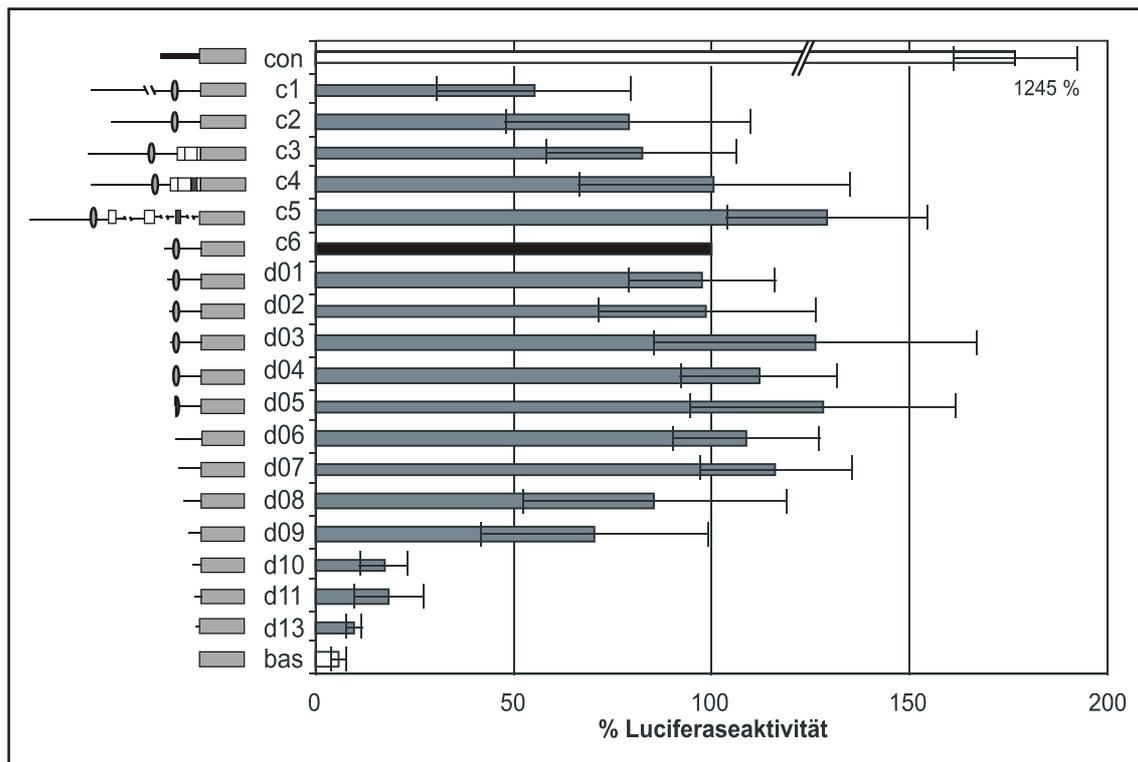


Abbildung 26 Funktionale Analyse des CGGBP1-Promotors in HeLa-Zellen

HeLa-Zellen wurden mit den auf der linken Seite schematisch abgebildeten Konstrukten wie beschrieben nukleofiziert. Vor dem offenen Leseraster der Luciferase (graues Rechteck) enthielt das Konstrukt c1 7,4 kb der 5'-flankierenden genomischen Sequenz des CGGBP1-Gens, die Konstrukte c2 bis c5 jeweils 1,35 kb. Die 5'-d(GCC)5-3'-Wiederholung ist durch eine Ellipse dargestellt. Die hoch konservierte 5'-UTR mit oder ohne Exon III wurde in die Konstrukte c4 bzw. c3 eingefügt (weiße Rechtecke). Im Konstrukt c5 waren neben den drei Exons insgesamt 3,5 kb der Introns der alternativ gespleißten 5'-UTR enthalten. Die zwischen Mensch und den Nagetieren konservierte Region wurde in allen weiteren Konstrukten durch die Exonuklease III schrittweise verkürzt. Die im Luminometer gemessenen relativen Lichteinheiten (RLU) wurden pro μg Protein und μg transfizierter DNA berechnet. Das Konstrukt c6 enthielt 0,6 kb genomische Sequenz und diente als Referenz. Zum Vergleich wurde die relative Aktivität des SV40-Promotors (con) und des Vektors ohne eukaryontischen Promotor (bas) eingetragen (weiße Balken).

D.4.5 Methylierungssensitivität des CGGBP1-Promotors

Die Konstrukte d09 und d03 (D.4.4.1.2) enthielten jeweils 18 bzw. 32 CpG-Dinukleotide. Im Konstrukt d09 waren zwei in jeweils einer HpaII und einer HhaI-Erkennungssequenz enthalten. Beim Konstrukt d03 lagen insgesamt 6 CpG-Dinukleotide in jeweils drei HpaII und drei HhaI-Erkennungssequenzen (Abb. 24 und 30). Bei einer *in-vitro*-Methylierung durch die Methylasen M-HhaI bzw. M-HpaII wurden in Anwesenheit des Methylgruppen-Donors S-Adenosyl-Methionin (SAM) nur die Desoxycytidine methyliert, die in der jeweiligen palindromischen Erkennungssequenz lagen, während die anderen CpGs unmethyliert blieben. Damit war es möglich, den Einfluß einer bzw. dreier Methylgruppen in der 5'-flankierenden Sequenz auf die Promotor-Aktivität zu testen. Die Konstrukte d03 und d09 wurden *in vitro* mit den Methylasen M-HhaI und M-HpaI methyliert. Im Falle des Konstruktes d03 wurden in der Promotorregion des CGGBP1-Gens also jeweils drei CpG-Dinukleotide methyliert, im Konstrukt d09 nur jeweils ein Dinukleotid. Nach einer Behandlung mit der SssI-Methylase waren alle vorhandenen 18 bzw. 32 CpG-Dinukleotide im Konstrukt d09 bzw. d03 methyliert.

Alle hier verwendeten Methylasen modifizierten auch alle im Vektor vorhandenen CpG-Dinukleotide, die in ihrer Erkennungssequenz lagen. Eine Restriktionsanalyse der *in-vitro*-methylierten Plasmide mit den methylierungssensitiven Restriktionsendonukleasen HhaI und HpaII gab dann Aufschluß darüber, daß die jeweiligen Erkennungssequenzen vollständig methyliert worden waren (Abb. 27). Zur SssI-Methylase gab es kein korrespondierendes Restriktionsenzym. Um einen Anhaltspunkt über das Ausmaß der Methylierung durch die

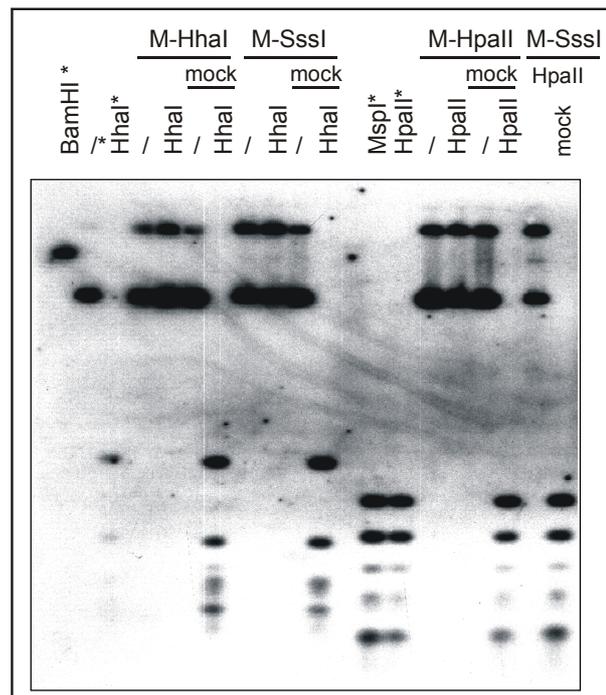


Abbildung 27 Überprüfung der *in-vitro*-Methylierung

Nach der Inkubation des Plasmids d09 mit den angegebenen Methyltransferasen und Restriktionsenzymen wurde ein Aliquot von 1 ng in einer DNA-Transfer-Analyse eingesetzt. Hier wird exemplarisch die Überprüfung der *in-vitro*-Methylierung des Konstruktes d09 gezeigt. Als Hybridisierungssonde diente das Plasmid d09. Die mit Sternchen versehenen Ansätze gingen auf unbehandeltes Plasmid zurück. Mit BamHI wurde das Konstrukt linearisiert.

SssI-Methylase zu erhalten, wurde ein SssI-methyliertes Konstrukt mit HpaII und HhaI inkubiert und nur dann eingesetzt, wenn beide Restriktionsenzyme das Plasmid nicht mehr spalten konnten. Die Restriktionen wurden anschließend mittels eines DNA-Transfer-Experimentes auf nicht methyliertes Plasmid hin untersucht (Abb. 27). In parallelen Ansätzen wurde der Methylgruppen-Donors SAM weggelassen (*mock*-Methylierung). Diese Ansätze dienten als Referenz in den Restriktionsanalysen, in denen die Vollständigkeit der Methylierung überprüft wurde. Bei der Bestimmung der Reporter-Gen-Aktivität wurden *in-vitro*-methylierte mit *mock*-methylierten Konstrukten verglichen (Abb. 28).

Zur Untersuchung der Empfindlichkeit des CGGBP1-Promotors gegenüber Methylierung wurden HeLa-Zellen schließlich jeweils mit der gleichen Menge an *in-vitro*- bzw. *mock*-methylierter DNA transfiziert. Die Aktivität der Luciferase wurde in Protein-extrakten der transfizierten Zellen gemessen, die nach 7 Stunden nach Transfektion geerntet worden waren (Abb. 28).

Der CGGBP1-Promotor wurde durch eine Methylierung aller 18 bzw. 32 CpG-Dinukleotide inaktiviert, während eine HhaI- bzw. HpaII-Methylierung die Aktivität kaum beeinflusste. Es konnte nicht ausgeschlossen werden, daß die Methylierung eines speziellen CpG-Dinukleotids den CGGBP1-Promotor nach SssI-Methylierung inaktivierte.

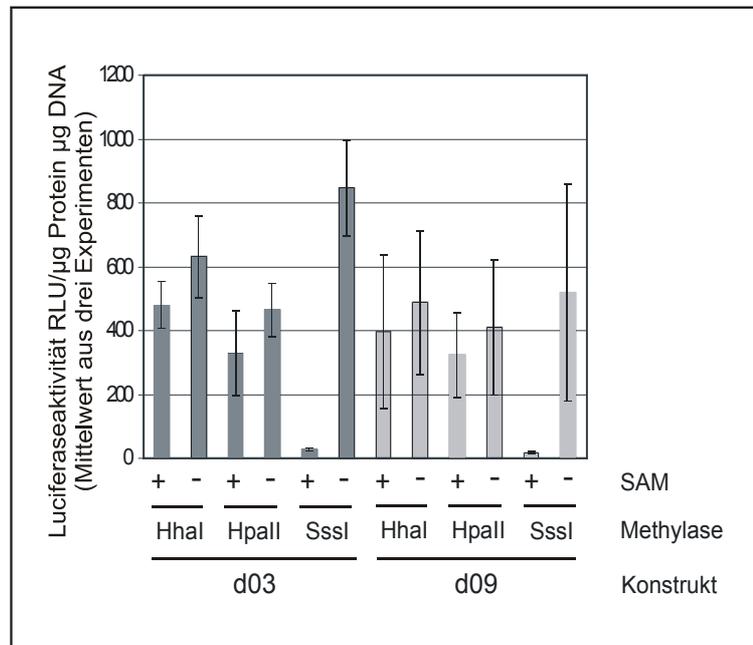


Abbildung 28 Methylierungssensitivität des CGGBP1-Promotors

Die Mittelwerte der Luciferase-Aktivität aus drei Experimenten wurden in transfizierten HeLa-Zellen bestimmt und auf µg DNA und µg Protein standardisiert. Die verwendeten Konstrukte, Methylasen und die Anwesenheit des Methylgruppen-Donors SAM wurden in das Balkendiagramm eingetragen.

D.4.6 Methylierungsstatus des CGGBP1-Gens

D.4.6.1 Lokalisierung von CpG-Dinukleotiden im genomischen Klon des CGGBP1-Gens

Das Dinukleotid CpG kommt im Genom seltener vor, als man aufgrund der Basenzusammensetzung erwarten würde (International Human Genome Sequencing Consortium, 2001). Dies drückt sich in einem durchschnittlichen CpG/GpC-Verhältnis von 0,2 aus (Antequera und Bird, 1993). Das Verhältnis liegt im Bereich um das erste Exon des CGGBP1-Gens jedoch durchschnittlich bei 0,85. Zwischen 50 und 80 % der im menschlichen Genom vorkommenden CpG-Dinukleotide sind methyliert. Eine Vorhersage des meist gewebespezifischen Methylierungsstatus ist bisher nicht möglich. Also wurde im Rahmen dieser Arbeit der Methylierungszustand der 5'-flankierenden Region des CGGBP1-Gens in zwei Zelltypen durch die Bisulfit-Methode des genomischen Sequenzierens bestimmt (Frommer et al., 1992; Clark et al., 1994).

D.4.6.2 Genomisches Sequenzieren nach der Bisulfitmethode

Ähnlich wie bei der basenspezifischen chemischen Spaltung von DNA (Church und Gilbert, 1984) wird bei der Bisulfit-Methode der Methylierungszustand aller Desoxycytidine unabhängig von der umgebenden Sequenz untersucht. Somit werden CpG-Dinukleotide sowie Desoxycytidine in anderen Nukleotidkombinationen erfaßt. Bei einer DNA-Transfer-Analyse mit methylierungssensitiven Restriktionsendonukleasen ist man bei der Detektion der Methylierung auf die in den Erkennungssequenzen vorkommenden Desoxycytidine beschränkt. Diese machen häufig nur einen geringen Prozentsatz der insgesamt vorkommenden Desoxycytidine aus.

Das genomische Sequenzieren mit Bisulfit beruht auf der katalytischen Wirkung von Bisulfit, Desoxycytidine in einzelsträngiger DNA hydrolytisch zu Desoxyuracilen zu desaminieren (Abb. 29). 5-Methyl-Desoxycytidin reagiert unter den gewählten Bedingungen nicht mit Bisulfit und bleibt somit unverändert. Die mit Bisulfit behandelte DNA wird dann in einer PCR eingesetzt, bei der das entstandene Desoxyuracil mit Desoxyadenin paart und schließlich in Desoxythymidin überschrieben wird. Die in der Sequenz verbliebenen Desoxycytidine

gehen, bis auf eine Fehlerquote von maximal 3,3 % (Zeschmig et al., 1997), auf methylierte Desoxycytidine zurück. Mit dem Sequenzieren von subklonierten PCR-Produkten wird der Methylierungszustand einzelner Allele analysiert.

D.4.6.3 Keine CpG-Methylierung im CGGBP1-Gen in primären Zellen und transformierten Zelllinien

Der Methylierungszustand der 5'-flankierenden Region und des Exon I des CGGBP1-Gens wurde in DNA aus HeLa-Zellen und peripheren weißen Blutzellen (PWBZ) mit der Bisulfit-Methode des genomischen Sequenzierens untersucht. Die im ersten Intron I vorkommenden CpG-Dinukleotide wurden ausschließlich in PWBZ analysiert. Hierzu wurde der kodierende Strang der Region um das erste Exon des CGGBP1-Gens mit vier Primerpaaren amplifiziert (Tab. 5 und Abb. 30). PCR-Produkte wurden subkloniert, und in der Regel 10 Klone pro analysierter DNA sequenziert.

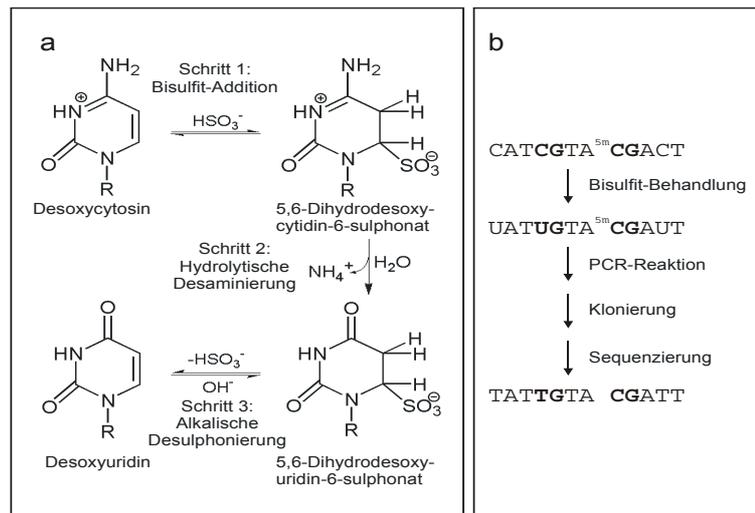


Abbildung 29 Genomisches Sequenzieren nach der Bisulfit-Methode

(a) Chemie der Bisulfit-Behandlung (modifiziert nach Hayatsu, 1976; Abb. 12) (b) Abfolge der Arbeitsschritte und Angabe der exemplarischen Nukleotidsequenz mit zwei CpG-Dinukleotiden (fett), von denen das zweite methyliert ist. Alle nicht methylierten Desoxycytidine werden in Desoxythymidine umgewandelt, 5-Methyl-Desoxycytidine erscheinen nach der PCR als als Desoxycytidine. (Abbildung übernommen aus Remus, 2000)

Primerpaar	Größe des PCR-Produktes (bp)	Bereich (nt in AF094481)	CpG n	GC-Gehalt %	Verhältnis CpG/GpC
p20 A	678	6141–6818	27	54	0,79
p20 B	530	6815–7344	49	65	0,8
p20 C1	211	7333–7543	29	69	0,93
Bestimmung des Methylierungsstatus mit p20 A–C1:			$\Sigma=102^*$		
p20 C2	584	7333–7916	68	73	0,94
Bestimmung des Methylierungsstatus mit p20 A–C2:			$\Sigma=141^*$		

Tabelle 5 Genomisches Sequenzieren nach der Bisulfit-Methode: Primer und amplifizierte Regionen

Lage der Primer und Eigenschaften der amplifizierten Regionen bei der Untersuchung der Methylierung im Promotorbereich des CGGBP1-Gens (früher als p20 bezeichnet). Die Sequenzen der Primer sind im Anhang (I.1) aufgeführt. *Da die PCR-Produkte an ihren Enden überlappten, war die Summe der insgesamt analysierten CpG-Dinukleotide kleiner, als die Summe der Anzahl der CpG-Dinukleotide in den einzelnen PCR-Fragmenten. Die 29 CpG Dinukleotide des PCR-Produktes p20 C1 sind alle auch im PCR-Produkt C2 enthalten (siehe Abb. 30)

Insgesamt wurde der Methylierungsstatus in 72 subklonierten PCR-Produkten bestimmt. In lediglich 20 Klonen wurden Desoxycytidine gefunden, die aus einem CpG-Dinukleotid der Originalsequenz stammten (Abb. 30). Die ermittelte Gesamtmethylierung lag somit bei 0,8 %. Keines der potentiell methylierten Desoxycytidine kam in zwei Klonen an der gleichen Stelle vor. Desgleichen traten 22 Desoxycytidine in den PCR-Produkten auf, die in der Originalsequenz nicht zu einem Desoxyguanosin benachbart gewesen waren (0,7 % aller Desoxycytidine). In zwei von 24 Klonen kam ein solches nicht in Desoxythymidin überschriebenes Desoxycytidin an der gleichen Stelle in einem CCC-Trinukleotid auf. Diese außerhalb eines CpG-Dinukleotids befindlichen Desoxycytidine könnten zwar auf eine in der Zelle vorhandene Methylierung hinweisen, bei den beobachteten geringen Prozentsätzen und dem Vorkommen (in der gleichen Position) in maximal zwei von 72 sequenzierten Klonen, handelt es sich wahrscheinlich um Artefakte der Methode. So könnte die Umsetzung mit Bisulfit in Einzelfällen unvollständig geblieben sein. Außerdem waren auch Mutationen denkbar, die durch die Taq-Polymerase eingeführt worden sein könnten, da sie über keine 3'-5'-Exonuklease-Aktivität verfügte. Deshalb enthielten die PCR-Produkte vereinzelt auch andere Abweichungen von der Originalsequenz.

Die 141 CpG-Dinukleotide in den Bereichen A, B, C1 und C2 waren demnach in allen Zelltypen nicht methyliert (Abb. 30). Da das Methylierungsmuster in PWBZ einer Probandin in den exemplarisch untersuchten Regionen B und C2 dem des Probanden entsprach, wurde auf eine Analyse des A-Fragmentes verzichtet. In KB-DNA wurde nur der Region C1 untersucht, in der ebenfalls alle CpG-Dinukleotide komplett unmethyliert vorlagen.

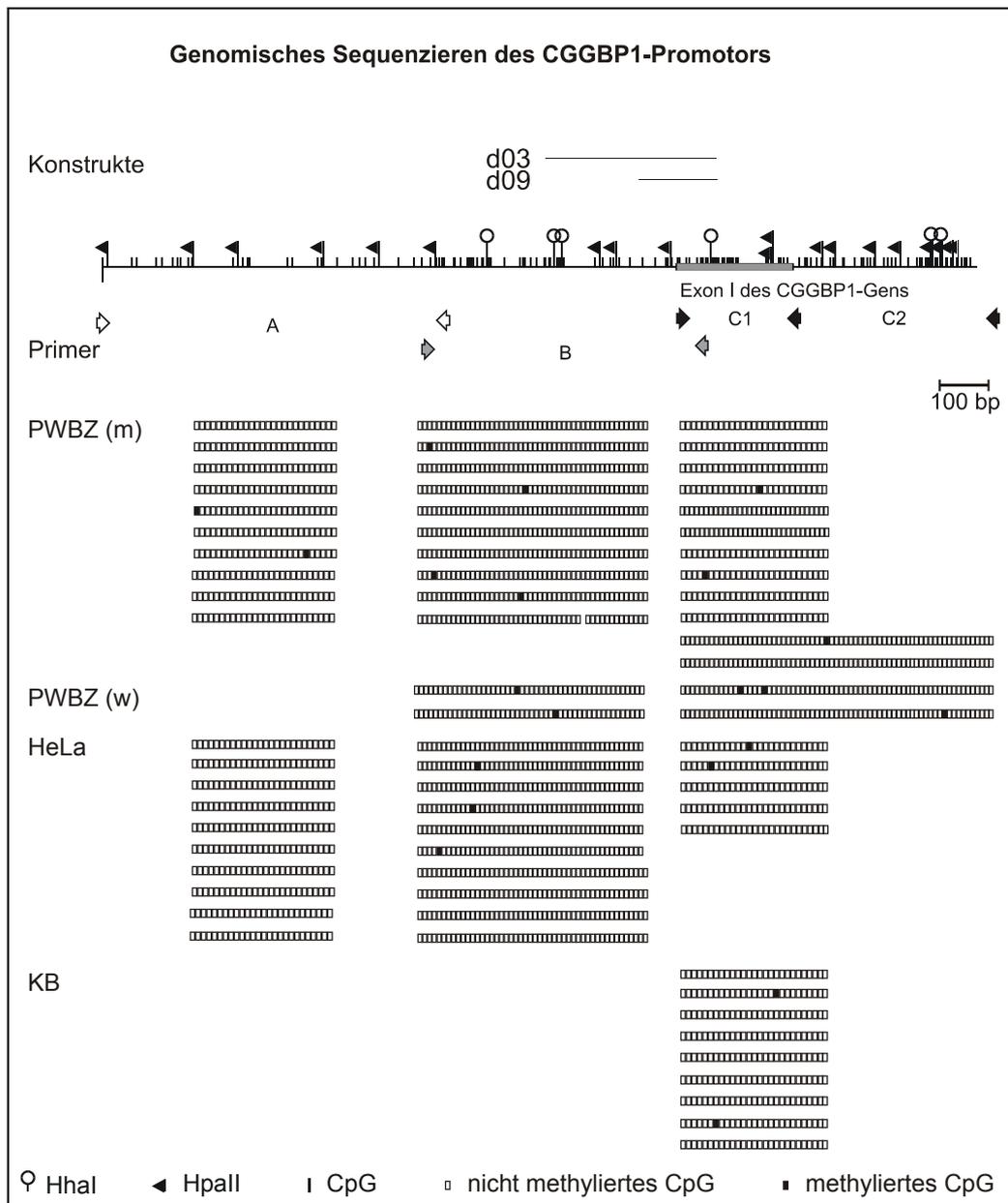


Abbildung 30 Genomisches Sequenzieren nach der Bisulfit-Methode: Karte des CGGBP1-Promotors

In der schematischen Karte wurde die Verteilung der CpG-Dinukleotide in der untersuchten Region eingetragen und ihre Position in Erkennungssequenzen von methylierungs-sensitiven Restriktionsendonukleasen HhaI und HpaII angegeben (Legende der verwendeten Symbole in der Abbildung). Die Lage der Konstrukte, die in Transfektions-Experimenten verwendet worden waren (D.4.5), wurden durch Striche über der Karte gekennzeichnet. Pfeile weisen auf die Position der verwendeten Primer hin (Tab. 5) Die Quellen der analysierten DNA wurde links in der Abbildung genannt: periphere weiße Blutzellen (PWBZ) eines Probanden (m) und einer Probandin (w) sowie die transformierte menschlichen Zelllinien HeLa und KB. Die Rechtecke stellen jeweils ein CpG-Dinukleotid dar, das sich nach der PCR als unmethyliert (leeres Rechteck) oder methyliert (schwarz) erwiesen hatte. Die Blöcke aus Rechtecken wurden unterhalb der verwendeten Primerpaare angeordnet.

D.5 Gezielte Geninaktivierung des *cggbp1* Gens

Im Rahmen dieser Arbeit sollte eine Strategie zur Inaktivierung des *cggbp1*-Gens entwickelt werden. Der Geninaktivierungs-Vektor wurde hier fertiggestellt, und mit der Transfektion von embryonalen Stammzellen der Maus wurde begonnen. Die technischen Grundlagen von Geninaktivierungs-Experimenten wurden hier auch unter Berücksichtigung der geplanten Tierexperimente vorgestellt, da sie für die Entwicklung der Strategie von großer Bedeutung waren.

D.5.1 Technische Grundlagen

D.5.1.1 Konventionelle Geninaktivierung

Die gezielte Inaktivierung eines Gens in der Maus kann zur Aufklärung seiner Funktion beitragen. Diese 1987 entwickelte Methode beruht auf homologer Rekombination, bei der ein Gen durch einen sogenannten Geninaktivierungs-Vektor ausgetauscht wird (Thomas und Capecchi, 1987). Der Geninaktivierungs-Vektor enthält dabei das Zielgen, das *in vitro* ausgeschaltet oder anderweitig modifiziert worden ist. Um die Voraussetzungen für einen sequenz-spezifischen Austausch zu schaffen, muß der Geninaktivierungs-Vektor zusätzlich mehr als 1,3 kb unveränderte Sequenz vom Genlocus des Zielgens enthalten (Hasty et al., 1991).

Nach einer Transfektion in pluripotente embryonale Stammzellen (ES-Zellen) (Evans und Kaufman, 1981; C.1.3.3) müssen diejenigen seltenen Klone identifiziert werden, die das Zielgen gegen den Geninaktivierungs-Vektor durch homologe Rekombination ausgetauscht haben. Mit Hilfe einer Positiv/Negativ-Selektion werden homologe Rekombinanten angereichert (Mansour et al., 1988). Bei der positiven Selektion überleben nur Zellen, die den Vektor aufgenommen und ins Genom integriert haben, da im Vektor das entsprechende Resistenzgen vorliegt. Der Selektionsmarker im Geninaktivierungs-Vektor muß auf beiden Seiten von genomischer DNA flankiert sein, die die homologe Rekombination in das Genom vermitteln kann.

Die Mehrheit der Klone, die die Positivselektion überleben, haben den Vektor jedoch durch illegitime Rekombination zufällig in das Genom integriert. Da für eine illegitime Rekombination keine hohe Sequenzähnlichkeit zwischen Integrationsort und der integrierten DNA bestehen muß, geschieht es häufig, daß der Geninaktivierungs-Vektor dabei samt der Plasmidsequenz in das Genom aufgenommen wird. Diese Beobachtung liefert die Grundlage für eine Anreicherung homologer Rekombinanten durch eine Negativselektion. Dafür trägt

der Geninaktivierungs-Vektor einen negativen Selektionsmarker, meist die Thymidinkinase (TK) des Herpes simplex-Virus. Transfizierte Zellen werden unter selektionierendem Agens kultiviert, der Tod derjenigen Zellen vermittelt, die den negativen Selektionsmarker an einer zufälligen Stelle in ihr Genom integriert haben. Bei Anwesenheit des TK-Gens kann das Guanodin-Analogon Ganciclovir zum Monophosphat phosphoryliert werden. Zelleigene Enzyme wandeln es in ein Triphosphat um, welches anstelle von Guanodin bei der Replikation in die DNA eingebaut wird, durch seine azyklische Seitenkette aber dann zum Abbruch der Polymerase-Reaktion führt. Für eine zufällige Integration darf der negative Selektionsmarker nicht auf beiden Seiten von genomischer DNA flankiert werden. Der Geninaktivierungs-Vektor wird also mit einer geeigneten Restriktionsendonuklease so linearisiert, daß der negative Selektionsmarker nur auf einer Seite an endogene DNA des Zielgens grenzt (Abb 34e). Somit ist er von einer homologen Rekombination ausgeschlossen, da diese nur bei hoher Sequenzähnlichkeit der zu rekombinierenden DNA-Moleküle stattfindet.

Homologe Rekombinanten sind doppelt-resistent und überleben beide Selektionen. Die genetische Veränderung der Klone wird durch geeignete molekularbiologische Methoden wie Southern-Transfer-Analysen oder PCR nachgewiesen. Zur Generation von Mäusen, in denen das Zielgen ausgeschaltet worden ist (*knock-out*-Maus) werden zunächst doppeltresistente, undifferenzierte ES-Zellen in Wildtyp-Blastozysten injiziert, die anschließend in eine scheinwangere Maus implantiert werden. Bei den entstehenden Mäusen handelt es sich um Chimären, deren Gewebe sich teilweise von Wildtyp- und teilweise von manipulierten ES-Zellen ableiten. Um eine transgene Mauslinie zu erhalten, die die eingeführte Modifikation gemäß den Mendelschen Regeln an ihre Nachkommen weitergibt, werden die Chimären mit Wildtyp-Mäusen verpaart. Bei geeigneter Wahl des Mausstamms, aus denen die Empfängerblastozysten entnommen worden waren, und des Mausstamms der manipulierten ES-Zellen kann man z. B. anhand der Fellfarbe leicht jene Nachkommen identifizieren, die von den manipulierten ES-Zellen abstammen. Man spricht dann von einer Keimbahntransmission des Geninaktivierungs-Vektors (Schwartzberg et al., 1989).

Die Folgen des Verlustes des Zielgens können dann im Zusammenhang des gesamten Organismus im hetero- sowie homozygoten Zustand studiert werden. Das Fehlen oder die Mutation des Zielgens kommt in homozygoten Tieren vom Zeitpunkt der Befruchtung an zum Tragen und ist während der gesamten Ontogenese in allen Zellen der Mutanten ausgeprägt.

D.5.1.2 Konditionale Geninaktivierung

Unter Umständen kann der Verlust eines Gens nach der in D.5.1.1 beschriebenen Methode zu starken Entwicklungsstörungen bis zum Absterben des Embryos führen (z. B. Brandon, et al., 1995a-c). Die konditionale Geninaktivierung ermöglicht es, die *in vitro* eingeführte Mutation auf bestimmten Zelltypen und/oder Zeitpunkte während der Entwicklung zu beschränken (zur Übersicht Gossen und Bujard, 2002). Sie beruht auf dem Cre/LoxP-Rekombinationssystem aus dem Bakteriophagen P1 (Gu et al., 1994; Kühn et al., 1995). Die 38 kDa große Cre-Rekombinase vermittelt die basenpaargenaue Rekombination zwischen zwei LoxP-Zielsequenzen (Sternberg und Hamilton, 1981; Sauer und Henderson, 1989; zur Übersicht Kilby et al., 1993). Bei tandem-artiger Orientierung der 34 bp langen, asymmetrischen Zielsequenz wird das DNA-Fragment, das von den Zielsequenzen flankiert wird, ausgeschnitten, bei entgegengesetzter Orientierung wird es invertiert.

Für die konditionale Geninaktivierung in der Maus wird eine funktionell wichtige Region des Zielgens im Geninaktivierungs-Vektor mit LoxP-Zielsequenzen flankiert. Nach der oben beschriebenen Transfektion und Selektion der transfizierten ES-Zellen muß kontrolliert werden, ob diese LoxP-Zielsequenzen beim Austausch des Zielgens zusammen mit dem Geninaktivierungs-Vektor in das Genom übertragen wurden.

Selektionsmarker besitzen häufig einen eigenen Promotor, der das Expressionsmuster des Zielgens und benachbarter Gene beeinflussen kann (Pham et al., 1996). Daher wird der positive Selektionsmarker ebenfalls mit zwei LoxP-Sequenzen versehen. Die Cre-Rekombinase wird dann in ES-Zellen transient exprimiert, um den positiven Selektionsmarker vor der Injektion doppelt-resistenter Klone in Blastozysten zu entfernen (Gu et al., 1993). Mit Klonen, in denen die Cre-Rekombinase neben dem Selektionsmarker auch das Zielgen deletiert, kann mit einem konditionalen Geninaktivierungs-Vektor das Zielgen auch konventionell (D.5.1.1) ausgeschaltet werden. Als Alternative zu dieser *in-vitro*-Deletion kreuzt man Mäuse, die den Geninaktivierungs-Vektor aufgenommen haben, auch mit der „Deleter“-Maus, die die Cre-Rekombinase bereits in Keimzellen exprimiert (Schwenk et al., 1995).

Eine große Auswahl an Cre-transgenen Mäusen erlaubt die zell- oder gewebespezifische und/oder induzierte Expression der Cre-Rekombinase (Nagy und Mar, 2001). Aus einer Kreuzung entstehen Nachkommen, in denen die Rekombination zwischen den LoxP-Sequenzen auf die Zelltypen beschränkt ist, in denen das Transgen Cre ausgeprägt wird (Kühn et al., 1995; Kühn und Schwenk, 1997).

Nach der Rekombination durch die Cre-Rekombinase verbleiben am modifizierten Genlocus nur die 34 bp einer LoxP-Sequenz. Durch die konditionale Mutagenese eines Gens kann die Funktion eines Gens also in einem bestimmten Zelltyp oder ab einem bestimmten Zeitpunkt untersucht werden (zur Übersicht Lewandoski, 2001).

D.5.2 Grundlegende Überlegungen zur gezielten Geninaktivierung des *cggbp1*-Gens

D.5.2.1 Art der Geninaktivierung

Für die gezielte Inaktivierung des *cggbp1*-Gens war die konventionelle sowie die konditionale Geninaktivierung denkbar. Bei der hohen Konservierung des *cggbp1*-Gens in Säugetieren und Vögeln bestand jedoch die Möglichkeit, daß das *cggbp1*-Gen eine grundlegende Rolle, z. B. während der Entwicklung, spielte. Dafür sprach auch die Expression von *cggbp1* in ES-Zellen (Abb. 11) und während der Embryonalentwicklung der Maus (Abb. 16). Deshalb wurde hier die konditionale Strategie der Geninaktivierung gewählt, um im Falle eines letalen Phänotyps, der nach einer konventionellen *cggbp1*-Inaktivierung hätte auftreten können, die Effekte eines auf bestimmte Zelltypen beschränkten Funktionsverlustes analysieren zu können.

D.5.2.2 Art der Mutation

In Geninaktivierungs-Experimenten muß sichergestellt werden, daß die eingeführte Mutation zu einem vollständigen Funktionsverlust des Zielgens führt. Im Falle des *cggbp1*-Gens war es möglich, das gesamte offene Leseraster, weil es nicht durch Introns unterbrochen war, zu entfernen. So war keine Hintergrund-Aktivität des Gens mehr zu erwarten. Bei der konditionellen Geninaktivierung muß gleichzeitig die normale Funktion des Gens in Anwesenheit der LoxP-Zielsequenzen gewährleistet sein. Wie unten beschrieben, lag die 5'-liegende LoxP-Sequenz 0,4 kb stromaufwärts des ORF im zweiten Intron des *cggbp1*-Gens. Die 3'-liegende LoxP-Sequenz wurde 0,2 kb stromabwärts des ersten Polyadenylierungssignals eingefügt. Diese beiden Regionen wurden gewählt, da die der Insertion benachbarten Sequenzen zwischen Maus und Mensch nicht konserviert und somit möglicherweise nicht von funktio-neller Bedeutung für die Expression des Gens sind.

D.5.3 Strategie der Geninaktivierung des *cggbp1*-Gens

Als Strategie eines Geninaktivierungs-Experiments bezeichnet man die Abstimmung des technischen Nachweises von homologen Rekombinanten auf die in das Zielgen einzuführenden Modifikationen. So muß das endogene Allel von dem modifizierten Allel in homologen Rekombinanten unterschieden werden können. Entscheidend für den Entwurf des Geninaktivierungs-Vektors ist dabei die Methode, mit der homologe Rekombinanten identifiziert werden sollen. Nach der basengenauen Rekombination durch die Cre-Rekombinase können in Abhängigkeit der Reihenfolge und der Vollständigkeit der Rekombination verschiedene Allele entstehen, die ebenfalls voneinander unterschieden werden müssen. Die Strategie beruht also auf der Kenntnis des zu verändernden Genortes und bezieht verschiedene Aspekte des Geninaktivierungs-Experimentes mit ein.

D.5.3.1 Wahl der Screening-Methode für die gezielte Inaktivierung des *cggbp1*-Gens

Für die Inaktivierung des *cggbp1*-Gens sollten transfizierte ES-Zellklone durch eine Southern-Transfer-Analyse überprüft werden. Dabei fallen erfahrungsgemäß weniger falsch-negative Ergebnisse an, als bei der auch üblichen PCR-Methode (Manolis Pasparakis, persönliche Mitteilung). Somit steigt die Wahrscheinlichkeit, daß homologe Rekombinanten entdeckt werden. Vor der Konstruktion des Geninaktivierungs-Vektors müssen geeignete Restriktionsendonukleasen identifiziert werden. Nach einer Restriktion genomischer DNA müssen in einer DNA-Transfer-Hybridisierung alle möglichen Allele – das Wildtypallel, das Allel homologer Rekombinanten und alle durch die Cre-Rekombinase veränderten Allele – voneinander unterschieden werden können.

D.5.3.2 Nachweis von Rekombinanten im *cggbp1*-Genlocus

Die Strategie alle auftretenden Allele nachzuweisen, basierte auf einer Restriktionskarte (Abb 31a). Gemäß der Strategie veränderte sich der Genlocus nach dem Austausch des endogenen *cggbp1*-Gens mit dem Geninaktivierungs-Vektor (Abb. 31b). Der positive Selektionsmarker für die Geneticinresistenz (*neo*) und das offene Leseraster des *cggbp1*-Gens sollten durch je zwei LoxP-Sequenzen flankiert werden. Im Ausgangsplasmid für die Klonierung (*easyflox*) wurde dies so gelöst, daß stromabwärts vom positiven Selektionsmarker zwei LoxP-Sequenzen vorlagen, denen jeweils für die Klonierung von DNA geeignete Schnittstellen folgten (Abb. 34a). Es wurden im Geninaktivierungs-Vektor für das *cggbp1*-Gen insgesamt drei LoxP-Sequenzen (A-C) verwendet, von denen zwei den positiven Selektionsmarker

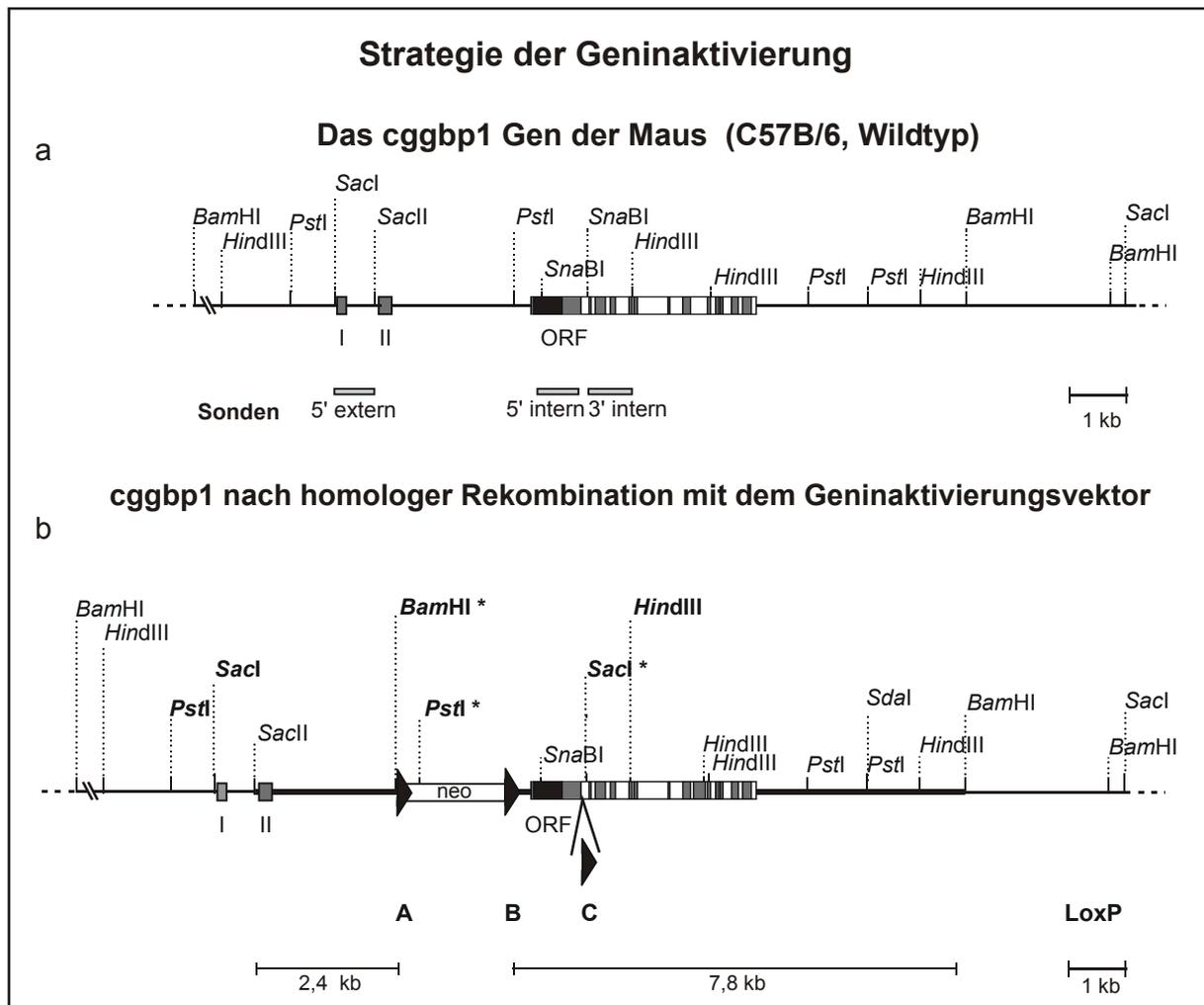


Abbildung 31 Strategie der Geninaktivierung

Schematische Darstellung der Restriktionskarte des Wildtyp-Allels des *cggbp1*-Gens in C57BL/6-Mäusen vor (a) und nach (b) der homologen Rekombination mit dem Gen-Inaktivierungs-Vektor. Die ORF des *cggbp1*-Gens wurde durch schwarze Rechtecke und die drei LoxP-Sequenzen A, B und C durch Dreiecke gekennzeichnet. Die grauen Boxen in der 5'- und 3'-UTR des *cggbp1*-Gens zeigen hohe Sequenzidentität zum menschlichen Gen an. Der Selektionsmarker der Geneticinresistenz wurde als weißes Rechteck (neo) dargestellt. Die Genstruktur des *cggbp1*-Gens wurde mit den für das Geninaktivierungs-Experiment relevanten Restriktionsschnittstellen angegeben. Die gesamte, dem chromosomalen Gen entsprechende Region des Geninaktivierungs-Vektors ist an der dicken Linie in der unteren Karte erkennbar. Den gezielt veränderten Genlokus kennzeichnen neu eingeführte Restriktionsschnittstellen, die mit einem Stern markiert wurden. Diejenigen Restriktionsschnittstellen, mit denen der Austausch mit dem Geninaktivierungs-Vektor nachgewiesen werden sollte, sind fett gedruckt. Die dritte LoxP-Sequenz wurde in eine weniger stark konservierte Region eingefügt.

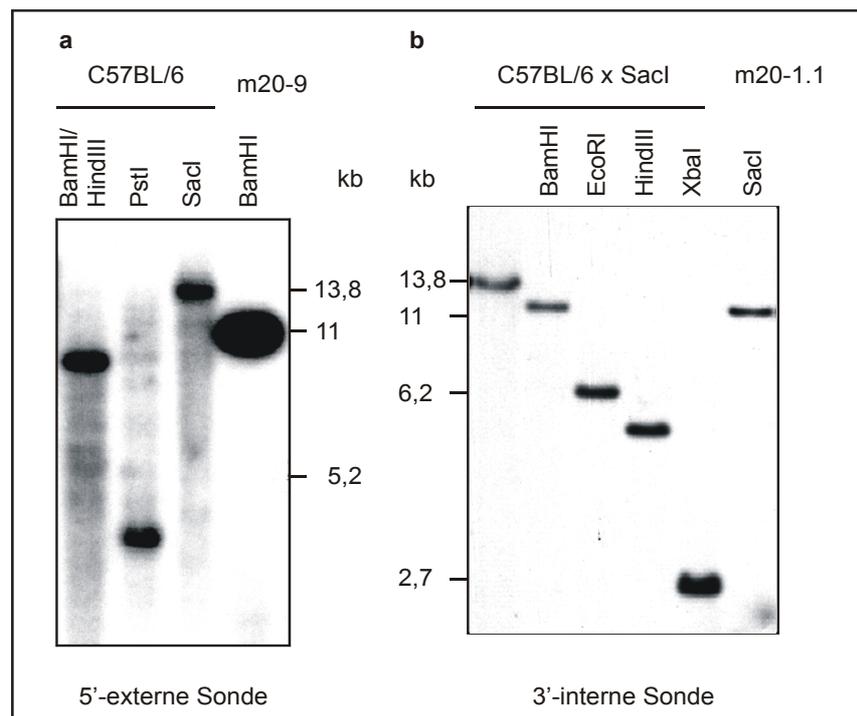
flankierten (A und B) (Abb. 31). Die mittlere LoxP-Sequenz (B) lag dann zwischen dem Selektionsmarker und dem offenen Leseraster. Stromabwärts des offenen Leserasters befand sich eine dritte LoxP-Sequenz (C), auf deren Anwesenheit durch die oben beschriebene

Positiv-Negativselektion nicht selektioniert wurde. Deshalb mußte ihre Anwesenheit nach der Identifizierung von doppelt-resistenten Klonen gesondert überprüft werden (D.5.3.3.1).

Der Austausch des endogenen *cggbp1*-Gens gegen den Geninaktivierungs-Vektor sollte unter Verwendung von drei Sonden nachgewiesen werden. Um homologe Rekombinanten von Klonen zu unterscheiden, die den Geninaktivierungs-Vektor zufällig ins Genom integriert hatten, wurde eine Sonde gewählt, die außerhalb der vom Geninaktivierungs-Vektor abgedeckten Region lag (5'-externe Sonde). Die resultierenden Allele aus einer sequenzspezifischen Rekombination durch die Cre-Rekombinase wurden mit den innerhalb des Geninaktivierungs-Vektors gelegenen 5'- und 3'-internen Sonden nachgewiesen. Mit allen Sonden konnte die Anwesenheit der LoxP-Sequenz C überprüft werden. Die Tauglichkeit der Sonden wurde im Vorfeld auf genomischer DNA von Wildtyp-Mäusen getestet (Abb. 32). Bei der 5'-internen Sonde handelt es sich um das ORF des *CGGBP1*-Gens, mit dem das *cggbp1*-Gen kloniert worden war (Abb. 5). Somit war bereits bekannt, daß die Sonde keine repetitiven Elemente enthielt.

Abbildung 32 Southern-Transfer-Analyse mit den Testsonden.

Genomische DNA von C57BL/6-Wildtyp-Mäusen wurde mit den angegebenen Restriktionsendonukleasen gespalten. Die Fragmente wurden in einem 0,8%igen Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Die DNA in (a) wurde mit der 5'-externen Sonde, die DNA in (b) mit der 3'-internen Sonde hybridisiert.



D.5.3.3 Detektion von homologen Rekombinanten nach der Transfektion von ES-Zellen

Die im Wildtypallel im zweiten Intron des CGGBP1-Gens vorkommende PstI-Schnittstelle wurde im Geninaktivierungs-Vektor durch eine BamHI-Schnittstelle ersetzt. Aufgrund des veränderten Restriktionsmusters kann dann der Austausch des endogenen *cggbp1*-Gens überprüft werden. Nach Hybridisierung mit der stromaufwärts liegenden 5'-externen Sonde würde im rekombinanten Allel im Vergleich zum Wildtyp-Allel ein größeres Fragment nachgewiesen werden (Tab. 6). Da das NotI-SacII-Fragment der 5'-externen Sonde außerhalb des Geninaktivierungs-Vektors lag, würden zufällig ins Genom integrierte Geninaktivierungs-Vektoren nicht gefunden werden. Alternativ zu einer Restriktion mit PstI dienten auch Restriktionen mit BamHI und SacI dem Nachweis von homologen Rekombinanten, da sich auch hier die Fragmentgrößen nach einer Hybridisierung mit der 5'-externe Sonde deutlich unterschieden (Tab. 6). Bei einer Restriktion mit SacI konnte zusätzlich die Anwesenheit der LoxP-Sequenz C überprüft werden, da diese im Geninaktivierungs-Vektor mit einer SacI-Schnittstelle versehen worden war.

Nachweis für homologe Rekombinanten mit der 5'-externen Sonde		
Restriktionsenzym	Gezielt verändertes Allel	Wildtyp-Allel
PstI	4,1 kb	3,6 kb
SacI	6,5 bk	13,8 kb
BamHI	17 bk	>23 kb

Tabelle 6 Nachweis für homologe Rekombinanten mit der 5'-externen Sonde.

Größen der in DNA-Transfer-Experimenten abgebildeten Fragmente des Allels nach homologer Rekombination des Geninaktivierungs-Vektors (gezielt verändert) und des Wildtyp-Allels.

D.5.3.3.1 Detektion der LoxP-Sequenz C

In einem gezielt veränderten *cggbp1*-Allel befinden sich drei LoxP-Sequenzen in tandem-artiger Orientierung. Da zwei davon den positiven Selektionsmarker flankieren, kann bei Vorliegen der Geneticinresistenz davon ausgegangen werden, daß die LoxP-Sequenzen A und B vorhanden sind. Die dritte LoxP-Sequenz lag 1,3 kb stromabwärts des Selektionsmarkers der Geneticinresistenz. Bei der homologen Rekombination des Geninaktivierungs-Vektors mit dem endogenen Gen könnte die LoxP-Sequenz durch ein stromaufwärts aufgelöstes *cross-over* ausgeschlossen werden. Es wurde stromaufwärts der LoxP-Sequenz C eine SacI-Schnittstelle eingeführt, um die Anwesenheit dieser LoxP-Sequenz überprüfen zu können. Eine DNA-Transfer-Analyse von homologen Rekombinanten nach einer Restriktion mit SacI ergäbe ein mit allen drei Sonden vom Wildtyp klar unterscheidbares Bandenmuster (Tab. 7).

Nachweis für LoxP C mit allen Sonden		
Restriktionsenzym	Gezielt verändertes Allel	Wildtyp-Allel
SacI	6,5 kb	13,8 kb

Tabelle 7 Nachweis für LoxP C mit allen Sonden

Größe des Wildtyp-Allels (wt) und des Allels nach homologer Rekombination des Geninaktivierungs-Vektors (gezielt verändert) nach DNA-Transfer-Hybridisierung mit jeder der drei Testsonden.

D.5.3.3.2 Klassifizierung der Allele nach der Rekombination durch die Cre-Rekombinase

Nach der Rekombination durch die Cre-Rekombinase müssen drei Allele voneinander unterschieden werden können: das unveränderte, mit LoxP-Sequenzen versehene Allel (neo⁺, floxed) und die Allele, in denen die LoxP-Sequenz A entweder mit der LoxP-Sequenz B oder mit C rekombiniert. Die Rekombination A-B führt zum alleinigen Verlust des Selektionsmarkers der Geneticinresistenz (Δ neo). Bei der Rekombination A-C werden das Neo-Resistenzgen und das LoxP-flankierte Zielgen gleichzeitig entfernt (Δ total). In beiden Fällen werden rekombinante ES-Zellklone anhand ihrer Empfindlichkeit gegenüber Neomycin erkannt. Ein Zellklon, in dem die LoxP-Sequenz B mit der LoxP-Sequenz C rekombiniert (Δ orf), wird aufgrund der weiterhin vorhandenen Neomycinresistenz erst gar nicht untersucht. Die 3'-interne Sonde dient der Analyse der möglichen resultierenden Allele nach der Rekombination durch die Cre-Rekombinase. Eine Restriktion mit PstI oder BamHI und HindIII führt dabei jeweils zu DNA-Fragmenten verschiedener Größe (Tab. 8).

Nachweis der Allele nach Rekombination durch Cre mit der 3'-internen Sonde		
Restriktionsenzym	Gezielt verändertes Allel	Größe (kb)
PstI (wt 5,1 kb)	Unverändertes Allel mit LoxP	6,6
	Δ neo	8,7
	Δ total	7,5
	(Δ ORF	5,3)
BamHI/HindIII (wt 7 kb)	Unverändertes Allel mit LoxP	4
	Δ neo	2
	Δ total	5,7
	(Δ ORF	2,7)

Tabelle 8 Nachweis der Allele nach Rekombination durch Cre mit der 3'-internen Sonde.

Größen der mit der 3'-internen Sonde in Southern-Transfer-Experimenten abgebildeten Fragmente des Allels nach homologer Rekombination des Geninaktivierungs-Vektors und des Wildtyp-Allels (wt).

D.5.3.4 Anwendbarkeit der Strategie im Maustamm 129

Die entwickelte Strategie sollte auch in ES-Zellen des 129 Mausstamms angewendet werden können, da ES-Zellen des 129 Stamms einfacher zu handhaben sind als die des C57BL/6-Stamms. Um Polymorphismen in den strategisch wichtigen Schnittstellen zu detektieren, wurde genomische DNA aus Mäusen von zwei 129-Substämmen mit den für die Geninaktivierungsstrategie entscheidenden Restriktionsendonukleasen gespalten und mit den verschiedenen Sonden hybridisiert (Abb. 33). Bis auf einen Polymorphismus in einer SacI-Schnittstelle waren die Fragmentgrößen zwischen C57BL/6- und 129-Mäusen identisch. Da das polymorphe Allel von 129/SvJ-Mäusen größer war als das von C57BL/6 und der Nachweis der dritten LoxP-Sequenz auf einer Verkleinerung eines SacI-Fragmentes beruhte, konnte die entwickelte Strategie auch in Mäusen des 129Sv-Stamms angewendet werden.

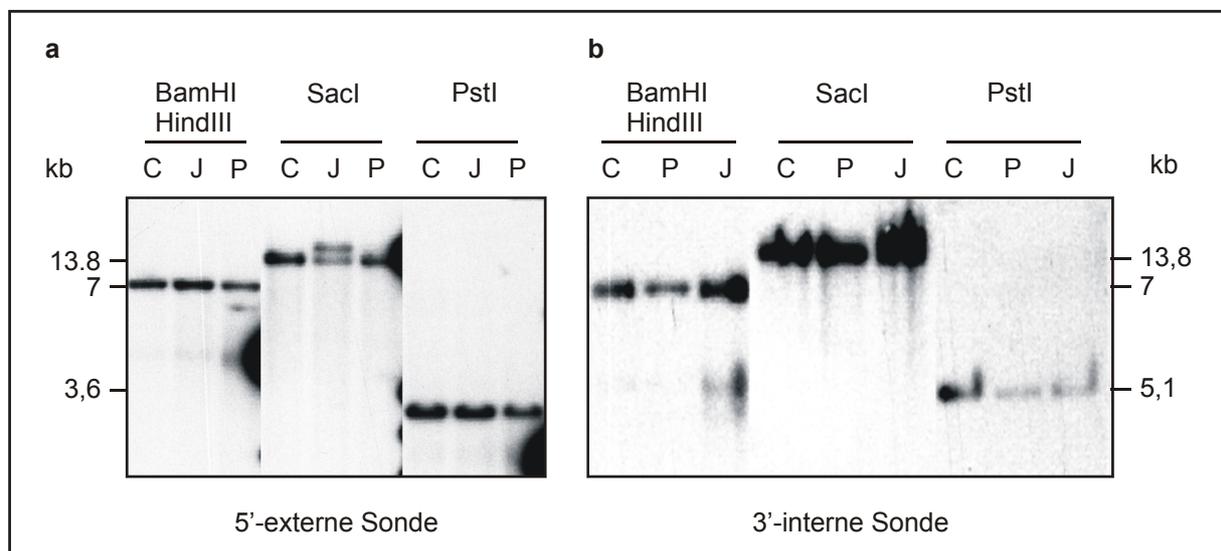


Abbildung 33 DNA-Transfer-Hybridisierung von DNA des Mausstamms 129Sv.

Genomische DNA von Wildtyp-Mäusen der Stämme C57BL/6 (C), 129Sv-Pasteur (P) und 129SvJ (J) wurden mit den angegebenen Restriktionsendonukleasen gespalten, auf einem 0,8%igen Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Anschließend wurde die DNA mit der 5'-externen (a) oder der 3'-internen Sonde (b) hybridisiert.

D.5.4 Klonierung des Geninaktivierungs-Vektors

D.5.4.1 Der Ausgangsplasmid easyflox

Für die Klonierung des Geninaktivierungs-Vektors wurde der Vektor easyflox (Manolis Pasparakis) verwendet. Er enthielt bereits die drei LoxP-Sequenzen A, B und C in gleicher Orientierung (Abb. 34a). Die Geneticinresistenz (Neomycinphosphotransferasegen) unter der Kontrolle des Phosphoglycerat-Promotors wurde von den LoxP-Sequenzen A und B flankiert. Der 3' vom Selektionsmarker gelegenen LoxP-Zielsequenz B folgte nach zwei für die Klonierung geeigneten Schnittstellen XbaI und Sall die LoxP-Sequenz C. Stromabwärts dieser LoxP-Sequenz ermöglichten die Schnittstellen HindIII und XhoI die Insertion eines weiteren DNA-Fragmentes. Das Plasmid kodierte auch für den negativen Selektionsmarker der Thymidinkinase (tk) aus Herpes simplex und die Ampicillinresistenz (amp-r) für die Vermehrung des Plasmids in Bakterien.

D.5.4.2 Klonierung des offenen Leserasters des cggbp1-Gens in easyflox (EF-I)

Das PstI-SnaBI-Fragment des cggbp1-Gens, das das offene Leseraster des cggbp1-Gens enthielt und gemäß der entwickelten Strategie von LoxP-Sequenzen flankiert werden sollte, wurde durch PCR amplifiziert. Der 3'-gelegene Primer enthielt dabei eine partielle SacI-Schnittstelle 5'-AGCTC-3'. Nach der Spaltung des easyflox-Vektors mit Sall (5'-G'TCGAC-3') und XbaI (5'-T'CTAGA-3') wurden die entstandenen Überhänge durch die S1-Nuklease abgebaut. Bei korrekter Orientierung des klonierten PCR-Produktes wurde die partielle SacI-Schnittstelle des Primers im Plasmid EF-I (Abb. 34b) durch das verbliebene Desoxyguanosin der Sall-Schnittstelle vervollständigt (5'-GAGCTC-3').

D.5.4.3 Klonierung des SacII/PstI-Fragmentes in EF-I (EF-II und III)

Die Effizienz der homologen Rekombination steigt bei zunehmender Länge der zur Verfügung stehenden chromosomalen DNA im Geninaktivierungs-Vektor (Thomas und Capecchi, 1987; Deng und Capecchi, 1992; Hasty et al., 1991). Der Geninaktivierungs-Vektor für das cggbp1-Gen enthielt insgesamt 10 kb genomische Maus-DNA. Ist der positive Selektionsmarker von den DNA-Enden des für die Transfektion linearisierten Geninaktivierungs-Vektors nicht weit genug entfernt, kann dies den Anteil an illegitimer Rekombination erhöhen (Thomas et al., 1992; Berinstein et al., 1992). Deshalb wurden 2,4 kb genomischer DNA zwischen dem DNA-Ende des Vektors und dem 5'-Ende des positiven Selektionsmarker eingefügt (Abb. 34c). Dafür wurde zunächst das PstI-Fragment aus dem

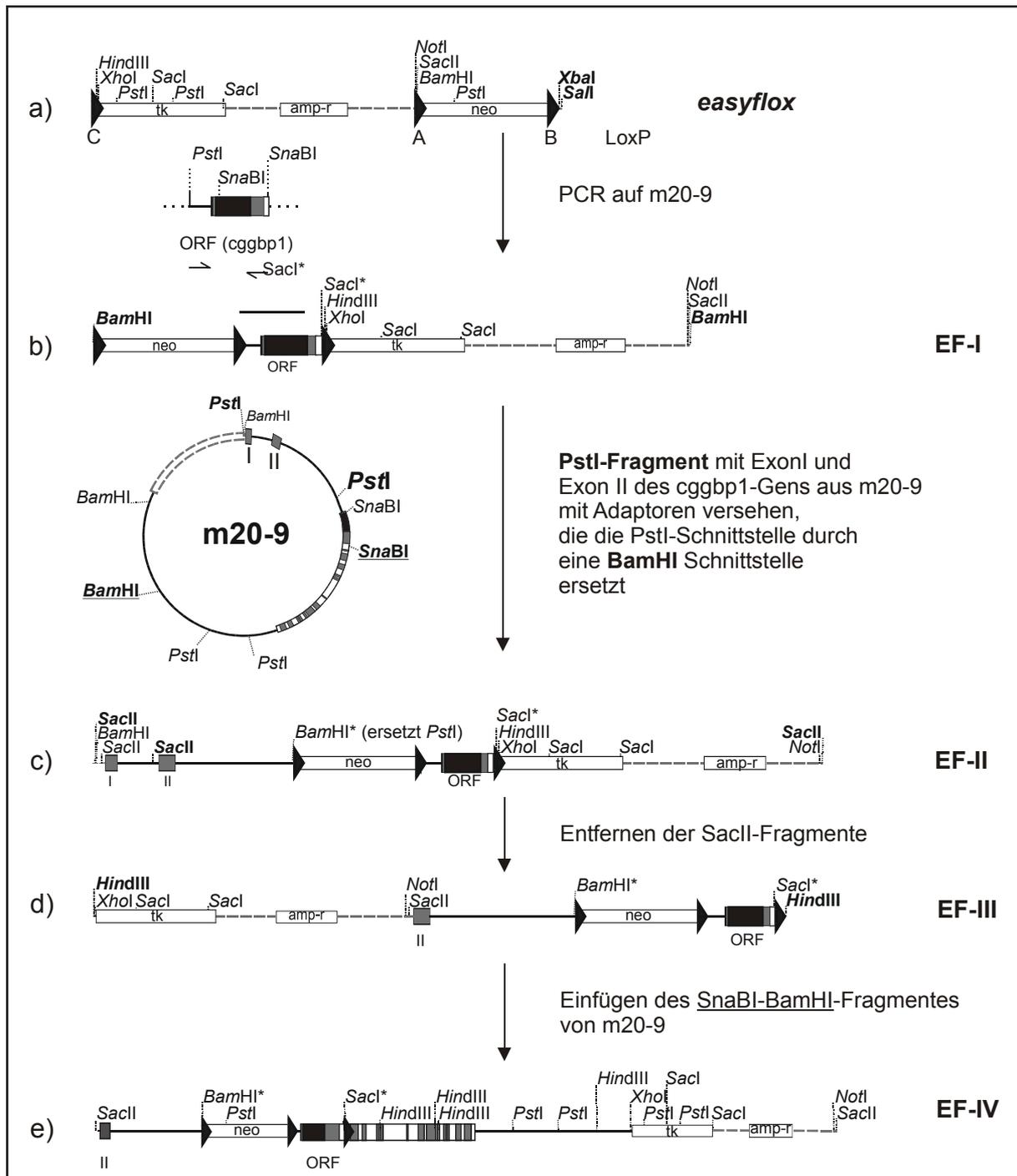


Abbildung 34 Klonierung des Gen-Inaktivierungs-Vektors.

Schematische Darstellung der Klonierung des Geninaktivierungs-Vektors. Die einzelnen Komponenten wurden wie in Abb. dargestellt. (a) Das Ausgangsplasmid *easyfloX* (gestrichelte, graue Linie), in dem drei *LoxP*-Sequenzen und alle benötigten Selektionsmarker (*neo* und *tk* und *amp-r*) enthalten waren. (b)–(e) Die vier Klonierungsschritte für die Konstruktion des Geninaktivierungs-Vektors. Dabei wurden die Positionen der Primer (horizontale Pfeile) für die Amplifikation des offenen Leserasters des *cggbp1*-Gens sowie die verwendeten Fragmente des genomischen Klon (m20-9) angegeben. Die jeweiligen Produkte sind hier so dargestellt, als ob sie mit dem Enzym gespalten worden sind (Fettdruck), in dessen Schnittstelle das neben dem Pfeil (Ligation) aufgeführte Fragment eingefügt, bzw. dessen Fragment entfernt wird.

Plasmid m20-9 über Adaptoren, die die PstI-Schnittstelle zerstörten und eine BamHI-Schnittstelle enthielten, in einen mit BamHI vorbereiteten pBS II KS(-) ligiert. Mit diesem Vektor konnten positive Klone durch eine Blau-Weiß-Selektion angereichert werden. Da die BamHI und die SacII-Schnittstelle im easyflox-Vektor so eng benachbart waren, daß der Vektor nicht mit beiden Enzymen gleichzeitig gespalten werden konnte, wurde zunächst das gesamte BamHI-Fragment in die BamHI-Schnittstelle von EF-I kloniert, wobei das Plasmid EF-II entstand. Anschließend wurde das SacII-Fragment entfernt und das Plasmid als EF-III bezeichnet (Abb. 34d).

D.5.4.4 Klonierung des SnaBI/BamHI-Fragmentes in EF-III (EF-IV)

Schließlich wurde der BamHI-Überhang des SnaBI/BamHI-Fragmentes des genomischen Klons des *cggbp1*-Gens mit Hilfe des Klenow-Fragments der DNA-Polymerase I aufgefüllt und mit der ebenfalls in stumpfe Enden überführten HindIII-Schnittstelle des EF-III-Plasmids ligiert (Abb. 34e). Das resultierende EF-IV-Plasmid konnte mit SacII oder NotI linearisiert werden. Im linearisierten Plasmid war der negative Selektionsmarker der Thymidinkinase nun von beiden Enden her vor einem exonukleolytischen Abbau geschützt. Das resultierende EF-IV-Plasmid wurde zur Kontrolle der Klonierung mit verschiedenen Restriktionsenzymen gespalten und auf ein Agarosegel aufgetragen (Abb. 35). Die LoxP-Sequenzen und die enthaltenen Exons wurden durch Sequenzierung auf Mutationen überprüft. Anschließend wurde CsCl-gereinigte DNA des EF-IV-Plasmids für die Transfektion vorbereitet (C.1.3.3.1). Die Vollständigkeit der Linearisierung mit dem Enzym SacII wurde durch eine Gelelektrophorese kontrolliert.

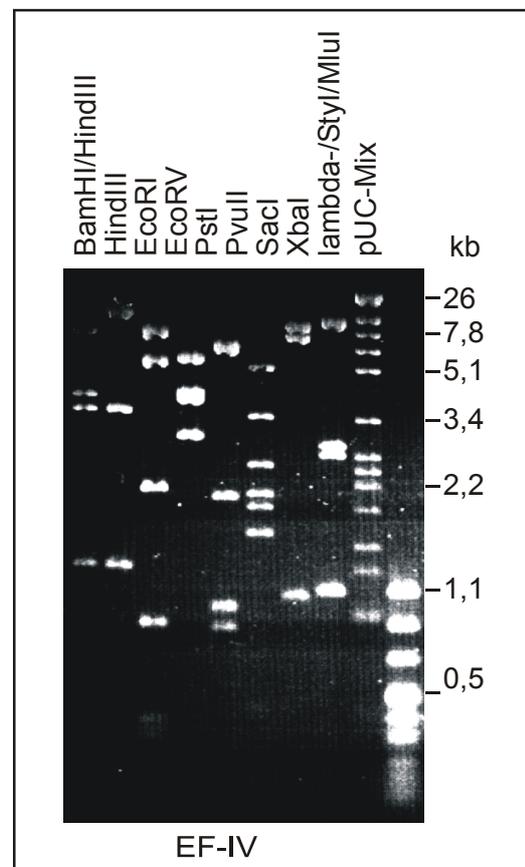


Abbildung 35 Restriktionsanalyse des Geninaktivierungs-Vektors EF-IV.

Jeweils 0,5 µg des EF-IV-Plasmids wurden mit den angegebenen Restriktionsenzymen gespalten und die Fragmente wurden in einem 0,8%igen Agarosegel elektrophoretisch getrennt.

D.5.5 Transfektion des LoxP-flankierten *cggbp1*-Gens in ES-Zellen

D.5.5.1 Transfektion in C57BL/6 ES-Zellen

In zwei Ansätzen wurden je 30 µg linearisierter Geninaktivierungs-Vektor in 1×10^7 ES-Zellen des Mausstamms C57BL/6 durch Elektroporation eingebracht. Etwa 2% der Zellen überlebten die Elektroporation und nur 54 Klone die Doppelselektion mit Geneticin und Ganciclovir. Nach der Positivselektion mit Geneticin wurden insgesamt 54 ES-Zellkolonien daraufhin untersucht, ob sie das *in vitro* modifizierte *cggbp1*-Gen durch homologe Rekombination ausgetauscht hatten. Die Negativselektion mit Ganciclovir führte zu keiner Anreicherung der Geneticin-resistenten ES-Zellkolonien. Keiner der einer Analyse zugänglichen Klone hatte das endogene *cggbp1*-Gen mit dem des Geninaktivierungs-Vektors ausgetauscht (Abb. 36).

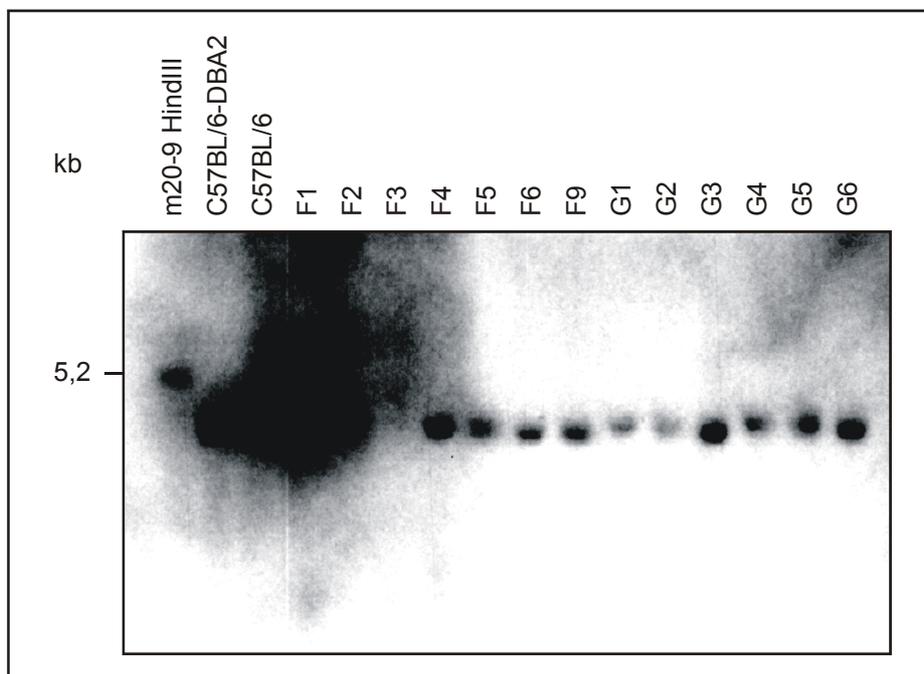


Abbildung 36 DNA-Transfer-Hybridisierung mit DNA aus transfizierten ES-Zellen des C57BL/6-Mausstamms.

Genomische DNA von doppelt-resistenten ES-Zellkolonien wurde mit PstI gespalten, die Fragmente wurden in einem 0,8%igen Agarosegel elektrophoretisch getrennt und auf eine Membran übertragen. Die Abbildung zeigt exemplarisch eine Hybridisierung der DNAs von 13 ES-Klonen mit der 5'-externen Sonde. Sie wurden gemäß ihrer Position in 96-Loch-Platten benannt.

D.5.5.2 Transfektion in 129-ES-Zellen (Ulrich Müller, Friedrich Miescher Institut, Basel)

Da Zellen des Mausstamms 129Sv leichter zu handhaben sind und das Labor von Ulrich Müller mehr Erfahrung in der Transfektion von ES-Zellen hatte, wurde in einer Kollaboration mit seiner Arbeitsgruppe der linearisierte Geninaktivierungs-Vektor zweimal in ES-Zellen des Mausstamms 129Sv eingebracht. Auch hier überlebte nur ein sehr geringer Prozentsatz der transfizierten Zellen die Doppelsektion. Vor der Analyse differenzierten zudem alle ES-Zellkolonien, so daß sie auch nicht auf homologe Rekombinanten hin untersucht wurden.

D.5.6 Modifizierung des Geninaktivierungs-Vektors

Die wiederholt gemachte Beobachtung in zwei verschiedenen Laboren, daß nur sehr wenige ES-Zellklone die Doppelsektion überlebten, könnte dadurch begründet sein, daß der Geninaktivierungs-Vektor ausschließlich zufällig ins Mausgenom integrierte. Da eine Resistenz gegen die Selektion durch Ganciclovir auch auf einem Funktionsverlust der Thymidinkinase (z.B. durch Punktmutationen) vor oder während eines Integrationsprozesses beruhen könnte, müssen doppelt-resistente Klone nicht auf homologe Rekombinanten zurückgehen.

Die Sequenz des SnaBI-BamHI-Fragmentes war zum Zeitpunkt der Klonierung des Geninaktivierungs-Vektors nicht vollständig bekannt. Es stellte sich nach der Bestimmung der Sequenz heraus, daß 1,65 kb am 3'-Ende des langen Armes durchgehend aus LINE- und SINE-Elementen bestand (Abb. 37). Diese repetitiven Elemente machen mehr als 20% des Mausgenoms aus (Mouse Genome Sequencing Consortium, 2002). Es war denkbar, daß diese ausgedehnte Region von hochrepetitiver DNA die homologe Rekombination mit dem *cggbp1*-Locus beeinträchtigte, d. h. zur homologen Rekombination an anderen, nicht relevanten Loci geführt haben könnte. Anhand der bestimmten Sequenz des gesamten genomischen Klons m20-9 wurde das Enzym SdaI identifiziert, das das EF-IV-Plasmid linearisierte. Nach einer zusätzlichen Restriktion mit der ebenfalls nur einmal stromaufwärts der Thymidinkinase schneidenden Restriktionsendonuklease XhoI wurde ein 1,7 kb DNA-Fragment mit den ununterbrochenen repetitiven Sequenzen aus dem Geninaktivierungs-Vektor entfernt (Abb. 37). Die verbleibenden repetitiven Elemente im Geninaktivierungs-Vektor EF-V waren maximal 260 bp lang.

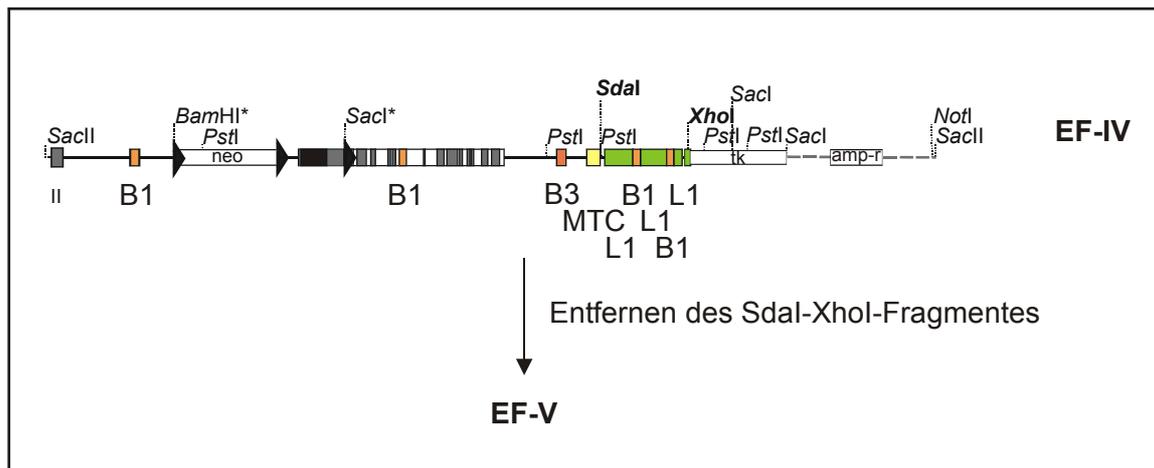


Abbildung 37 Modifizierung des Geninaktivierungs-Vektors.

Der Geninaktivierungs-Vektor mit allen funktionellen Elementen (LoxP, neo, tk, amp-r) und der vollständigen Genstruktur des *cggbp1*-Gens. Das offene Leseraster des *cggbp1*-Gens ist als schwarzes Rechteck gekennzeichnet. Graue Rechtecke in den UTR weisen auf eine Sequenzidentität von mindestens 78% zwischen dem CGGBP1- Gen des Menschen und der Maus hin. Enthaltene repetitive Elemente sind analog zur Abb. 10 eingetragen und klassifiziert.

Der modifizierte Vektor EFIV-77 wurde mit verschiedenen Restriktionsendonukleasen gespalten und auf ein Agarosegel aufgetragen (Abb. 38). Mutationen in allen LoxP-Sequenzen und Exons wurden durch Sequenzierung ausgeschlossen. Die Transfektion des modifizierten Geninaktivierungs-Vektors in ES-Zellen steht aufgrund des Umzugs der Arbeitsgruppe Müller im Januar 2003 an das Scripps Research Institut in San Diego noch aus.

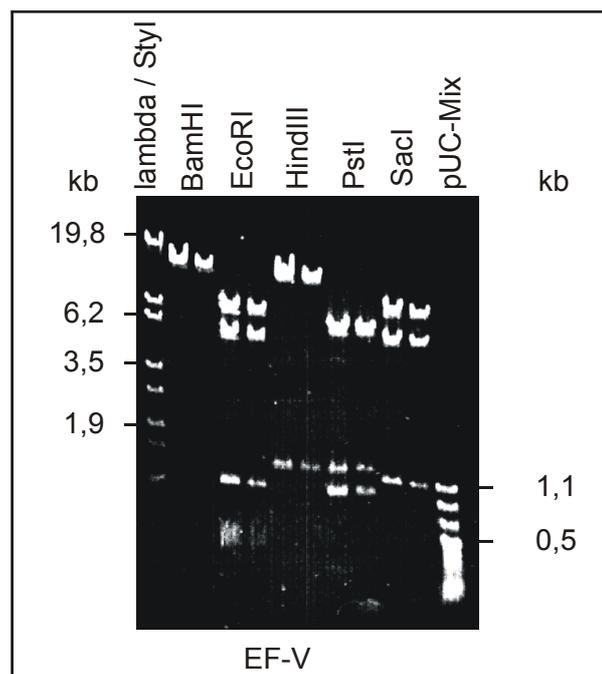


Abbildung 38 Restriktionsanalyse des Geninaktivierungs-Vektors EF-V.

Jeweils 0,5 μ g zweier Klone des EF-V-Plasmids wurden mit den angegebenen Restriktionsendonukleasen gespalten und die Fragmente wurden in einem 0,8%igen Agarosegel elektrophoretisch getrennt.

D.6 Charakterisierung der Bindungssequenz von CGGBP1

D.6.1 Bindung an 5'-d(CGG)_n-3' (n ≤ 8)

Die im menschlichen Genom vorkommenden 5'-d(CGG)_n-3'-Wiederholungen sind häufig kürzer als n = 8 (Tóth et al., 2000). Bei der bisherigen Charakterisierung der Bindungssequenz von CGGBP1 wurden meist Oligodesoxyribonukleotide eingesetzt (Tab. 1), die reine 5'-d(CGG)_n-3'-Wiederholungen mit n ≥ 8 enthielten (Deissler et al., 1996 und 1997). Die 5'-d(CGG)_n-3'-Einheiten waren in den Oligodesoxyribonukleotiden ohne flankierende Sequenz eingesetzt worden, weshalb die DNA-Fragmente mit weniger 5'-d(CGG)-3'-Wiederholungen gleichsam an Länge verloren. Deshalb bestand die Möglichkeit, daß CGGBP1 auch an kürzere Wiederholungen binden konnte, dafür aber flankierende Sequenz benötigte. So band CGGBP1 sehr gut an die strukturelle 28S rRNA Sequenz ab Nukleotid 2119, in der eine 5'-d(CGG)₇-3'-Wiederholung auf beiden Seiten von nicht 5'-d(CGG)_n-3'-haltiger DNA flankiert wurde (Müller-Hartmann et al., 2000; Tab. 1). An ein Oligodesoxyribonukleotid, in dem die vom FraXF-Locus stammende 5'-d(CGG)₈-3'-Wiederholung am 3'-Ende des Moleküls lag, band CGGBP1 dagegen nicht (Deissler et al., 1996).

Name	Sequenz eines Stranges 5'-3'
FMR1-5	AGG GGG CGT GCG GCA GCG (CGG)₅ CTG GGC CTC GAG CGC CCG
FMR1-8	CGT GCG GCAGCG (CGG)₈ CTG GGC CTC GAG CGC
FMR1-12	GCG GCA GCG (CGG)₁₂ CTG GGC
mfmr1	GGG GGG CGT GCG GTA ACG (CGG)₆ CGA (CGG)₂ CTG GGC
CGGBP1-5	CAG CGC GTA CCA GCC CCT GC (GCC)₅ GTA GAT GCC TAG GCT A
mcggbp1	CCA GCC CCT GC (GCC)₆ GCT (GCC)₃ GCT GCCTCA G
(CGG) ₁₇	(CGG)₁₇

Tabelle 9 Sequenz der in EMSA verwendeten doppelsträngigen Oligodesoxyribonukleotide

Es wurden die Bezeichnung der Oligodesoxyribonukleotide und die Sequenz des oberen Stranges angegeben. Die CGG-Wiederholung ist fett gedruckt.

Es wurden hier natürlich vorkommende kürzere Varianten von 5'-d(CGG)-3'-Wiederholungen als Sonde in Gelretardations-Experimenten (Abb. 40) oder in Kompetitions-Experimenten (Abb. 39) eingesetzt, um den Einfluß von flankierender Sequenz auf die Komplexbildung zu testen. Die 5'-d(CGG)_n-3'-Wiederholung im 5'-UTR des FMR1-Gens des Menschen wird

mindestens fünfmal wiederholt (Maddalena et al., 2001). Deshalb wurden hier 51 bp lange Oligodesoxyribonukleotide ausgewählt, deren zentrale 5'-d(CGG)-3'-Folge aus 5, 8 oder 12 Einheiten bestand (FMR1-5, FMR1-8 und FMR1-12) und von genomischer DNA des menschlichen FMR1-Gens flankiert wurde (Tab. 9).

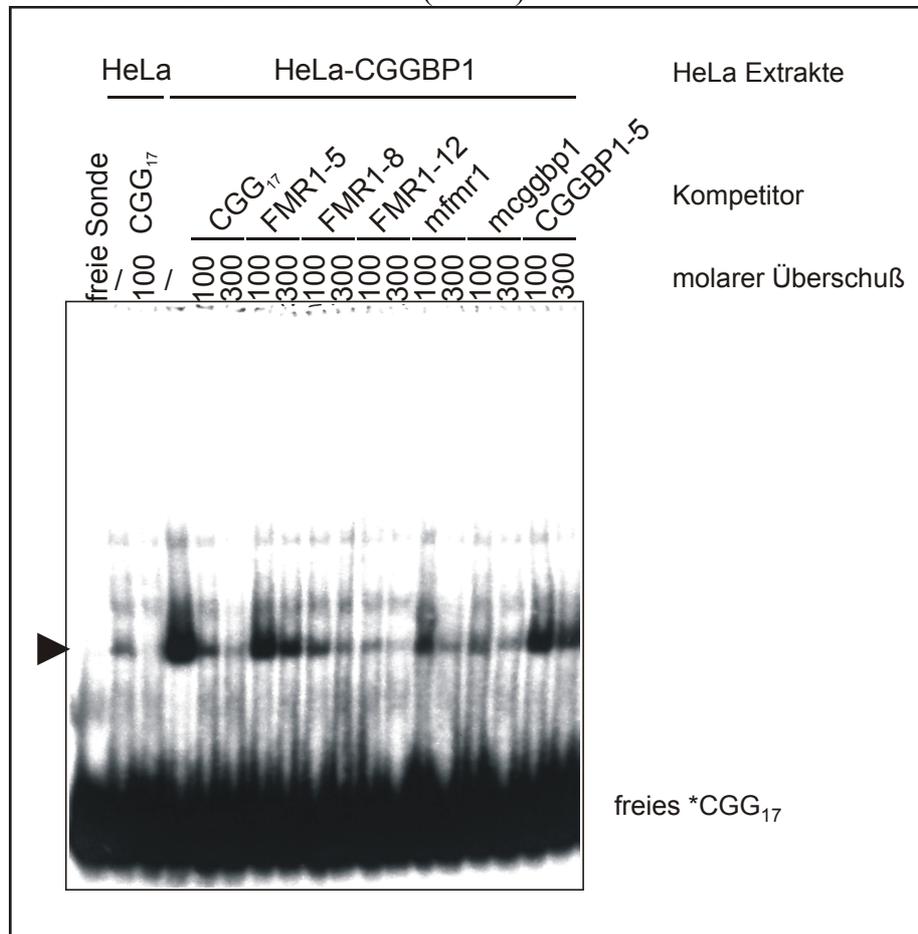


Abbildung 39 Kompetitions-Experimente

Zellextrakte aus CGGBP1-überexprimierenden HeLa-Zellen (HeLa-CGGBP1) wurden mit der (CGG)₁₇-Sonde inkubiert und mit den angegebenen nicht radioaktiv markierten Oligodesoxyribonukleotiden (Tab. 9) kompetiert, die in 100- und 300-fachen molaren Überschuss zugegeben worden waren. Die Protein-DNA-Komplexe wurden auf ein 3,5%ige PAA-Gel aufgetragen.

Außerdem wurde der Teil des CGGBP1-Promotors verwendet, in dem eine 5'-d(CCG)₅-3'-Folge lag (CGGBP1-5). Hierzu wurden die radioaktiv markierten doppelsträngigen Oligodesoxyribonukleotide mit Zellextrakten von CGGBP1-überexprimierenden Zellen inkubiert und auf ein 3,5 %iges PAA-Gel aufgetragen (Abb. 39-40). Der zu beobachtende Protein-DNA-Komplex wurde mit einem im Überschuss angebotenen, nicht radioaktiv markierten 5'-d(CGG)₁₇-3' Oligodesoxyribonukleotid kompetiert (Abb. 40). Im Vergleich zum Komplex aus CGGBP1 und dem FMR1-12-Oligodesoxyribonukleotid war die Bindung an fünf 5'-

d(CGG)-3'-Wiederholungen wesentlich schwächer. Die Stärke des CGG-spezifischen Komplex aus CGGBP1 und FMR1-8 lag zwischen der der Komplexe mit FMR1-5 und FMR1-12. Analog dazu kompetierten CGGBP1-5 und FMR1-5 die Bildung des Komplexes aus CGGBP1 und der Sonde 5'-d(CGG)₁₇-3' selbst bei einem 300-fachen molaren Überschuß nur schwach (Abb. 39). Die Konkurrenz von FMR1-8 war mit der von 5'-d(CGG)₁₇-3' vergleichbar.

CGGBP1 band somit auch an 5'-CGG_n-3'-Folgen, wenn das Trinukleotid weniger als acht mal wiederholt wurde. Eine Bindung an ein 5'-d(CGG)₅-3'-Oligodesoxyribonukleotid war unabhängig von der es flankierenden Sequenz im Vergleich zur Bindung an 5'-d(CGG)₈-3' deutlich schwächer.

D.6.2 Bindung an unterbrochene 5'-d(CGG)_n-3'-Trinukleotid-Folgen

Um die Toleranz von CGGBP1 gegenüber Unterbrechung einer 5'-d(CGG)_n-3'-Wiederholung mit davon abweichenden Sequenzen zu testen, wurden natürlich vorkommende unterbrochene 5'-(CGG)_n-3'-Wiederholungen wie oben beschrieben in Gelretardations-Experimenten eingesetzt (Tab. 9 und Abb. 39-40). Die reine Wiederholung von 5'-(CGG)-3' wurde in bisherigen Experimenten nach neun tandem-artig wiederholten 5'-(CGG)-3'-Trinukleotiden mit 5'-(AGG)-3' unterbrochen (Deissler et al., 1997). Diese Unterbrechung, und ein nach sieben 5'-(CGG)-3' plaziertes 5'-(TGG)-3' beeinträchtigte die Komplexbildung mit CGGBP1 nicht (Müller-Hartmann et al., 2000). Allerdings hatte sich ein Oligodesoxyribonukleotid, in dem 5'-(CGG)-3' mit 5'-(TGG)-3' alterniert, schon in anderen Experimenten als schwacher Konkurrent erwiesen, so daß man bei der Unterbrechung mit 5'-(TGG)-3' nicht von einer die Bindung störenden Unterbrechung sprechen konnte (Deissler et al., 1996).

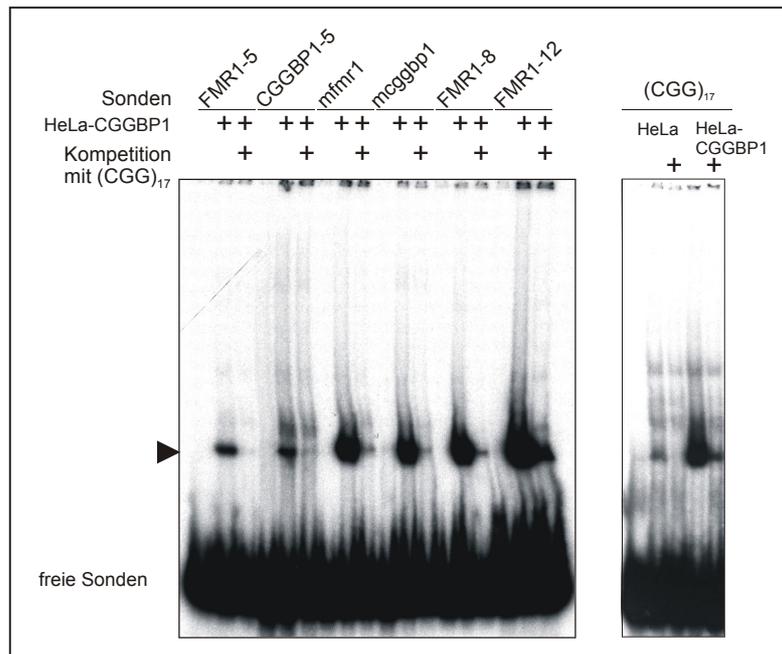


Abbildung 40 Gelretardation mit CGGBP1

Zellextrakte aus CGGBP1-überexprimierenden HeLa-Zellen (HeLa-CGGBP1) mit den angegebenen Sonden (Tab. 9), die jeweils einmal ohne Zellextrakte auf das 3,5%ige PAA-Gel aufgetragen worden waren. Die DNA-Komplexe wurden durch Zugabe eines 100-fachen molaren Überschusses von nicht radioaktiv markiertem (CGG)₁₇ kompetiert. Zellextrakte von nicht transfizierten Zellen (HeLa) wurden zum Vergleich mit CGGBP1-überexprimierenden HeLa-Zellen mit der (CGG)₁₇-Sonde inkubiert.

Im *fmr1*-Gen des BALB/c Mausstammes ist die 5'-(CGG)₈-3'-Folge nach sechs 5'-(CGG)-3' Wiederholungen durch das Trinukleotid 5'-(CGA)-3' unterbrochen (Ashley et al., 1993). Ein 51 bp langes Oligodesoxyribonukleotid wurde danach ausgewählt und als *mfmr* bezeichnet. Die 5'-(GCC)₅-3'-Folge im Promotor des menschlichen CGGBP1 Gens war in der Maussequenz konserviert (Abb. 24). Im C57BL/6-Stamm wurde 5'-(GCC)-3' insgesamt neun mal mit einer Unterbrechung durch das Trinukleotid 5'-d(CGA)-3' wiederholt (*mcggbp1*; Tab. 9). Beide Oligodesoxyribonukleotide bildeten einen starken Komplex mit CGGBP1, der sich in seinen Kompetitionseigenschaften nicht von dem mit 5'-(CGG)₁₇-3' unterschied. Demzufolge tolerierte CGGBP1 Unterbrechungen schon nach sechs 5'-(CGG)-3'-Wiederholungen. Es ist hier allerdings nicht auszuschließen, daß eine starke Komplexbildung mit CGGBP1 schon ab einem Schwellenwert von sechs Wiederholungen von 5'-(CGG)-3' erfolgte.

E Diskussion

Die Amplifikation bestimmter Mikrosatelliten führt beim Menschen zu einer Reihe von genetischen Krankheiten (Cummings und Zoghbi, 2000; Margolis und Ross, 2001). Die Funktionen der betroffenen Gene, die molekulare Pathogenese und der Mechanismus solcher Amplifikationen sind von grundsätzlichem Interesse. Die Analyse von Proteinen, die repetitive Sequenzen erkennen, könnte zur Aufklärung der Funktion(en) von Mikrosatelliten beitragen. Es ist möglich, daß solche Proteine auch eine Rolle bei der stabilen Vererbung dieser Sequenz-Wiederholungen spielen. Es wurde deshalb in unserem Labor ein 5'-d(CGG)_n-3'-bindendes Protein (CGGBP1) gereinigt und teilweise charakterisiert (Deissler et al., 1996 und 1997; Müller-Hartmann et al., 2000).

Thema dieser Dissertation war die Aufklärung der Genstrukturen und der Expression der CGGBP1-Gene des Menschen und der Maus. Es wurde zudem beim menschlichen Gen die Promotorregion definiert und der Methylierungsstatus in zwei Zelltypen bestimmt. Außerdem wurde eine Strategie zur konditionalen Inaktivierung des *cggbp1*-Gens der Maus entworfen. In Gelretardations-Experimenten (EMSA) wurde die DNA-Bindung von CGGBP1 an weitere natürlich vorkommende 5'-d(CGG)_n-3'-Wiederholungen getestet. Dabei wurde untersucht, ob 5'-d(CGG)_n-3'-Wiederholungen mit weniger als 8 Einheiten als Bindungssequenzen für CGGBP1 dienten. Außerdem wurde der Einfluß von Unterbrechungen in der perfekten Wiederholung von 5'-d(CGG)_n-3'-Trinukleotiden analysiert.

E.1 Der genomische Klon des *cggbp1*-Gens

Ein 13.828 bp großes DNA-Fragment wurde aus einer Bank aus genomischer DNA des C57BL/6-Mausstamms isoliert und komplett sequenziert. Es enthielt die drei Exons des *cggbp1*-Gens der Maus. Der genomische Klon des CGGBP1-Gens des Menschen war bereits im Rahmen meiner Diplomarbeit isoliert worden (Naumann, 1998). Bei Datenbankrecherchen wurde hier auch das *cggbp1*-Gen der Ratte identifiziert und in vergleichende Sequenzanalysen mit einbezogen. Die Genstrukturen der CGGBP1-Gene der beiden Nagetiere waren identisch, während das menschliche Gen ein zusätzliches Exon in der 5'-UTR enthielt. Das offene Leseraster dieses Gens in den drei Spezies war nicht durch Introns unterbrochen. Da die 5'-UTR jedoch mindestens zwei Introns aufwies, wurde ausgeschlossen, daß es sich bei den klonierten Sequenzen um prozessierte Pseudogene handelte.

Die genomischen Sequenzen, die die CGGBP1-Gene flankierten, enthielten verschiedene repetitive Elemente, z. B. auch die der *short* und *long interspersed elements* (SINEs bzw. LINEs). Das Muster der meisten repetitiven Elemente unterschied sich bei allen drei

untersuchten genomischen Sequenzen. Die Position einiger SINE-Elemente der B1-Familie waren zwischen Maus und Ratte jedoch identisch. Stromabwärts des *cggbp1*-Gens traten mehr LINE-Elemente auf als im genomischen Klon des Menschen. Möglicherweise beruht die höhere Anzahl von LINE-Elemente im genomischen Klon der Maus auf ihrer im Vergleich zum menschlichen Genom höheren Aktivität im Mausgenom (Mouse Genome Sequencing Consortium, 2002).

E.2 Zur Expression des CGGBP1-Gens

Das CGGBP1-Gen scheint eine grundlegende Rolle in der Physiologie der Zellen zu spielen, da es nach bisherigen Untersuchungen in allen untersuchten Geweben des Menschen, sowie während der Embryonalentwicklung der Maus transkribiert wird. Aufgrund der ubiquitären Transkription in menschlichen Geweben konnten die möglichen Funktionen des CGGBP1-Gens jedoch nicht eingegrenzt werden.

Das CGGBP1-Gen wird oft in Geweben abgelesen, in denen auch das FMR1-Gen stark transkribiert wird, z. B. im Gehirn und in Hoden (Hinds et al., 1993). In HeLa-Zellen war der FMR1-Promotor jedoch durch eine Überexpression von CGGBP1 inhibiert worden (Müller-Hartmann et al., 2000). Die Expression des FMR1-Gens war dabei in Abhängigkeit der Länge einer 5'-d(CGG)-3'-Wiederholung durch CGGBP1 verringert worden. Das FMR1-Gen wird in Blutzellen von Trägern der Prämutation, bei der die 5'-(CGG)-3'-Wiederholung in der 5'-UTR im FMR1-Gen über 50 Einheiten aufweist, jedoch stärker transkribiert als von einem Allel mit weniger als 50 Einheiten (Tassone, 2000c). Die Inhibition des FMR1-Promotors durch CGGBP1 in HeLa-Zellen könnte auf zelltyp-spezifischen Interaktionspartnern von CGGBP1 beruhen.

Andererseits ist auch nicht auszuschließen, daß die in RNA-Transfer-Hybridisierungen nachgewiesene mRNA-Menge des CGGBP1-Gens nicht mit der Proteinmenge korreliert. Sehr lange und in Säugetieren hoch konservierte nicht-translatierte Regionen im CGGBP1-Gen ließen eine komplexe posttranskriptionale Kontrolle vermuten. Möglicherweise sind den konservierten Sequenzen Funktionen zugeordnet, die für die zeitgerechte und gewebe-spezifische Expression des CGGBP1-Gens *in vivo* notwendig sind.

E.3 Das 5'-d(CGG)_n-3'-bindende Protein

E.3.1 Eigenschaften des offenen Leserasters

Das CGGBP1 war bei verschiedenen Säugetieren, und in der verfügbaren C-terminalen Hälfte der Proteinsequenz bei Hühnern, hoch konserviert. Die Primärsequenz scheint also für die Tertiärstruktur des 167 Aminosäuren langen Proteins sehr wichtig zu sein. So entsprechen die sehr niedrigen dN/dS-Werte (Yang und Nielsen, 2000) zwischen Maus und Mensch den Werten, die bei orthologen Domänen, v.a. von nukleär lokalisierten Proteinen beobachtet werden (Mouse Genome Sequencing Consortium, 2002). Ähnlich groß angelegte Studien wie die des Mouse Genome Sequencing Consortiums zu dN/dS-Werten zwischen den anderen genannten Spezies liegen noch nicht vor. Die dN/dS-Werte zwischen Maus und Ratte entsprechen in etwa den Ergebnissen einer kleineren Studie (Wolfe und Sharp, 1993).

Das CGGBP1-Protein ist ein Kernprotein und enthält gemäß dem Vergleich mit der Datenbank Interpro (Mulder et al., 2003) zwei Domänen bzw. Motive: eine zweiteilige Kernlokalisationssequenz und ein Zink-Finger-Motiv mit zwei Cystein- und zwei Histidinresten. Die Funktionalität des zweiteiligen Kernlokalisierungssignals war durch Deletionsmutanten bewiesen worden (Müller-Hartmann et al., 2000). Die modifizierten Proteine sollten dabei auch dazu verwendet werden, die DNA-Bindungsdomäne zu lokalisieren. Sie schien mit der Kernlokalisationssequenz zu überlappen. Eine Rolle des identifizierten Zink-Finger-Motivs für die DNA-Bindung konnte in den Studien von Müller-Hartmann et al. nicht ausgeschlossen werden. Angesichts der hohen Konservierung der Sequenz in verschiedenen Spezies ist es nun wahrscheinlich, daß die eingeführten Deletionen das Protein so stark verändern, daß es seine Funktion nicht mehr wahrnehmen konnte. Dafür spricht auch, daß eine C-terminale Deletionsmutante zwar noch in den Kern transportiert worden war, jedoch nicht mehr an DNA band (Müller-Hartmann et al., 2000), obwohl sowohl das Zink-Finger-Motiv als auch die basische Region des Kernlokalisierungssignals im Protein erhalten waren (Abb.41).

Durch den Sequenzvergleich von CGGBP1 mit Zink-Finger-Motiven aus DNA-bindenden Proteinen ließen sich wenig Anhaltspunkte für ein möglicherweise funktionelles Zink-Finger-Motiv in CGGBP1 finden (Müller-Hartmann, 2000). So enthalten alle beschriebenen DNA-bindenden Zink-Finger-Proteine mindestens drei dieser Bindungsdomänen, im Gegensatz zu einer in CGGBP1 (Iuchi, 2001). Hier wurde die Ähnlichkeit des Zink-Finger-Motivs von CGGBP1 mit dem der Unterklasse der RNA-bindenden U1-ähnlichen oder Matrin-ähnlichen Zink-Finger-Proteinen aufgezeigt (Nelissen et al., 1991; Matsushima et al., 1997; Bedford et

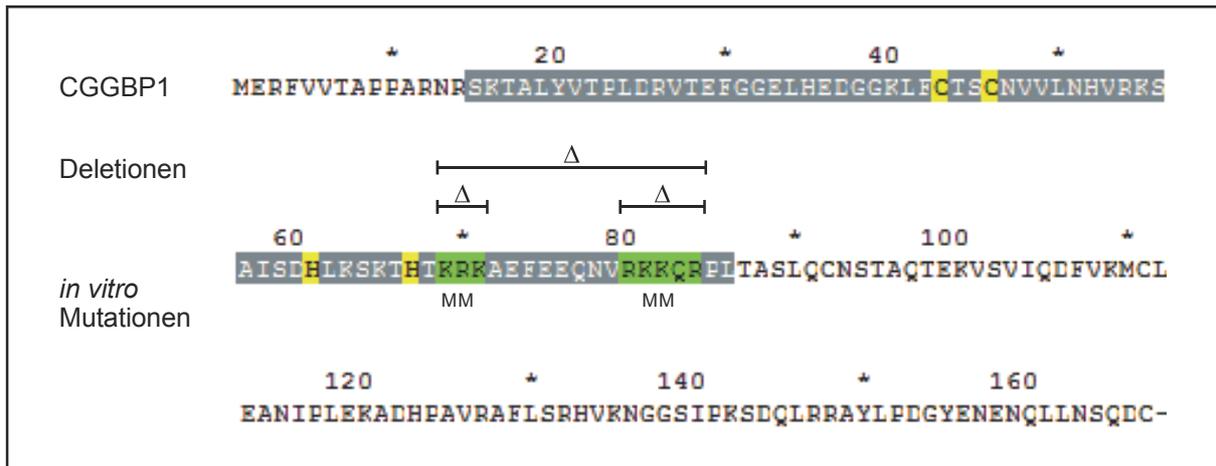


Abbildung 41 In-vitro-Mutationen von CGGBP1

Proteinsequenz von CGGBP1 unter Angabe der eingeführten kleinen Deletionen (Δ). Der grau hinterlegte Bereich fehlte in der "N-terminalen"-Deletion. Mit Ausnahme der Deletion der Aminosäuren 80-84 band keine der Mutanten an DNA. Vier Aminosäuren wurden in vier verschiedenen Mutanten in Methionin umgewandelt. Die DNA-Bindung der Mutanten K69M, und R70M war im Vergleich zum Wildtyp leicht eingeschränkt, die DNA-Bindung der anderen Mutanten war unverändert (Müller-Hartmann, 2000).

al., 1998). Die bisher beschriebenen Proteine der Unterklasse der U1-ähnlichen Zink-Finger-Proteine enthalten maximal zwei Zink-Finger-Motive (Eintrag in der PROSITE-Datenbank (Hofman et al., 1999) <http://www.expasy.org/cgi-bin/prosite-search-ac?PS50171>).

E.3.2 Funktionelle Überlegungen

Bisher hatten wir uns auf die doppelsträngigen DNA-Bindungseigenschaften des CGGBP1 konzentriert, da CGGBP1 in EMSA nicht an einzelsträngige DNA band (Heidrun Deissler, persönliche Mitteilung). Die Ähnlichkeit des Zink-Finger-Motivs in CGGBP1 mit dem von RNA-bindenden Proteinen könnte ein Hinweis auf eine mögliche RNA-Bindung durch CGGBP1 sein. Es sind weitere Experimente nötig, um eine solche Bindung des CGGBP1 an RNA zu diskutieren. Der hier aufgezeigte Anhaltspunkt für eine mögliche RNA-Bindung durch CGGBP1 ist in Hinblick auf den mittlerweile entdeckten Phänotyp von einigen Trägern der Prämutation der 5'-(CGG)-3'-Wiederholung im FMR1-Gen interessant (Hagerman und Hagerman, 2002). Das FMR1-Gen wird in Abhängigkeit der Länge der 5'-(CGG)-3'-Wiederholung stärker transkribiert, wobei das Transkript jedoch weniger effizient translatiert wird (Tassone et al., 2000c; Kenneson et al., 2001; Primerano et al., 2002). Könnte ein kern-lokalisierendes, hypothetisch RNA-bindendes CGGBP1 die effiziente Translation z. B. durch Beeinflussung des mRNA-Transports beeinträchtigen? Nach einer möglichen Etablierung von

RNA-bindenden Eigenschaften von CGGBP1 durch EMSA oder North-Western-Experimente könnte dieser Frage mittels RNA-*in situ*-Hybridisierung von Zellen von Prämutations-Trägern nachgegangen werden.

Bei einigen Individuen mit Prämutations-Allelen am FMR1-Locus, die ein neurodegeneratives Syndrom entwickelten (Hagerman et al., 2001), wurden in bestimmten Gehirnarealen intranukleäre Einschlüsse bisher ungeklärter Zusammensetzung beobachtet (Greco et al., 2002). Es wird vermutet, daß diese Einschlüsse auf der erhöhten FMR1-RNA-Menge in Prämutations-Trägern beruhen. In einer Kooperation mit Ben Oostras Labor (Rotterdam) wird nun untersucht, ob das CGGBP1 Bestandteil dieser Einschlüsse ist.

E.4 Eigenschaften der 5'-nicht translatierten Regionen der CGGBP1-Gene

Die 5'-UTR war beim menschlichen Gen aus drei Exons aufgebaut, von denen das dritte alternativ gespleißt wurde. Das alternativ gespleißte Exon III wurde als zusätzliche mRNA-Population in allen untersuchten Zelltypen gefunden. Wahrscheinlich werden zwischen 40 - 60% der bisher bekannten menschlichen Gene alternativ gespleißt (International Human Genome Sequencing Consortium, 2001; Übersicht in Modrek und Lee, 2002). Während das alternative Spleißen innerhalb der translatierten Region zu Proteinen unterschiedlicher Funktion führen kann, ist das ebenfalls weit verbreitete alternative Spleißen in der 5'-UTR weniger gut verstanden (Mironov et al., 1999).

Die nicht-translatierten Regionen können Elemente enthalten, denen experimentell wichtige Aufgaben bei der posttranskriptionalen Kontrolle der Genexpression zugewiesen wurden (Übersichten in Kozak, 1996; Gray und Hentze, 1994; Curtis et al., 1995; Meijer und Thomas, 2002). Es wurde vorgeschlagen, daß Sekundärstrukturen oder zusätzliche Startcodons in der 5'-UTR die Translation des eigentlichen ORFs verringern können (Kozak, 1991). Gemäß des Scanning-Modells der Translation werden Ribosomen durch solche Elemente behindert, wenn sie die mRNA von der Cap-Struktur am 5'-Ende aus nach dem Startcodon für das kodierte Protein absuchen (Kozak, 1989). Bei der statistischen Auswertung von 5'-UTR Sequenzen von hoch bzw. niedrig exprimierten Proteinen beobachtet man, daß sich die 5'-UTR Sequenzen hinsichtlich der Länge, des GC-Gehalts und der Häufigkeit von uORFs signifikant unterscheiden (Kozak, 1987; 1996; Kochetov et al., 1998).

Früher wurde angenommen, daß Elemente wie potentielle Sekundärstrukturen oder Startcodons nur in einem geringem Prozentsatz (10 %) von mRNAs vorkamen (Kozak 1987), in denen wachstums-regulierende Gene angereichert waren. Mittlerweile steigt mit zunehmender Menge der zur Verfügung stehenden Sequenzinformationen auch der Anteil an

mRNAs, die diese Elemente enthalten (Suzuki et al., 2000, Pesole et al., 2000). Da es sich bisher meist um auf Sequenzeigenschaften basierenden Studien handelte, steht der experimentelle Nachweis für die posttranskriptionale Kontrolle für die meisten Transkripte noch aus. Man geht mittlerweile davon aus, daß es sich bei der posttranskriptionalen Kontrolle der Genexpression eher um eine Modulation, als – wie bisher angenommen – um eine Inhibition der Translation handelt (Meijer und Thomas, 2002).

Die 5'-UTR aller untersuchten CGGBP1-Gene wiesen Eigenschaften auf, durch die die Translationseffizienz moduliert werden könnte. In Transfektions-Experimenten, in denen die 5'-UTR des CGGBP1-Gens mit dem Reporter-gen Luciferase verbunden worden war, wurde jedoch kein Einfluß der 5'-UTR auf die Reporter-gen-Aktivität festgestellt. Dieses Ergebnis war unerwartet angesichts der hoch konservierten 5'-UTR Sequenzen in mehreren Säugetieren.

So war die 5'-UTR doppelt so lang wie durchschnittliche 5'-UTR (International Human Genome Sequencing Consortium, 2001, Mouse Genome Sequencing Consortium, 2002). Das erste Exon war zudem sehr GC-reich, so daß sich wahrscheinlich stabile Sekundärstrukturen ausbilden können. Im menschlichen Gen waren fünf und in den CGGBP1-Genen der anderen Säugetiere zwei zusätzliche Startcodons in der 5'-UTR enthalten. Meist entsprach die Sequenzumgebung dieser Startcodons jedoch nicht der Consensussequenz für funktionelle Startcodons (Kozak, 1987).

Das erste ATG, das zudem weniger als die anderen von der Kozak-Sequenz abwich, führte in allen untersuchten Spezies zu einem 25 Aminosäuren langen Peptid. Die Peptidsequenz war in den untersuchten Säugetieren zu mindestens 88% konserviert, da die Hälfte der Nukleotidaustausche innerhalb dieses uORFs synonyme Positionen betraf. Die hohe Konservierung der Peptidsequenz könnte darauf beruhen, daß Proteine sequenzspezifisch die zugrundeliegende RNA-Sequenz binden, so daß mehr oder davon abweichende Nukleotidaustausche nicht toleriert würden. Die translationsregulierende Wirkung von in der 5'-UTR liegenden Minicistrons hing bei anderen Genen auch von der kodierten Peptidsequenzen (zur Übersicht Meijer und Thomas, 2002; Mize 1998; Wang et al., 1998).

Die posttranskriptionale Kontrolle der Genexpression spielt eine entscheidende Rolle bei der Genexpression in Abwesenheit von Transkription, so z. B. nach der Befruchtung und im frühen Embryo (Meijer und Thomas, 2002; Curtis et al., 1995). Ob ein Transkript in Protein übersetzt wird, hängt neben globalen Effekten, wie der Verfügbarkeit von Translationsfaktoren auch häufig von zelltypspezifischen (Ruan et al., 1994) oder von physiologischen Bedingungen ab, wie es z. B. für die Ferritin-mRNA gezeigt worden ist (Übersicht in

Klausner et al., 1993). Der fehlende Einfluß des 5'-UTRs des CGGBP1-Gens auf die Expression des Reportergens in HeLa könnte also auf zelltyp-spezifischen Einflüssen beruhen. Außerdem ist es möglich, daß die 5'-UTR nur die Stabilität des Transkriptes erhöhte, was aufgrund der Analyse nur eines Zeitpunktes nach der Transfektion nicht nachgewiesen werden konnte.

E.5 Alternative Polyadenylierung im CGGBP1-Gen

Die CGGBP1-mRNA wurde alternativ polyadenyliert, so daß in RNA-Transfer-Experimenten mindestens zwei Transkripte detektiert wurden. Das Protein wurde durch die alternative Polyadenylierung nicht verändert, während sich die Länge der 3'-UTR um etwa 3,4 kb veränderte. In 3'-UTR sind häufig Elemente enthalten, die die Stabilität der mRNA oder die Effizienz, mit der die mRNA translatiert wird, beeinflussen (Jackson, 1993; Mitchell und Tollervey, 2001). Außerdem kann die 3'-UTR die Lokalisation der mRNA innerhalb der Zelle beeinflussen (Lipshitz und Smibert, 2000; Jansen, 2001).

Ähnlich wie beim alternativen Spleißen werden in zumindest teilweise auf Datenbanken basierenden Projekten immer mehr Transkripte bei Mensch und Maus entdeckt, die alternativ polyadenyliert werden (Gautheret et al., 1998; Pauws et al., 2001; Iseli et al., 2002). In einer Übersicht über experimentell nachgewiesenes alternatives Polyadenylieren fiel – wie auch im Falle des CGGBP1-Gens – die Wahl des Polyadenylierungssignals in Hoden häufig anders aus, als in anderen Geweben (Edwalds-Gilbert et al., 1997).

E.6 Der Promotor des CGGBP1-Gens

Die stromaufwärts des CGGBP1-Gens gelegenen Sequenzen waren ähnlich den nicht-translatierten Regionen der CGGBP1-Gene zwischen Mensch, Maus und Ratte hoch konserviert. Regulatorische Regionen von Genen können sich so stark unterscheiden, daß man funktionell entscheidende Signale durch das sogenannte phylogenetische *footprinting* identifizieren kann (Gumucio et al., 1996; Duret und Bucher, 1997). Im Vergleich des menschlichen Genoms und des Mausgenoms waren durchschnittlich 200 bp stromaufwärts der Transkriptionsstartpunkte zu 73,9% konserviert (Mouse Genome Sequencing Consortium, 2002). In Transfektions-Experimenten in HeLa-Zellen hatte sich eine 235 bp große Region der 5'-flankierenden Sequenz des CGGBP1-Gens als ausreichend für eine Promotorfunktion erwiesen.

Möglicherweise geben Transfektionsstudien in der Tumorzelllinie HeLa jedoch nicht alle *in vivo* herrschenden Verhältnisse wieder. Die regulatorischen Elemente für die Expression des Gens während der Embryonalentwicklung wären ein Beispiel. So waren bei Mensch, Maus und Ratte insgesamt 600 bp stromaufwärts des Exon I zu mindestens 85 % konserviert und nicht nur die etwa 200 bp große Region, die für eine starke Expression in HeLa-Zellen benötigt worden war. Da sich weiter stromaufwärts gelegene Sequenz oder auch die Sequenz des zweiten Introns zwischen Mensch und beiden Säugetieren stark unterschied, war davon auszugehen, daß Veränderungen in der konservierten Region nicht toleriert wurden, da sie wahrscheinlich eine wichtige Rolle bei der Genexpression des CGGBP1-Gens erfüllten (Duret et al., 1993).

Die teilweise konservierte Sequenz im ersten Intron könnten darauf hinweisen, daß dieses Intron, wie in anderen Genen häufig beobachtet, weitere regulatorische Signale oder einen zweiten Promotor enthielt (Majewski und Ott, 2002). Die Verwendung von alternativen Promotoren ist, wie die posttranskriptionale Kontrolle auch, ein häufiges Mittel, die Genexpression flexibler zu gestalten (Ayoubi und van den Wen, 1996).

E.7 Zur Methylierung des CGGBP1-Gens

In Säugetiergenomen wird fast ausschließlich das Desoxycytidin im Dinukleotid CpG am C-5 Atom des Pyrimidinrings methyliert, wobei die Desoxycytidinreste beider Stränge modifiziert werden (Riggs und Jones, 1983; Doerfler 1981, 1983). Das Methylierungsmuster bestimmter Regionen des menschlichen Genoms kann sich in verschiedenen Gewebe- oder Zelltypen unterscheiden, wobei das zelltyp-spezifische Muster zwischen verschiedenen Individuen oftmals übereinstimmt (Kochanek et al., 1990; 1991; Doerfler et al., 1990, Behn-Krappa et al., 1991; Achten et al., 1991). Das Dinukleotid CpG ist in verschiedenen Genomen statistisch gesehen unterrepräsentiert (Subak-Sharpe et al., 1966; Russel et al., 1976; International Human Genome Sequencing Consortium, 2001). Die meisten CpG-Dinukleotide treten im menschlichen Genom gehäuft in repetitiven Elementen, v.a. in Alu-Elementen, und in so genannten CpG-Inseln auf (Bird et al., 1986; International Human Genome Sequencing Consortium, 2001). Zunächst sah es danach aus, daß v.a. „Haushaltsgene“ mit CpG-Inseln assoziiert waren (Antequera und Bird, 1993). Mittlerweile werden jedoch immer mehr Gene entdeckt, die dieser Klassifizierung nicht entsprechen (Suzuki et al.; 2002, Shiota und Yanigamachi, 2002).

Das Ausmaß der Methylierung im gesamten Genom variiert während der Entwicklung (Razin und Shemer, 1995). So beobachtet man z. B. nach der Befruchtung eine starke Demethylierung des Genoms, das durch eine folgende *de-novo*-Methylierung in weiten Teilen ein neues oder das alte Methylierungsmuster erhält. Methylierung ändert sich auch mit verschiedenen physiologischen Zuständen, z.B. verändert sich das Methylierungsmuster in Krebszellen oder im Alter (Übersicht in Jaenisch und Bird, 2003). Es besteht ein gut dokumentierter Zusammenhang zwischen Methylierung und Promotor-Aktivität; eine Vorhersage des Methylierungsstatus und seine Auswirkungen auf die Promotor-Aktivität ist jedoch nicht möglich (Übersichten in Munnes und Doerfler, 1997; Shiota und Yanagimachi, 2002). Methylierung stellt somit wahrscheinlich ein wichtiges Element bei der Organisation des Genoms dar. Es werden deshalb Anstrengungen unternommen, nach der Bestimmung der Sequenz des Menschen nun auch das - man verzeihe diesen Neologismus - „Methylom“ zu untersuchen (Feinberg, 2001; Novik et al., 2002).

Es gibt verschiedene Ansätze um zu erklären, wie Methylierung die Aktivität eines Promotors reguliert. Methylierung kann die Bindung von Transkriptionsfaktoren verhindern oder die Zusammensetzung der Proteine, die an einen bestimmten Promotor binden, verändern (Übersicht in Bird und Wolffe, 1999). Ein weiterer Zusammenhang besteht zwischen dem Methylierungsstatus und der Chromatinstruktur eines Gens, die für die Transkription eines Gens ebenfalls eine Rolle spielt (El-Osta und Wolffe, 2000). Werden Promotorsequenzen *in vitro* methyliert, hängt die Auswirkung auf die Promotor-Aktivität vom Muster der Methylgruppen in der Promotorsequenz ab, und nicht von der Menge der methylierten CpG-Dinukleotide (Munnes et al., 1998; Muiznieks und Doerfler, 1994).

Der Promotor des CGGBP1-Gens war in weißen Blutzellen von männlichen und weiblichen Probanden, in HeLa-Zellen und – soweit analysiert – in KB-Zellen nicht methyliert. Der Methylierungsstatus wurde hier mit der Bisulfit-Methode des genomischen Sequenzierens bestimmt, mit der alle methylierten Desoxycytidine unabhängig von ihrem Sequenzzusammenhang erfaßt werden können. Mit isoschizomeren Restriktionsendonukleasen, die sich in ihrer Empfindlichkeit gegenüber Methylierung unterscheiden, hätte nur Auskunft über 18 % der hier analysierten CpG-Dinukleotide gewonnen werden können.

Die Beobachtung, daß der Methylierungszustand in HeLa- und KB-Zellen mit dem von primären Zellen übereinstimmte, war insofern bemerkenswert, weil sich das Methylierungsmuster in Zelllinien in Kultur sonst häufig unterscheidet (Doerfler, 1993; Behn-

Krappa et al., 1991; Achten et al., 1991; Antequera et al., 1990). In RNA-Transfer Hybridisierungen und RT-PCR-Experimenten wurde nachgewiesen, daß das CGGBP1-Gen auch tatsächlich in allen untersuchten Zelltypen transkribiert wurde.

In Transfektions-Experimenten wurde die Empfindlichkeit des CGGBP1-Promotors auf *in-vitro*-Methylierung untersucht. Die zwei Konstrukte d03 und d09, die in der CGGBP1-Promotorregion jeweils 32 bzw. 18 CpG-Dinukleotide trugen, wurden mit den Methylasen M-HhaI, M-HpaII und M-SssI behandelt. Im Konstrukt d03 wurden dabei jeweils drei HhaI bzw. HpaII-Erkennungssequenzen methyliert, während die gleichen Methylasen im Konstrukt d09 nur jeweils ein CpG-Dinukleotid modifizierten. Durch die vollständige Methylierung aller 18 bzw. 32 CpG-Dinukleotide durch die M-SssI wurde der CGGBP1-Promotor inaktiviert. Die Methylierung von nur einem oder drei CpG-Dinukleotiden verringerte die Promotor-Aktivität nur geringfügig. Es ist nicht auszuschließen, daß der CGGBP1-Promotors nur durch bestimmte der 18 bzw. 32 CpG-Dinukleotide inaktiviert wurde. Die *in-vitro*-Methylierung von CpG-Dinukleotiden, die außerhalb von Erkennungssequenzen von Methylasen liegen, ist technisch jedoch sehr aufwendig und war nicht Gegenstand dieser Studie.

E.8 Zur konditionalen Inaktivierung des *cggbp1*-Gens

Das *cggbp1*-Gen wurde im Rahmen dieser Arbeit kloniert, um zukünftig die Inaktivierung dieses Gens in der Maus zu ermöglichen. Geninaktivierungs-Experimente (*knock-out*-Mäuse) können zur Aufklärung der Funktion eines Gens beitragen (Capecchi, 1989). Ein Geninaktivierungs-Vektor, in dem das *cggbp1*-Gen von LoxP-Zielsequenzen (Sternberg und Hamilton, 1981) flankiert wurde, ist fertiggestellt worden. Die Strategie kann in zwei Mausstämmen, dem C57BL/6 sowie dem 129-Mausstamm angewendet werden. Mit der Transfektion des Vektors in ES-Zellen wurde begonnen, bisher sind jedoch keine homologen Rekombinanten entdeckt worden. Die Effizienz der homologen Rekombination von Geninaktivierungs-Vektoren mit dem endogenen Gen ist abhängig vom Genlocus und kann selbst innerhalb eines Genlocus bei Verwendung unterschiedlicher DNA-Abschnitte stark variieren (Joyner, 1993). Es ist möglich, daß der Geninaktivierungs-Vektor, der zunächst auch einen 1,65 kb langen, ununterbrochenen repetitiven Abschnitt enthielt, die Identifizierung homologer Rekombinanten verhinderte. Deshalb wurde dieser Bereich aus dem Geninaktivierungs-Vektor entfernt. Die Transfektion in ES-Zellen im Labor von Ulrich Müller ist in Planung.

Es ist meist nicht möglich, den Phänotyp einer *knock-out*-Maus vorherzusagen (Bradley, 2002). Je nach beobachtetem Phänotyp können viele verschiedene Parameter untersucht werden, die zur Aufklärung der Funktion von *cggbp1* beitragen könnten. Da CGGBP1 nach Überexpression den menschlichen FMR1-Promotor, der eine 5'-d(CGG)_n-3'-Wiederholung enthält (Verkerk et al., 1991), inhibiert (Müller-Hartmann et al., 2000), wäre ein Aspekt, der in *knock-out*-Mäusen untersucht werden könnte, das zelluläre Expressionsmuster. Des Weiteren könnte die Stabilität von 5'-(CGG)_n-3'-Wiederholungen in somatischen Geweben sowie bei der Vererbung verfolgt werden.

Es wurden bei Datenbankrecherchen neben dem Zink-Finger-Motiv (Matsushima et al., 1997) keine dem *cggbp1*-Protein ähnlichen Proteine entdeckt. Trotzdem besteht die Möglichkeit, daß das *cggbp1*-Gen eine redundante Funktion erfüllt, die im Falle des *cggbp1*-Verlustes von einem anderen Protein übernommen wird, so daß eine *knock-out*-Maus des *cggbp1*-Gens keinen sichtbaren Phänotyp entwickelt (Übersicht in Müller, 1999). Eine embryonal-letale Konsequenz des Verlustes des *cggbp1*-Gens ist im Vorfeld auch nicht auszuschließen. Das Gen ist in vielen Säugetieren sehr hoch konserviert, was für eine starke negative Selektion gegen einen Funktionsverlust des *cggbp1*-Gens für den Organismus spricht. Da das Gen zudem in embryonalen Stammzellen und während der Entwicklung transkribiert wird, könnte es eine wichtige Rolle während der Entwicklung spielen. Mit der hier angewendeten konditionalen Strategie zur Ausschaltung des *cggbp1* könnten die Mutanten auch bei schweren Entwicklungsstörungen oder bei einem letalen Phänotyp untersucht werden (Übersicht in Rajewsky et al., 1996; Kühn und Schwenk, 1997; Sauer, 1998). Durch Einkreuzen von transgenen Mausstämmen, die die Cre-Rekombinase unter der Kontrolle von gewebespezifischen oder induzierbaren Promotoren exprimieren (Nagy und Mar, 2001), wird es möglich sein, das *cggbp1*-Gen zu unterschiedlichen Zeitpunkten und spezifisch in verschiedenen Geweben auszuschalten.

Die Zielsequenz LoxP ist nur 34 bp lang, und die Spezifität der Cre-Rekombinase ist nicht absolut (Hoess et al., 1986; Sauer, 1996). Pseudo-LoxP-Sequenzen treten auch im Säugetiergenom auf, die durch die Cre-Rekombinase auch rekombiniert werden können (Thyagarajan et al., 2000). In Zellkultur beobachtet man eine zeit- und dosisabhängige Schädigung der Zellen durch die Cre-Rekombinase (Loonstra et al., 2001). Bei einer Cre-transgenen Mauslinie wurde auch von Chromosomen-Rearrangierung berichtet (Schmidt et al., 2000). Neben Wildtyp-Mäusen muß deshalb der Phänotyp der *knock-out*-Mäuse unbedingt auch mit dem der allein Cre-transgenen Mäuse verglichen werden. Die Möglichkeit eines Einflusses der Expression der Cre-Rekombinase auf den beobachteten Phänotyp einer *knock-*

out-Maus kann allerdings nie ganz ausgeschlossen werden. Außerdem dürfen mögliche Konsequenzen der Manipulation der ES-Zellen in Kultur, oder das Einbringen fremder DNA ins Mausgenom, z .B. auf das zelluläre Methylierungsmuster nicht vernachlässigt werden (Remus et al., 1999; Hertz et al., 1999; Müller et al., 2001; Doerfler et al., 2001).

E.9 Die Bindung von CGGBP1 an kurze und unterbrochene 5'-(CGG)_n-3'-Wiederholungen

In Gelretardations-Experimenten band CGGBP1 hier an kürzere 5'-(CGG)-3'-Wiederholungen, als bisher beschrieben (Deissler et al., 1996 und 1997; Müller-Hartmann et al., 2000). Dies beruhte wahrscheinlich darauf, daß hier die CGG-Wiederholung von zusätzlichen Sequenzen flankiert wurde. Zudem tolerierte CGGBP1 Unterbrechungen in der Bindungssequenz, wie sie in natürlich vorkommenden Wiederholungen häufig auftreten (Katti et al., 2001, Tóth et al., 2000; Mangel et al., 1998). Damit wird das Spektrum möglicher Bindungsstellen für CGGBP1 erweitert und damit die Anzahl der Gene, deren Expression durch CGGBP1 reguliert werden könnte. Nach Überexpression oder Ausschalten des CGGBP1-Gens könnte diese These mit Hilfe der DNA-Array-Technologie geklärt werden. Entsprechende Experimente sind bereits in Planung.

E.10 Ausblick

Die in dieser Arbeit vorgestellten Aspekte zur Genstruktur und Expression des CGGBP1-Gens wurden durchgeführt, um eine solide Grundlage für die weitere Erforschung der Funktion des CGGBP1-Gens zu schaffen. Zu den spannendsten Aspekten zählt dabei, wie viele Gene in ihrer Expression durch das CGGBP1-Gen reguliert werden. Neben der transkriptionsregulierenden Eigenschaft könnte auch den Hinweisen auf eine mögliche RNA-Bindung – und ihrer Konsequenz für die Translation des gebundenen Transkripts – von CGGBP1 nachgegangen werden. Nach einer weiteren Verfeinerung der Anforderungen an die Bindungssequenz des CGGBP1 könnten potentielle Zielgene eventuell auch durch Sequenzabgleich in der bisher ermittelten Genomsequenz des Menschen identifiziert werden.

Es wäre auch interessant zu wissen, ob das CGGBP1 während des Zellzyklus gleichmäßig exprimiert wird, oder ob es zu bestimmten Zeitpunkten (oder durch extrazelluläre Signale) induziert werden kann. Aufgrund der Hinweise auf posttranskriptionale Regulation der Expression sollte hierbei eher auf die Anwesenheit des Proteins und nicht auf die der mRNA geachtet werden.

Molekulare Interaktionspartner des CGGBP1-Gens könnten durch biochemische Analysen wie z. B. die Immunopräzipitation identifiziert werden. Des Weiteren sind *one-* oder *two-hybrid*-Experimente denkbar. Wenn die Funktion der Interaktionspartner bekannt wäre, könnte dies eventuell Rückschlüsse auf die Rolle des CGGBP1 erlauben.

Die Analyse des Phänotyps der *knock-out*-Maus des *cggbp1*-Gens wird vielleicht weitere Ansatzpunkte für die Funktion des *cggbp1*-Gens und damit möglicherweise auch des CGGBP1-Gens des Menschen liefern.

F Zusammenfassung

Die Amplifikation von repetitiven Elementen in oder in der Nähe von Strukturgenen gilt als neuer Mutationsmechanismus bei Erbkrankheiten des Menschen. Das Fragile-X-Syndrom, eine der häufigsten Ursachen für erbliche geistige Behinderung, ist mit der Amplifikation einer 5'-d(CGG)_n-3'-Wiederholung in der 5'-nicht translatierten Region des FMR1-Gens (*fragile X mental retardation gene 1*) assoziiert (Verkerk et al., 1991; Fu et al., 1991). Bei den meisten Patienten ist das FMR1-Gen komplett methyliert und das FMR1-Protein wird nicht exprimiert (Pieretti et al., 1991; Sutcliffe et al., 1992).

Die Funktion von kurzen Sequenz-Wiederholungen, die im Genom weit verbreitet und bei Säugetieren oft konserviert sind, ist nicht geklärt. Auch werden die genetischen Mechanismen, die zur ihrer pathogenen Amplifikation führen, noch untersucht. Das Verrutschen eines DNA-Stranges (*DNA slippage*) während der Replikation, der Rekombination und/oder der Reparatur wurde als möglicher Mechanismus vorgeschlagen, da nur homogene Folgen der repetierten Einheit, die *in vitro* anomale DNA-Strukturen annehmen können (Sinden et al., 2002), instabil vererbt werden.

Die Analyse von Proteinen, die sequenzspezifisch an 5'-d(CGG)_n-3'-Wiederholungen binden, könnte möglicherweise zur Aufklärung der Funktion und/oder Vererbung dieser instabilen Trinukleotid-Wiederholung beitragen. Ein 5'-d(CGG)_n-3'-bindendes Protein (CGGBP1) wurde in unserem Labor gereinigt und charakterisiert (Deissler et al., 1996; 1997). Funktionelle Studien zeigten, daß das CGGBP1 den FMR1-Promotor inhibiert, wenn es in HeLa-Zellen überexprimiert wird (Müller-Hartmann et al., 2000).

In dieser Dissertation wurde die genomische Organisation und Expression des CGGBP1-Gens untersucht. Das menschliche Gen bestand aus 4 Exons und lag auf dem kurzen Arm von Chromosom 3 in der Nähe des Centromers. Das CGGBP1-Gen der Maus wurde hier isoliert und das 14 kb große Fragment genomischer DNA wurde komplett sequenziert. Die Struktur des Maus-Gens wurde aus Sequenzähnlichkeiten zum menschlichen Gen abgeleitet und durch EST-Klon- und RNA-Analysen belegt. Die translatierten und große Bereiche der nicht-translatierten Regionen (UTR) des CGGBP1-Gens waren zwischen Mensch und Maus hoch konserviert. Die Proteinsequenz des CGGBP1 der Maus stimmte bis auf drei Aminosäuren mit der des menschlichen Proteins überein. EST-Klon-Sequenzen des CGGBP1-Gens von anderen Säugetieren und von Hühnern sowie die genomische Sequenz des CGGBP1-Gens der Ratte waren in Datenbanken zugänglich und wiesen ebenfalls einen hohen Grad an Ähnlichkeit zum menschlichen Gen auf. Bei drei Säugetieren (Ratte, Schwein und Rind) war die abgeleitete Aminosäuresequenz zu der Aminosäuresequenz des menschlichen CGGBP1 sogar identisch.

Die translatierte Region der CGGBP1-Gene des Menschen, der Maus und der Ratte war nicht durch Introns unterbrochen. Die 5'-UTR im Menschen bestand aus drei Exons, die der beiden Nagetiere aus je zwei Exons. RT-PCR-Experimente zeigten, daß das Exon in der 5'-UTR des Menschen, zu dem keine Entsprechung im CGGBP1-Gen der Maus oder der Ratte festgestellt wurde, in menschlichen Zellen alternativ gespleißt wurde. Gemäß RNA-Transfer-Experimenten wurden vom ubiquitär exprimierten CGGBP1-Gen des Menschen und der Maus mRNAs von etwa 4,5 und 1,2 kb Länge gebildet. Das CGGBP1-Gen der Maus wurde in embryonalen Stammzellen sowie während der Embryonalentwicklung exprimiert. Das Verhältnis der Mengen der beiden Transkripte fiel in adulten und fötalen Geweben des Menschen oft unterschiedlich aus. EST-Klon-Analysen deuteten auf eine alternative Polyadenylierung der Transkripte hin.

Der CGGBP1-Promotor wurde durch das Einführen von fortlaufenden Deletionen in die 5'-flankierende Sequenz des menschlichen Gens lokalisiert. Promotortest-Konstrukte mit dem Reportergen Luciferase enthielten zwischen 52 bp und 7,4 kb der 5'-flankierenden genomischen Sequenz des CGGBP1-Gens und wurden in HeLa-Zellen transfiziert. Die Promotorregion der CGGBP1-Gene von Mensch, Maus und Ratte waren mit 87% Sequenzidentität innerhalb einer 600 bp langen Region hoch konserviert. In HeLa-Zellen reichte davon eine 235 bp große Region für eine hohe Promotor-Aktivität aus. Die Reporter-gen-Aktivität veränderte sich nicht, wenn die Promotortest-Konstrukte die hoch konservierte 5'-UTR enthielten.

In Transfektions-Experimenten war der Promotor nur gegenüber der Methylierung aller CpG-Dinukleotide empfindlich. Während die Methylierung von einzelnen HpaII- und HhaI-Erkennungssequenzen die Promotor-Aktivität kaum beeinflusste, inaktivierte eine SssI-Methylierung den CGGBP1-Promotor vollständig. Der Methylierungsstatus von 141 CpG-Dinukleotiden in der 5'-Region des CGGBP1-Gens wurde mittels der Bisulfit-Methode des genomischen Sequenzierens bestimmt. In mononukleären Blutzellen sowie in HeLa-Zellen lag das CGGBP1-Gen komplett unmethyliert vor.

In Gelretardierungs-Experimenten (EMSA) wurde die Bindung von CGGBP1 an solchen 5'-d(CGG)_n-3'-Wiederholungen getestet, die auch in den menschlichen und murinen Sequenzen der FMR1- und CGGBP1-Gene vorkommen. Dabei stellte sich heraus, daß CGGBP1 auch an kürzere 5'-d(CGG)_n-3'-Wiederholungen (n = 5) als bisher beschrieben (n > 8) (Deissler et al., 1996; 1997) binden konnte. Trinukleotid-Folgen aus den murinen FMR1- und CGGBP1-Genen, die durch Trinukleotide anderer Sequenz unterbrochen werden, erwiesen sich als starke Bindungsstellen für das CGGBP1.

Das CGGBP1-Gen ist bei Säugetieren und Vögeln hoch konserviert. Die Transkription des Maus-Gens während der Embryonalentwicklung könnte ebenfalls auf eine wichtige Funktion des CGGBP1 hinweisen. Deshalb wurde hier, basierend auf dem Cre/LoxP-System, eine Strategie entworfen, mit der das CGGBP1-Gen der Maus konditional ausgeschaltet werden kann. Damit wurde die Voraussetzung für die Erforschung des CGGBP1-Gens durch Geninaktivierungs-Experimente geschaffen. Der Geninaktivierungs-Vektor wurde fertiggestellt und mit den Transfektionen in embryonale Stammzellen wurde begonnen. Bisher wurden keine homologen Rekombinanten identifiziert.

Die detaillierten strukturellen Analysen des CGGBP1-Gens des Menschen und der Maus wurden durchgeführt, um eine solide Grundlage für die weitere Erforschung der biologischen Funktion(en) des CGGBP1-Gens zu schaffen. CGGBP1 könnte eine Rolle bei der Regulation von Genen spielen, die 5'-d(CGG)_n-3'-Wiederholungen in ihren Promotor- oder 5'-nicht-translatierten Regionen tragen. In *knock-out*-Mäusen könnte zudem eine mögliche Rolle von CGGBP1 bei der stabilen Vererbung von 5'-d(CGG)_n-3'-Wiederholungen überprüft werden.

G Summary

The amplification of repetitive elements which are located within or close to genes has been recognized as a new mutational mechanism in human disease. Fragile X Syndrome, one of the most frequent causes of heritable mental retardation in humans, is associated with an expanded 5'-(CGG)_n-3' repeat in the 5'-untranslated region of the FMR1 (fragile X mental retardation) gene (Verkerk et al., 1991; Fu et al., 1991). In affected individuals, the gene is completely methylated such that a normal level of the encoded protein is not produced (Pieretti et al., 1991; Sutcliffe et al., 1992).

The function of simple sequence repeats which are common in eukaryotic genomes and frequently conserved among mammals is still unknown. The genetic mechanisms leading to the pathogenic expansions are still under investigation. Strand slippage during replication, repair or recombination has been proposed as a possible mechanism of amplification, because only homogenous repeats are unstable, which can adopt unusual DNA structures *in vitro* (Sinden et al., 2002).

The analysis of proteins which specifically interact with 5'-(CGG)_n-3' repeats might contribute to the elucidation of the function and/or the inheritance of this unstable trinucleotide repeat. A 5'-(CGG)_n-3'-binding protein (CGGBP1) had been purified and cloned in our laboratory (Deissler et al. 1996; 1997). Functional studies showed that CGGBP1 inhibits the FMR1 promoter when it is overexpressed in HeLa cells (Müller-Hartmann et al., 2000).

Here, the genomic organization and expression of the human and the murine CGGBP1 genes were analyzed. The human gene of CGGBP1 consists of 4 exons and had been mapped to chromosome 3p close to the centromere. During this study, a 14 kbp DNA fragment with the murine CGGBP1 gene was isolated from a subgenomic library and the complete nucleotide sequence was determined. The murine gene structure was deduced from sequence similarity to the human gene and was verified by EST clone and RNA analyses. The translated regions as well as large parts of the untranslated regions were highly conserved between the human and the murine genes. Only three amino acids in the murine protein differed from the human sequence. EST clones of the CGGBP1 gene of other mammalian species and the genomic sequence of the CGGBP1 gene of rat were available in GenBank and demonstrated a high degree of sequence similarity to the human gene as well. The translated region of the CGGBP1 genes of human, mouse and rat were not interrupted by introns. The 5'-untranslated region (UTR) of the two rodent genes contained two exons instead of three exons present in the human gene. RT-PCR experiments revealed that the exon of the human gene which did not have an equivalent in the genes of mouse and rat was alternatively spliced in human cells. According to RNA transfer experiments, mRNAs of 1.2 and 4.5 kb were transcribed from the ubiquitously expressed human and murine CGGBP1 genes. The murine gene was transcribed

in embryonic stem cells and during embryonic development. The relative amount of the two transcripts varied in different adult as well as fetal human tissues. EST clone analyses suggested alternative polyadenylation.

The promoter of the human CGGBP1 gene was defined by deletion mapping. Constructs containing DNA-fragments of 7.4 kb to 52 bp from the 5'-flanking region of the CGGBP1 gene upstream of the luciferase gene were transfected into HeLa cells. A region of 600 bp upstream of exon I displayed 87% sequence identity between the human and the two rodent genes. Of this upstream region, a fragment of 235 bp was sufficient for strong promoter activity in HeLa cells. Inclusion of the highly conserved 5'-UTR into the promoter constructs did not influence reporter gene activity.

The promoter was sensitive only to complete CpG methylation. Whereas *in vitro* methylation of HhaI and HpaII sites did not alter the promoter activity, SssI methylation completely inactivated the promoter. The methylation status of 141 CpG dinucleotides in the 5'-region of the CGGBP1 gene was determined by the bisulfite protocol of genomic sequencing. In mononuclear blood cells as well as in HeLa cells, the CGGBP1 gene was completely unmethylated.

Testing short and interrupted 5'-d(CGG)_n-3'-repeats from the human and murine FMR1 and CGGBP1 gene loci in electrophoretic mobility shift assays revealed that CGGBP1 bound to shorter repeats ($n = 5$) than previously reported ($n > 7$) (Deissler et al., 1996; 1997; Müller-Hartmann et al., 1997). The murine sequences of the FMR1 and CGGBP1 genes, harboring short interrupted 5'-(CGG)_n-3' repeats were identified as strong binding sites.

The CGGBP1 gene is highly conserved among mammals and birds. Expression during embryonic development might also hint at an important function of the CGGBP1. Therefore, a conditional strategy of gene inactivation based on the Cre/LoxP recombination system was then developed in order to prepare a future gene targeting experiment in mice. The appropriate targeting construct was completed. First transfection experiments did not yield homologous recombinants.

The detailed structural analyses described here were performed to obtain a solid base for further work on the elucidation of the biological function of the CGGBP1 gene. CGGBP1 might be of interest when studying the transcriptional control of genes harboring 5'-d(CGG)_n-3'-repeats in their promoter and 5'-untranslated regions. A CGGBP1 *knock out* mouse will provide a means to test a possible role of CGGBP1 in the stable inheritance of the trinucleotide repeat.

H Literatur

- Aaronson J. S., Eckman B., Blevins R. A., Borkowski J. A., Myerson J., Imran S. und Elliston K. O. (1996) Toward the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence data. *Genome Res.* 6: 829-845
- Achten S., Behn-Krappa A., Jucker M., Sprengel J., Holker I., Schmitz B., Tesch H., Diehl V., Doerfler W. (1991) Patterns of DNA methylation in selected human genes in different Hodgkin's lymphoma and leukemia cell lines and in normal human lymphocytes. *Cancer Res.* 51: 3702-3709
- Adams M. D., Kelley J. M., Gocayne J. D., Dubnick M., Polymeropoulos M. H., Xiao H., Merrill C. R., Wu A., Olde B., Moreno R. F., Kerlavage A. R., McCombie W. R., Venter J. C. (1991) Complementary DNA-sequencing: expressed sequence tags and Human Genome Project. *Science* 252: 1651-1656
- Akashi H. (1995) Inferring weak selection from patterns of polymorphisms and divergence a „silent“ sites in *Drosophila* DNA. *Genetics* 139: 1067-1076
- Allingham-Hawkins D. J., Babul-Hirji R., Chitayat D., Holden J. J., Yang K. T., Lee C., Hudson R., Gorwill H., Nolin S. L., Glicksman A., Jenkins E. C., Brown W. T., Howard-Peebles P. N., Becchi C., Cummings E., Fallon L., Seitz S., Black S. H., Vianna-Morgante A. M., Costa S. S., Otto P. A., Mingroni-Netto R. C., Murray A., Webb J., Vieri F., et al. (1999) Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study-preliminary data. *Am. J. Med. Genet.* 83: 322-325
- Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. und Lipman D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402
- Antequera F., Boyes J., Bird A. (1990) High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 62: 503-514
- Antequera F. und Bird A. (1993) CpG islands. *EXS* 64: 169-185
- Ashley C. T., Sutcliffe J. S., Kunst C. B., Leiner H. A., Eichler E. E., Nelson D. L. und Warren S. T. (1993) Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nature Genet.* 4: 244-251
- Ashley Jr. C. T. und Warren S. T. (1995) Trinucleotide repeat expansion and human disease. *Ann. Rev. Genet.* 29: 703-728
- Ashley-Koch A. E., Robinson H., Glicksman A. E., Nolin S. L., Schwartz C. E., Brown W. T., Turner G., Sherman S. L. (1998) Examination of factors associated with instability of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 63: 776-785
- Ayoubi T. A., van De Ven W. J. (1996) Regulation of gene expression by alternative promoters. *FASEB J.* 10: 453-460
- Bacolla A., Pradhan S., Larson J. E., Roberts R. J., Wells R. D. (2001) Recombinant human DNA (Cytosine-5) Methyltransferase. *J. Biol. Chem.* 276: 18605-18613

- Bardoni B. und Mandel J.-L. (2002) Advances in understanding of fragile X pathogenesis and FMRP function, and in identification of X linked mental retardation genes. *Curr. Opin. Genet. Dev.* 12: 284-293
- Bauer I., Gencik M., Laccone F., Peters H., Weber B. H., Feder E. H., Weirich H., Morris-Rosendahl D. J., Rolfs A., Gencikova A., Bauer P., Wenning G. K., Epplen J. T., Holmes S. E., Margolis R. L., Ross C. A., Riess O. (2002) Trinucleotide repeat expansions in the junctophilin-3 gene are not found in Caucasian patients with a Huntington's disease-like phenotype. *Ann Neurol.* 51: 662
- Beckmann J. S. und Weber J. L. (1992) Survey of human and rat microsatellites. *Genomics* 12: 627-631
- Bedford M. T., Reed R. und Leder P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc. Natl. Acad. Sci. USA* 95: 10602-10607
- Behn-Krappa A., Hölker I., Sandaradura de Silva U. und Doerfler W. (1991) Patterns of DNA-methylation are indistinguishable in different individuals over a wide range of human DNA sequences. *Genomics* 11: 1-7
- Behn-Krappa A. und Doerfler W. (1994). Enzymatic amplification of synthetic oligodeoxyribonucleotides: implications for triplet repeat expansions in the human genome. *Hum. Mutat.* 3: 19-24
- Benson D. A., Boguski M. S., Lipman D. J., Ostell J., Ouellette B. F. (1998) GenBank. *Nucleic Acids Res.* 26: 1-7
- Berinstein N., Pennell N., Ottaway C. A., Shulman M. J. (1992) Gene replacement with one-sided homologous recombination. *Mol. Cell. Biol.* 12: 360-367
- Bhagwati S., Ghatpande A. und Leung B. (1996) Identification of two proteins which bind to RNA CUG repeats: significance for myotonic dystrophy. *Biochem. Biophys. Res. Com.* 228: 55-62
- Bird A. (1986) CpG rich islands and the function of DNA methylation. *Nature* 321: 209-213
- Bird A. P. und Wolffe A. P. (1999) Methylation-induced repression - belts, braces, and chromatin. *Cell* 99: 451-454
- Blackburn E. H. (1990) Telomeres and their synthesis. *Science* 249: 489-490
- Boardman P. E., Sanz-Ezquerro J., Overton I. M., Burt D. W., Bosch E., Fong W. T., Tickle C., Brown W. R. A., Wilson S. A., Hubbard S. J. (2002) A comprehensive collection of chicken cDNAs. *Curr. Biol.* 12: 1965-1969
- Boguski M. S., Lowe T. M. J., Tolstoshev C. M. (1993) dbEST - database for „expressed sequence tags“ *Nature Genet.* 4: 332-333
- Bonner J. et al. (1973) Functional organization of the mammalian genome. *Cold Spring Harbor Symp. Quant. Bio.* 38: 303-310

- Bontekoe C. J. M., de Graaff E., Nieuwenhuizen I. M., Willemsen R. und Oostra B. A. (1997) FMR1 premutation allele is stable in mice. *Eur. J. Hum. Genet.* 5: 293-298
- Bontekoe C. J., Bakker C. E., Nieuwenhuizen I. M., van der Linde H., Lans H., de Lange D., Hirst M. C., Oostra B. A. (2001) Instability of a (CGG)₉₈ repeat in the Fmr1 promoter. *Hum Mol Genet.* 10: 1693-1699
- Borrmann L., Seebeck B., Rogalla P. und Bullerdiek J. (2003) Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat. *Oncogene* 22: 756-760
- Bradford M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254
- Bradley A. (2002) Mining the mouse genome. *Nature* 420: 512-514
- Brahmachari S. K., Meera G., Sarkar P. S., Balagurumoorthy P., Tripathi J., Raghavan S., Shaligram U., Pataskar S. (1995) Simple repetitive sequences in the genome: structure and functional significance. *Elektrophoresis* 16: 1705-1714
- Brais B., Bouchard J. P., Xie Y. G., Rochefort D. L., Chretien N., Tome F. M. S., Lafreniere R. G., Rommens J. M., Uyama E., Nohira O., Blumen S., Korczyn A. D., Heutink P., Mathieu J., Duranceau A., Codere F., Fardeau M. und Rouleau G. A. (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nature Genet.* 18: 164-167
- Brandon E. P., Idzerda R. L. und McKnight G. S. (1995a) Targeting the mouse genome: a compendium of knockouts (part I). *Curr. Biol.* 5: 625-634
- Brandon E. P., Idzerda R. L. und McKnight G. S. (1995b) Targeting the mouse genome: a compendium of knockouts (part I). *Curr. Biol.* 5: 758-765
- Brandon E. P., Idzerda R. L. und McKnight G. S. (1995c) Targeting the mouse genome: a compendium of knockouts (part I). *Curr. Biol.* 5: 873-881
- Brinkmann B., Klintschar M., Neuhuber F., Hühne J. und Rolf B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* 62: 1408-1415
- Britten R. J. und Kohne D. E. (1968) Repeated sequences in DNA. *Science* 161: 529-540
- Brock G. J. R., Anderson N. H. und Monckton D. G. (1999) Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum. Mol. Genet.* 8: 1061-1067
- Brook J. D., McCurrach M. E., Harley H. G., Bruckler A. J., Church D., Aburatani H., Hunter K., Stanton V. P., Thirion J.-P., Hudson T., Sohn R., Zemelmann B., Snell R.G., Rundle S. A., Crow S., Davies J., Shellbourn P., Buxon J., Jones C., Juvonen V., Johnson K., Harper P. S., Shaw D. J. und Housman D. E. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68: 799-808
- Brown W. T., Houck G. E. Jr., Ding X., Zhong N., Nolin S., Glicksman A., Dobkin C., Jenkins E. C. (1996) Reverse mutations in the fragile X syndrome. *Am. J. Med. Genet.* 64: 287-292

- Brown T. W. (2002) The molecular biology of the fragile X mutation. In Hagerman R. J. und Hagerman P. J. (eds.), Fragile X syndrome: Diagnosis, Treatment and Research, Johns Hopkins University Press, Baltimore, pp. 110-135
- Bucher P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 212: 563-578
- Burman R. W., Popovich B. W., Jacky P. B. und Turker M. S. (1999) Fully expanded FMR1 CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation. *Hum. Mol. Genet.* 8: 2293-2302
- Burman R. W., Anoe K. S., Popovich B. W. (2000) Fragile X full mutations are more similar in siblings than in unrelated patients: further evidence for a familial factor in CGG repeat dynamics. *Genet. Med.* 2: 242-248
- Campuzano V., Montermini L., Molto M. D., Pianese L., Cossee M., Cavalcanti F., Monros E., Rodius F., Duclos F., Monticelli A., Zara F., Canizares J., Koutnikova H., Bidichandani S. I., Gellera C., Brice A., Trouillas P., De Michele G., Filla A. L., De Frutos R., Palau F., Patel P. L., Di Donato S., Mandel J. L., Coccozza S., Koenig M. und Pandolfo M. (1996). Friedreich's Ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion *Science* 271: 1423-1427
- Capecchi M. R. (1989) Altering the genome by homologous recombination. *Science* 244: 1288-1292
- Chen F., MacDonald C. C. und Wilusz J. (1995) Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res.* 23: 2614-2620
- Chen X., Mariappan S. V. S., Moyzis R. K., Bradbury E. M. und Gupta G. (1998) Hairpin induced slippage and hyper-methylation of the Fragile X DNA Triplets. *J. Biomol. Struct. & Dyn.* 15: 745-756
- Church G. M., Gilbert W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81: 1991-1995
- Chiurazzi P., Pomponi M. G., Pietrobono R., Bakker C. E., Neri G., Oostra B. A. (1999) Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the FMR1 gene. *Hum. Mol. Genet.* 8: 2317-2323
- Chiurazzi P., Neri G. (2001) Pharmacological reactivation of inactive genes: the fragile X experience. *Brain. Res. Bull.* 56: 383-387
- Chomczynski P. (1992) One hour downward alkaline capillary transfer for blotting of DNA or RNA. *Anal. Biochem.* 201: 134-139
- Clark S. J., Harrison J., Paul C. L., Frommer M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22: 2990-2997
- Cleary J. D., Nichol K., Wang Y.-H. und Pearson C. E. (2002) Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet.* 31: 37-46

- Coffee B., Zhang F., Warren S. T. und Reines D. (1999) Acetylated histones are associated with FMR1 in normal but not in fragile X-syndrome cells. *Nature Genet.* 22: 98-101
- Coffee B., Zhang F., Ceman S., Warren S. T., Reines D. (2002) Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile X syndrome. *Am. J. Hum. Genet.* 71: 923-932
- Corbin F., Bouillon M., Fortin A., Morin S., Rousseau F. und Khandjian E. W. (1997) The fragile X mental retardation protein is associated with poly(A)⁺ mRNA in actively translating polyribosomes. *Hum. Mol. Genet.* 6: 1465-1472
- Cox R. und Mirkin S. M. (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl. Acad. Sci. USA* 94: 5237-5242
- Crawford D. C., Schwartz C. E., Meadows K. L., Newman J. L., Taft L. F., Gunter C., Brown W. T., Carpenter N. J., Howard-Peebles P. N., Monaghan K. G., Nolin S. L., Reiss A. L., Feldman G. L., Rohlfis E. M., Warren S. T., Sherman S. L. (2000) Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and single-nucleotide-polymorphism haplotypes in an African American population. *Am. J. Hum. Genet.* 66: 480-493
- Crawford D. C., Acuna J. M., Sherman S. L. (2001) FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet. Med.* 3: 359-371
- Cummings C. J. und Zoghbi H. Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* 9: 909-916
- Curtis D., Lehmann R., Zamore P. D. (1995) Translational regulation in development. *Cell* 81: 171-178
- Darlow J. M. und Leach D. R. F. (1998) Secondary structures in d(CGG) and d(CCG) repeat tracts. *J. Mol. Biol.* 275: 3-16
- Darnell J. C., Jensen K. B., Jin P., Brown V., Warren S. T., Darnell R. B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell.* 107: 489-499
- David G., Abbas N., Stevanin G., Durr A., Yvert G., Cancel G., Weber C., Imbert G., Saudou F., Antoniou E., Drabkin H., Gemmill R., Giunti P., Benomar A., Wood N., Ruberg M., Agid Y., Mandel J.-L. und Brice A. (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature Genet.* 17: 65-70
- Davidson E. H., Hough B. R., Amenson C. S. und Britten R. J. (1973) General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* 77: 1-23
- Davis B. M., McCurrach M. E., Taneja K.L., Singer R. H., Housman D. E. (1997) Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. U S A.* 94: 7388-7393

- DeBouille K., Verkerk A. J. M. H., Reyniers E., Vits L., Hendrickx J., Van Roy B., Van Den Bos F., de Graaff E., Oostra B. A. und Willems P. J. (1993) A point mutation in the FMR1 gene associated with fragile X mental retardation. *Nature Genet.* 3: 31-35
- Debrauwere H., Gendrel C. G., Lechat S., Dutreix M. (1997) Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. *Biochimie* 79: 577-586
- Délot E., King L. M., Briggs M. D., Wilcox W. R., Cohn D. H. (1999) Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene. *Hum. Mol. Genet.* 8: 123-128.
- Deissler H., Behn-Krappa A. und Doerfler W. (1996) Purification of nuclear proteins from human HeLa cells that bind specifically to the unstable tandem repeat (CGG)_n in the human FMR1 gene. *J. Biol. Chem.* 271: 4327-4334
- Deissler H., Wilm M., Genç B., Schmitz B., Ternes T., Naumann F., Mann M. und Doerfler W. (1997) Rapid protein sequencing by tandem mass spectrometry and cDNA cloning of p20-CGGBP. *J. Biol. Chem.* 272: 16761-16768
- Deng C., Capecchi M. R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* 12: 3365-3371
- Dibb N. J. (1993) Why do genes have introns? *FEBS Letters* 325: 135-139
- Djian P., Hancock J. M. und Chana H. S. (1996) Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. *Proc. Natl. Acad. Sci. USA* 93: 417-421
- Doerfler W. (1981) DNA methylation - a regulatory signal in eukaryotic gene expression. *J. Gen. Virol.* 57: 1-20
- Doerfler W. (1983) DNA methylation and gene activity. *Annu. Rev. Biochem.* 52: 93-124
- Doerfler W., Toth M., Kochanek S., Achten S., Freisem-Rabien U., Behn-Krappa A., Orend G. (1990) Eukaryotic DNA methylation: facts and problems. *FEBS Lett.* 268: 329-333
- Doerfler W. (1993) Pattern of *de novo* DNA methylation and promoter inhibition: Studies on the adenovirus and the human genomes. In Jost J. P. und Saluz H. P. (eds.) *DNA methylation: Molecular biology and biological significance*. Birkhäuser Verlag, Basel
- Doerfler W. (1996) A new concept in (adenoviral) oncogenesis: integration of foreign DNA and its consequences. *Biochim. Biophys. Acta* 1288: F79-99
- Doerfler W., Remus R., Muller K., Heller H., Hohlweg U., Schubbert R. (2001) The fate of foreign DNA in mammalian cells and organisms. *Dev. Biol. (Basel)* 106: 89-97; discussion 143-60
- Dombrowski C., Lévesque S., Morel M. L., Rouillard P., Morgan K. und Rousseau F. (2002) Premutation and intermediate-size FMR1 alleles in 10,572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* 11: 371-378

- Doolittle W. F. und Sapienza C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603
- Duan J., Wainwright M. S., Comeron J. M., Saitou N., Sanders A. R., Gelernter J., Gejman P. V. (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* 12: 205-216
- Dugaiczuk A., Boyer H. W., Goodman H. M. (1975) Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* 96: 171-184
- Duret L., Dorkeld F. und Gautier C. (1993) Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acids Res.* 21: 2315-2322
- Duret L., Bucher P. (1997) Searching for regulatory elements in human noncoding sequences. *Curr. Opin. Struct. Biol.* 7: 399-406
- Duval A., Hamelin R. (2002a) Genetic instabilities in human mismatch repair deficient cancers *Ann. Genet.* 45: 71-75
- Duval A., Hamelin R. (2002b) Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.* 62: 2447-2454
- Edwards-Gilbert G., Veraldi K. L., Milcarek C. (1997) Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* 25: 2547-2561
- Eichler E. E., Holden J. J. A., Popovich B. W., Reiss A. L., Snow K., Thibodeau S. N., Richards C. S., Ward P. A. und Nelson D. L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.* 8: 88-94
- Eichler E. E., Kunst C. B., Lugenbeel K. A., Ryder O. A., Davison D., Warren S. T. und Nelson D. L. (1995) Evolution of the cryptic FMR1 CGG repeat. *Nature Genet.* 11: 301-308
- El-Osta A. und Wolffe A. P. (2000) DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease. *Gene Expression* 9: 63-75
- Ellegren H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genet.* 16: 551-558
- Epplen J. T., Kyas A., Mäueler W. (1996) Genomic simple repetitive DNAs are targets for differential binding fo nuclear proteins. *FEBS Letters* 389: 92-95
- Evans M. J. und Kaufmann M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292: 154-156
- Feinberg A. P. und Vogelstein B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301: 89-91
- Feinberg A. P. (2001) Methylation meets genomics. *Nat. Genet.* 27: 9-10

- Feng Y., Zhang F., Lokey L. K., Chastein J. L., Lakkis L., Eberhard D. und Warren S. T. (1995) Translational suppression by trinucleotide repeat expansion at FMR1. *Science* 268: 731-734
- Fisch G. S., Snow K., Thibodeau S. N., Chalifaux M., Holden J. J., Nelson D. L., Howard-Peebles P. N., Maddalena A. (1995) The fragile X premutation in carriers and its effect on mutation size in offspring. *Am. J. Hum. Genet.* 56: 1147-1155
- Fortune M. T., Vassilopoulos C., Coolbaugh M. I., Siciliano M. J. und Monckton D. G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.* 9: 439-445
- Freshney R. I. (1987) Culture of animal cells. A manual of basic technique, 2. edition. Alan R. Liss, New York
- Frommer M., McDonald L. E., Millar D. S., Collis C. M., Watt F., Grigg G. W., Molloy P. L., Paul C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* 89: 1827-1831
- Fry M., Loeb L. A. (1994) The fragile X syndrome d(CGG)_n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U.S.A.* 91: 4950-4954
- Fu Y. H., Kuhl D. P. A., Pizzuti A., Pieretti M., Sucliffe J. S., Richards S., Verkerk A. J. M. H., Holden J. J. A., Fenwick Jr R.G., Warren S. T., Oostra B. A., Nelson D. L. und Caskey T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the sherman paradox. *Cell* 67: 1047-1058
- Galvao R., Mendes-Soares L., Camara J., Jaco I., Carmo-Fonseca M. (2001) Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis. *Brain Res. Bull.* 56: 191-201
- Gautheret D., Poirot O., Lopez F., Audic S. und Claverie J.-M. (1998) Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering. *Genome Res.* 8: 524-530
- GCG Package version 10.x Genetics Computer Group (GCG), Madison, Wisc., USA
- Genç B., Müller-Hartmann H., Zeschnigk M., Deissler H., Schmitz B., Majewski F., von Gontard A. und Doerfler W. (2000) Methylation mosaicism of 5'-d(CGG)_n-3' repeats in fragile X, premutation and healthy individuals. *Nucleic Acids Res.* 28: 2141-2152.
- Gerber H. P., Seipel K., Georgiev O., Hofferer M., Hug M., Rusconi S. und Schaffner W. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263: 808-811
- Gläser D., Wöhrle D., Salat U., Vogel W. und Steinbach P. (1999) Mitotic behaviour of expanded CGG repeats studied on cultured cells: further evidence for methylation-mediated triplet repeat stability in fragile X-syndrome. *Am. J. Med. Genet.* 84: 226-228
- Glover T. W. (1998) Instability at chromosomal fragile sites. *Recent Results Cancer Res.* 154: 185-199

- Godde J. S., Kass S. U., Hirst M. C. und Wolffe A. P. (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promotor. *J. Biol. Chem.* 271: 24325-24328
- Goffeau A., Barrell B. G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galivert F., Hoheisel J. D., Jacq D., Johnston J., Louis E. J., Mewes H. W., Murakami Y., Phillipsen P., Tettelin H., Oliver S. G. (1996) Life with 6000 Genes. *Science* 274: 546-567
- Gordenin D. A., Kunkel T. A., Resnick M. A. (1997) Repeat expansion--all in a flap? *Nat. Genet.* 16: 116-118
- Gossen M., Bujard H. (2002) Studying gene function in eukaryotes by conditional gene inactivation. *Annu. Rev. Genet.* 36: 153-173
- Gossler A., Doetschman T., Korn R., Serfling E. und Kemler R. (1986) Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc. Natl. Acad. Sci. USA* 83: 9065-9069
- Grabczyk E., Kumari D., Usdin K. (2001) Fragile X syndrome and Friedreich's ataxia: two different paradigms for repeat induced transcript insufficiency. *Brain Res. Bull.* 56: 367-373
- Gray N. K., Hentze M. W. (1994) Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.* 19: 195-200
- Greco C. M., Hagerman R. J., Tassone F., Chudley A. E., Del Bigio M. R., Jacquemont S., Leehey M., Hagerman P. J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* 125: 1760-1771
- Gu H., Zou Y. R., Rajewsky K. (1993) Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* 73: 1155-1164
- Gu H., Marth J. D., Orban P. C., Mossmann H., Rajewsky K. (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265: 103-106
- Gumucio D. L., Shelton D. A., Zhu W., Millinoff D., Gray T., Bock J. H., Slightom J. L., Goodman M. (1996) Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the beta-like globin genes. *Mol. Phylogenet. Evol.* 5: 18-32
- Gunter C., Paradee W., Crawford D. C., Meadows K. A., Newman J., Kunst C. B., Nelson D. L., Schwartz C., Murray A., Macpherson J. N., Sherman S. L., Warren S. T. (1998) Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum. Mol. Genet.* 7: 1935-1946
- Haberer K. (1966) Measurements of beta activities in aqueous samples utilizing Cerenkov radiation. *Packard Technical Bulletin* 16

- Hagerman R. J., Hull C. E., Safanda J. F., Carpenter I., Staley L. W., O'Connor R. A., Seydel C., Mazzocco M. M. M., Snow K., Thibodeau S. N., Kuhl D., Nelson D. L., Caskey C. T. and Taylor A. K. (1994) High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *Am. J. Med. Genet.* 51: 298-308
- Hagerman R. J., Leehey M., Heinrichs W., Tassone F., Wilson R., Hills J., Grigsby J., Gage B., Hagerman P. J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57: 127-130
- Hagerman R. J. (2002) The physical and behavioral phenotype. In Hagerman R. J. und Hagerman P. J. (eds.), *Fragile X syndrome: Diagnosis, Treatment and Research*, Johns Hopkins University Press, Baltimore, pp. 70-73
- Hagerman R. J. und Hagerman P. J. (2002) The fragile X premutation: into the phenotypic fold. *Curr. Opin. Genet. Dev.* 12: 278-283
- Hamada H., Seidman M., Howard B. H., Gorman C. M. (1984) Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Mol. Cell. Biol.* 4: 2622-2630
- Hammond L. S., Macias M. M., Tarleton J. C. und Pai G. S. (1997) Fragile X syndrome and deletions in FMR1: new case and review of the literature. *Am. J. Med. Genet.* 72: 430-434
- Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166: 557-580
- Harr B., Schlötterer C. (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* 155: 1213-1220
- Harris S., Moncrieff C., Johnson K. (1996) Myotonic dystrophy: will the real gene please step forward! *Hum Mol Genet.* 5: 1417-1423
- Hartl D. L. (2000) Molecular melodies in high and low C. *Nature Rev. Genet.* 1: 145-149
- Hasty P., Rivera-Perez J., Chang C., Bradley A. (1991) Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.* 11: 4509-4517
- Hayatsu H. (1976) Bisulfite modification of nucleic acids and their constituents. *Prof. Nucleic Acid Res. Mol. Biol.* 16: 75-124
- Hertz J. M., Schell G., Doerfler W. (1999) Factors affecting *de novo* methylation of foreign DNA in mouse embryonic stem cells. *J. Biol. Chem.* 274: 24232-24240
- Hillier L., Lennon G., Becker M., Bonaldo M. F., Chiapelli B., Chissoe S., Dietrich N., DuBuque T., Favello A., Gish W., Hawkins M., Hultman M., Kucaba T., Lacy M., Le M., Le N., Mardis E., Moore B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T., Schellenberg K., Soares M. B., Tan F., Thierry-Meg J., Trevaskis E., Underwood K., Wohldman P., Waterston R., Wilson R. und Marra M. (1996) Generation and analysis of 280.000 human expressed sequence tags. *Genome Res.* 6: 807-828

- Hinds H. L., Ashley C. T., Sutcliffe J. S., Nelson D. L., Warren S. T., Housman D. E. und Schalling M. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nature Genet.* 3: 36-43
- Hirst M. C. (1995) FMR1 triplet arrays: paying the price for perfection. *J. Med. Genet.* 32: 761-763
- Hirst M. C., White P. J. (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of *in vivo* lagging strand secondary structure. *Nucleic Acids Res.* 26: 2353-2358
- Hoess R. H., Wierzbicki A., Abremski K., (1986) The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res.* 14: 2287-2300
- Hofmann K., Bucher P., Falquet L. und Bairoch A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27: 215-219
- Holmes D. S., Quigley M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114: 193-197
- Holmes S. E., O'Hearn E. E., McInnis M. G., Gorelick-Feldman D. A., Kleiderlein J. J., Callahan C., Kwak N. G., Ingersoll-Ashworth R. G., Sherr M., Sumner A. J., Sharp A. H., Ananth U., Seltzer W. K., Boss M. A., Viera-Saecker A. M., Epplen J. T., Riess O., Ross C. A., Margolis R. L. (1999) Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat. Genet.* 23: 391-392
- Holmes S. E., O'Hearn E., Rosenblatt A., Callahan C., Hwang H. S., Ingersoll-Ashworth R. G., Fleisher A., Stevanin G., Brice A., Potter N. T., Ross C. A., Margolis R. L. (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat. Genet.* 29: 377-378
- Hornstra I. K., Nelson D. L., Warren S. T. und Yang T. P. (1993) High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Hum. Mol. Genet.* 2: 1659-1665
- Hubbard T., Barker D., Birney E., Cameron G., Chen Y., Clark L., Cox T., Cuff J., Curwen V., Down T., Durbin R., Eyras E., Gilbert J., Hammond M., Huminiecki L., Kasprzyk A., Lehvaslaiho H., Lijnzaad P., Melsopp C., Mongin E., Pettett R., Pocock M., Potter S., Rust A., Schmidt E., Searle S., Slater G., Smith J., Spooner W., Stabenau A., Stalker J., Stupka E., Ureta-Vidal A., Vastrik I., Clamp M. (2002) The Ensembl genome database project. *Nucleic Acids Res.* 30: 38-41
- Hundscheid R. D. L., Siermans E. A., Thomas C. M. G., Braat D. D. M., Straatman H., Kiemeneij L. A. L. M., Oostra B. A. und Smits A. P. T. (2000) Imprinting effect in premature ovarian failure confined to paternally inherited fragile X premutations. *Am. J. Hum. Genet.* 66: 413-418
- Imagawa M., Ishikawa Y., Shimano J., Osada S. und Nishihara T. (1995) CTG triplet repeat in mouse growth inhibitory factor/metallothionein III gene promoter represses the transcriptional activity of the heterologous promoters. *J. Biol. Chem.* 270: 20898-20900

- Imbert G., Kretz C., Johnson K., Mandel J. L. (1993) Origin of the expansion mutation in myotonic dystrophy. *Nat. Genet.* 4: 72-76
- Imbert G., Saudou F., Yvert G., Devys D., Trottier Y., Garnier J. M., Weber C., Mandel J. L., Cancel G., Abbas N., Durr A., Didierjean O., Stevanin G., Agid Y., Brice A. (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat. Genet.* 14: 285-291
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921
- Ionov Y., Peinado M. A., Malkhosyan S., Shibata D. und Perucho M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363: 558-561
- Iseli C., Stevenson B. J., de Souza S. J., Samaia H. B., Camargo A. A., Buetow K. H., Strausberg R. L., Simpson A. J. G., Bucher P. und Jongeneel C. V. (2002) Long-range heterogeneity at the 3'-ends of human mRNAs. *Genome Res.* 12: 1068-1074
- Iuchi S. (2001) Three classes of C2H2 zinc finger proteins. *Cell. Mol. Life Sci.* 58: 625-635
- Iyer R. R., Wells R. D. (1999) Expansion and deletion of triplet repeat sequences in *Escherichia coli* occur on the leading strand of DNA replication. *J. Biol. Chem.* 274: 3865-3877
- Jackson R. J. (1993) Cytoplasmic regulation of mRNA function: the importance of the 3'-untranslated region. *Cell* 74: 9-14
- Jaenisch R. und Bird A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet. Suppl.* 33: 245-154
- Jansen R.-P. (2001) mRNA localization: message on the move. *Nature Rev. Mol. Cell Biol.* 2: 247-256
- Jeffreys A.J., Turner M., Debenham P. (1991) The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *Am. J. Hum. Genet.* 48: 824-840
- Jeffreys A. L., Tamaki K., MacLeod A., Monckton D. G., Neil D. L., und Armour J. A. L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* 6: 136-145
- Jin P., Warren S. T. (2000) Understanding the molecular basis of fragile X syndrome. *Hum. Mol. Genet.* 9: 901-908
- Jones C., Penny L., Mattina T., Yu S., Baker E., Voullaire L., Langdon W. Y., Sutherland G. R., Richards R. I. und Tunnacliffe A. (1995) Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* 376: 145-149
- Joyner (1993) Gene targeting. A practical approach.
- Jurka J. (1998) Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* 8: 333-337

- Kaiser A. D. und Hogness D. S. (1960) The transformation of *Escherichia coli* with deoxyribonucleic acid isolated from bacteriophage λ dg. *J. Mol. Biol.* 2: 392-415
- Kang S., Ohshima K., Shimizu M., Amirhaeri S. und Wells R. D. (1995) Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J. Biol. Chem.* 270: 27014-27021
- Kashi Y., King C., Soller M. (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genet.* 13: 74-78
- Katti M. V., Ranjekar P. K., Gupta V. S. (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* 18: 1161-1167
- Kawaguchi Y., Okamoto T., Taniwaki M., Aizawa M., Inoue M., Katayama S., Kawakami H., Nakamura S., Nishimura M., Akiguchi I. et al. (1994) CAG expansion in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genet.* 8: 221-228
- Kenneson A., Zhang F., Hagedorn C. H., Warren S. T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.* 10: 1449-1454
- Kilby N. J., Snaith M. R., Murray J. A. (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet.* 9: 413-421
- Klausner R. D., Rouault T. A., Harford J. B. (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* 72: 19-28
- Knight S. J. L., Flannery A. V., Hirst M. C., Campbell L., Christodoulou Z., Phelps S. R., Pointon J., Middleton-Price H. R., Barnicoat A., Pembrey M. E., Holland J., Oostra B. A., Bobrow M. und Davies K. E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 74: 127-134
- Kochanek S., Toth M., Dehmel A., Renz D., Doerfler W. (1990) Interindividual concordance of methylation profiles in human genes for tumor necrosis factors alpha and beta. *Proc. Natl. Acad. Sci. USA* 87: 8830-8834
- Kochanek S., Radbruch A., Tesch H., Renz D., Doerfler W. (1991) DNA methylation profiles in the human genes for tumor necrosis factors alpha and beta in subpopulations of leukocytes and in leukemias. *Proc. Natl. Acad. Sci. USA* 88: 5759-5763
- Kochetov A. V., Ischenko I. V., Vorobiev D. G., Kel A. E., Babenko V. N., Kisselev L. L., Kolchanov N. A. (1998) Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features. *FEBS Lett.* 440: 351-355
- Koetsier P. A., Schorr J. und Doerfler W. (1993) A rapid optimized protocol for downward alkaline Southern blotting of DNA. *BioTechniques* 15: 260-261
- Koide R., Ikeuchi T., Onodera O., Tanaka H., Igarashi S., Endo K., Takahashi H., Kondo R., Ishikawa A., Hayashi T., Saito M., Tomoda A., Miike T., Naito H., Ikuta F. und Tsuji S. (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nature Genet.* 6: 9-13

- Köntgen F., Süß G., Stewart C., Steinmetz M. und Bluethmann H. (1993) Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. *International Immunol.* 5: 957-964
- Koob M. D., Moseley M. L., Schut L. J., Benzow K. A., Bird T. D., Day J. W., Ranum L. P. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8) *Nat. Genet.* 21: 379-384
- Kooy F. R. (2003) Of mice and the fragile X syndrome. *Trends Genet.* 19: 148-154
- Kornberg A., Bertsch L.-R. L., Jackson J. F., Khorana H. G. (1964) Enzymatic synthesis of deoxyribonucleic acid, XVI. Oligonucleotides as templates and the mechanisms of their replication. *Proc. Natl. Acad. Sci. USA* 51: 315-323
- Kornberg A. (1965) Synthesis of DNA-like polymers *de novo* or by reiterative replication. In Bryson B., Vogel H. J. (Herausgeber): *Evolving genes and proteins*. New York: Academic Press, 403-417.
- Koshy B. T. und Zoghbi H. Y. (1997) The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis. *Brain Pathology* 7: 927-942
- Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15: 8125-8148
- Kozak M. (1989) The Scanning Model for Translation: An Update. *J. Cell. Biol.* 108: 229-241
- Kozak M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* 266: 19867-19870
- Kozak M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* 7: 563-574
- Kühn R., Schwenk F., Aguet M., Rajewsky K. (1995) Inducible gene targeting in mice. *Science* 269: 1427-1429
- Kühn R., Schwenk F. (1997) Advances in gene targeting methods. *Curr. Opin. Immunol.* 9: 183-188
- Künzler P., Matsuo K., Schaffner W. (1995) Pathological, physiological and evolutionary aspects of short unstable DNA repeats in the human genome. *Biol. Chem. Hoppe-Seyler* 376: 201-211
- Kunst C. B. und Warren S. T. (1994) Cryptic and polar variation of the Fragile X repeat could result in predisposing normal alleles. *Cell* 77: 853-861
- Lafreniere R. G., Rochefort D. L., Chretien N., Rommens J. M., Cochius J. I., Kalviainen R., Nousiainen U., Patry G., Farrell K., Soderfeldt B., Federico A., Hale B. R., Cossio O. H., Sorensen T., Pouliot M. A., Kmiec T., Uldall P., Janszky J., Pranzatelli M. R., Andermann F., Andermann E., Rouleau G. A. (1997) Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat. Genet.* 15: 298-302
- Lagercrantz U., Ellegren H., Andersson L. (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 21: 1111-1115

- Lantto J., Ohlin M. (2002) Uneven distribution of repetitive trinucleotide motifs in human immunoglobulin heavy variable genes. *J. Mol. Evol.* 54: 346-353
- La Spada A. R., Roling D. B., Harding A. E., Warner C. L., Spiegel R., Hausmanowa-Petrusewicz I., Yee W. C. und Fischbeck K. H. (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet.* 2: 301-304
- La Spada A. R. (1997) Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathology* 7: 943-963
- Lavedan C. N., Garrett L., Nussbaum R. L. (1997) Trinucleotide repeats (CGG)₂₂TGG (CGG)₄₃TGG(CGG)₂₁ from the fragile X gene remain stable in transgenic mice. *Hum. Genet.* 100: 407-414
- Levinson G. und Gutman G. A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203-221
- Lewandoski M. (2001) Conditional control of gene expression in the mouse. *Nat. Rev. Genet.* 2: 743-755
- Lipshitz H. D. und Smibert C. A. (2000) Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genet. Dev.* 10: 476-488
- Liquori C., Ricker K., Moseley M. L., Jacobsen J. F., Kress W., Naylor S., Day J. W., Ranum L. P. W. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293: 864-867
- Loonstra A., Vooijs M., Beverloo H. B., Allak B. A., van Drunen E., Kanaar R., Berns A., Jonkers J. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl. Acad. Sci. USA* 98: 9209-9214
- Lorenzetti D., Watase K., Xu B., Matzuk M. M., Orr H. T., Zoghbi H. Y. (2000) Repeat instability and motor incoordination in mice with a targeted expanded CAG repeat in the *Scal* locus. *Hum. Mol. Genet.* 9: 779-785
- Love J. M., Knight A. M., McAleer M. A. und Todd J. A. (1990) Towards construction of a high resolution map of the mouse genome using PCR-analyzed microsatellites. *Nucleic Acids Res.* 18: 4123-4130
- Lugenbeel K. A., Peier A. M., Carson N. L., Chudley A. E. und Nelson D. L. (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in the fragile X syndrome. *Nature Genet.* 10: 483-485
- Maddalena A., Richards C. S., McGinnis M. J., Brothman A., Desnick R. J., Grier R. E., Hirsch B., Jacky P., McDowell G. A., Popovich B., Watson M., Wolff D. J. (2001) Technical standards and guidelines for fragile X: the first of a series of disease-specific supplements to the standards and guidelines for clinical genetics laboratories of the American College of Medical Genetics. *Genet. Med.* 3: 200-205

- Mäueler W., Bassili G., Epplen C., Keyl H. G., Epplen J. T. (1999) Protein binding to simple repetitive sequences depends on DNA secondary structure(s). *Chromosome Res.* 7: 163-166
- Mäueler W., Bassili G., Hardt C., Keyl H. G., Epplen J. T. (2001) A complex containing at least one zinc dependent HeLa nuclear protein binds to the intronic (gaa)(n) block of the frataxin gene. *Gene* 270: 131-143
- Majewski J., Ott J. (2002) Distribution and characterization of regulatory elements in the human genome. *Genome Res.* 12: 1827-1836
- Malter H. E., Iber J. C., Willemsen R., de Graaff E., Tarleton J. C., Leisti J., Warren S. T. und Oostra B. A. (1997) Characterization of the full fragile X syndrome mutation in fetal gametes. *Nature Genet.* 15: 165-169
- Mangel L., Ternes T., Schmitz B. und Doerfler W. (1998) New 5'-(CGG)_n-3' repeats in the human genome. *J. Biol. Chem.* 273: 30466-30471
- Mangiarini L., Sathasivam K., Mahal A., Mott R., Seller M., Bates G. P. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.* 15: 197-200
- Mansour S. L., Thomas K. R., Capecchi M. R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336: 348
- Margolis R. L., McInnis M. G., Rosenblatt A., Ross C. A. (1999) Trinucleotide repeat expansion and neuropsychiatric disease. *Arch. Gen. Psychiatry* 56: 1019-1031
- Margolis R. L. und Ross C. A. (2001) Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends Mol. Med.* 7: 479-482
- Martin J. P. und Bell J. (1943) A pedigree of mental defect showing sex-linkage. *J. Neurol. Psychiatry* 6: 154-157
- Matsushima Y., Matsumura K., Kitagawa Y. (1997) Zinc finger-like motif conserved in a family of RNA binding proteins. *Biosci. Biotechnol. Biochem.* 61: 905-906
- Matys V., Fricke E., Geffers R., Gossling E., Haubrock M., Hehl R., Hornischer K., Karas D., Kel A. E., Kel-Margoulis O. V., Kloos D. U., Land S., Lewicki-Potapov B., Michael H., Munch R., Reuter I., Rotert S., Saxel H., Scheer M., Thiele S., Wingender E. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31: 374-378
- McDonnell M. W., Simon M. N., Studier F. W. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* 110: 119-146
- McLaughlin B. A., Spencer C., Eberwine J. (1996) CAG trinucleotide RNA repeats interact with RNA-binding proteins. *Am. J. Hum. Genet.* 59: 561-569
- McMurray C. T. (1995) Mechanisms of DNA expansion. *Chromosoma* 104: 2-13

- McMurray C. T. (1999) DNA secondary structure: A common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. USA* 96: 1823-1825
- Meijer H. A. und Thomas A. A. M. (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem. J.* 367: 1-11
- Metzgar D., Bytof J. und Wills C. (2000) Selection against frameshift mutations limits microsatellites expansion in coding DNA. *Genome Res.* 10: 72-80
- Mironov A. A., Fickett J. W., Gelfand M. S. (1999) Frequent alternative splicing of human genes. *Genome Res.* 9: 1288-1293
- Mirsky A. E. und Ris H. (1951) The DNA content of animal cells, and its evolutionary significance. *J. Gen. Physiol.* 34: 451-462
- Mirkin S. M. und Smirnova E. V. (2002) Positioned to expand. *Nature Genet.* 31: 5-6
- Mitas M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* 25: 2245-2253
- Mitchell P., Tollervey D. (2001) mRNA turnover. *Curr. Opin. Cell. Biol.* 13: 320-325
- Mize G. J., Ruan H., Low J. J. und Morris D. R. (1998) The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J. Biol. Chem.* 273: 32500-32505
- Modrek B., Lee C. (2002) A genomic view of alternative splicing. *Nature Genet.* 30:13-19
- Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562
- Moutou C., Vincent M.-C., Biancalana V. und Mandel J.-L. (1997) Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. *Hum. Mol. Genet.* 6: 971-979
- Müller K., Heller H., Doerfler W. (2001) Foreign DNA integration. Genome-wide perturbations of methylation and transcription in the recipient genomes. *J. Biol. Chem.* 276: 14271-14278
- Müller W., Kühn R., Rajewsky K. (1991) Major histocompatibility complex class II hyperexpression on B cells in interleukin 4-transgenic mice does not lead to B cell proliferation and hypergammaglobulinemia. *Eur. J. Immunol.* 21: 921-925
- Müller U. (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech. Dev.* 82: 3-21
- Müller W., Kühn R. und Rajewsky K. (1991) Major histocompatibility complex class II hyperexpression on B cells in interleukin 4 transgenic mice does not lead to B cell proliferation and hypergammaglobulinemia. *Eur. J. Immunol.* 21: 921-925
- Müller-Hartmann H., Deissler H., Naumann F., Schmitz B., Schröer J. und Doerfler W. (2000) The human 20 kDa 5'-(CGG)_n-3'-binding protein is targeted to the nucleus and affects the activity of the FMR1-promotor. *J. Biol. Chem.* 275: 6447-6452

- Müller-Hartmann H. (2000) Das menschliche 5'-(CGG)-3'-bindende Protein p20 ist kernlokalisiert, inhibiert den Promotor des Fragilen-X-Gens und bindet an vielen anderen Stellen im Genom. Dissertation, Köln Muiznieks I., Doerfler W. (1994) The impact of 5'-CG-3' methylation on the activity of different eukaryotic promoters: a comparative study. *FEBS Let.* 344: 251-254
- Mulder N. J., Apweiler R., Attwood T. K., Bairoch A., Barrell D., Bateman A., Binns D., Biswas M., Bradley P., Bork P., Bucher P., Copley R. R., Courcelle E., Das U., Durbin R., Falquet L., Fleischmann W., Griffiths-Jones S., Haft D., Harte N., Hulo N., Kahn D., Kanapin A., Krestyaninova M., Lopez R., Letunic I., Lonsdale D., Silventoinen V., Orchard S. E., Pagni M., Peyruc D., Ponting C. P., Selengut J. D., Servant F., Sigrist C. J., Vaughan R., Zdobnov E. M. (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* 31: 315-318
- Mundlos S., Otto F., Mundlos C., Mulliken J. B., Aylsworth A. S., Albright S., Lindhout D., Cole W. G., Henn W., Knoll J. H., Owen M. J., Mertelsmann R., Zabel B. U., Olsen B. R. (1997) Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* 89: 773-779
- Munnes M., Doerfler W. (1997) DNA methylation in mammalian genomes: promoter activity and genetic imprinting. In *Encyclopedia of Human Biology*, second edition, Vol. 3; Academic Press
- Munnes M., Patrone G., Schmitz B., Romeo G. und Doerfler W. (1998) A 5'-CG-3'-rich region in the promoter of the transcriptionally frequently silenced RET protooncogene lacks methylated cytidine residues. *Oncogene* 17: 2573-2583
- Muragaki Y., Mundlos S., Upton J., Olsen B. R. (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272: 548-551
- Nagy A., Mar L. (2001) Creation and use of a Cre recombinase transgenic database. *Methods Mol. Biol.* 158: 95-106
- Nakamoto M., Nakano S., Kawashima S., Ihara M., Nishimura Y., Shinde A., Kakizuka A. (2002) Unequal crossing-over in unique PABP2 mutations in Japanese patients: a possible cause of oculopharyngeal muscular dystrophy. *Arch. Neurol.* 59: 474-477
- Nakamura K., Jeong S. Y., Uchihara T., Anno M., Nagashima K., Nagashima T., Ikeda S., Tsuji S., Kanazawa I. (2001) SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum. Mol. Genet.* 10: 1441-1448
- Nancarrow J. K., Kremer E., Holman K., Eyre H., Doggett N. A., Le Paslier D., Callen E. F., Sutherland G. R., Richards R. I. (1994) Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. *Science* 264: 1938-1941
- Nathans D. und Smith H. O. (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. *Ann. Rev. Biochem.* 44: 273-293
- Naumann F. (1998) Genomische Organisation des Gens für das 5'-d(CGG)_n-3'-bindende Protein p20 des Menschen. Diplomarbeit, Universität zu Köln

- Naumann F., Müller-Hartmann H., Deissler H., Doerfler W. (2001) On the function of the CGG-binding protein. *Gene Funct. Dis.* 2: 89-94
- Nelissen R. L. H., Heinrichs V., Habets W. J., Simons F., Lührmann R. und van Venrooij W. J. (1991) Zin finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res.* 19: 449-454
- Nelson D. L., Warren S. T. (1993) Trinucleotide repeat instability: when and where? *Nat. Genet.* 4: 107-108
- Neville C. E., Mahadevan M. S., Barcelo J. M., Korneluk R. G. (1994) High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum. Mol. Genet.* 3: 45-51
- Nichol K. und Pearson C. E. (2002) CpG methylation modifies the genetic stability of cloned repeat sequences. *Genome Res.* 12: 1246-1256
- Nicholas K. B., Nicholas H. B. Jr. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments.
- Nolin S. L., Glicksman A., Houck G. E. Jr., Brown W. T., Dobkin C. S. (1994) Mosaicism in fragile X affected males. *Am. J. Med. Genet.* 51: 509-512
- Nolin S. L., Lewis F. A. 3rd, Ye L. L., Houck G. E. Jr., Glicksman A. E., Limprasert P., Li S. Y., Zhong N., Ashley A. E., Feingold E., Sherman S. L., Brown W. T. (1996) Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 59: 1252-1261
- Nolin S. L., Brown W. T., Glicksman A., Houck Jr. G. E., Gargano A. D., Sullivan A., Biancalana V., Brondum-Nielsen K., Hjalgrim H., Holinski-Feder E., Kooy F., Longshore J., Macpherson J., Mandel J. L., Matthijs G., Rousseau F., Steinbach P., Vaisanen M. L., von Koskull H., Sherman S. L. (2003) Expansion of the Fragile X CGG Repeat in Females with Premutation or Intermediate Alleles. *Am. J. Hum. Genet.* 72: 454-464
- Novik K.L., Nimmrich I., Genc B., Maier S., Piepenbrock C., Olek A., Beck S. (2002) Epigenomics: genome-wide study of methylation phenomena. *Curr. Issues Mol. Biol.* 4: 111-128
- Oberlé I., Rousseau F., Heitz D., Kretz C., Devys D., Hanauer A., Boue J., Bertheas M. F., Mandel J.-L. (1991) Instability of a 550 base pair DNA segment and abnormal methylation in fragile X chromosome. *Science* 252: 1097-1102
- O'Donnell W. T., Warren S. T. (2002) A decade of molecular studies of fragile X syndrome. *Annu. Rev. Neurosci.* 25: 315-338
- Ohno S. (1972) in *Evolutio of Genetic Systems*, Brookhaven Symp. Biol. (ed. Smith H. H.) 366-370
- Okubo K., Hori N., Matoba R., Niiyama T., Fukushima A., Kojima Y., Matsubara K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 2: 173-179

- Opitz J. M., Holt M. C., Spano L. M. (1986) Bibliography of X-linked mental retardation and related subjects. III. *Am. J. Med. Genet.* 23: 69-99
- Orgel L. E. und Crick F. H. C. (1980) Selfish DNA: the ultimate parasite. *Nature* 284: 604-607
- Orr H. T., Chung M., Banfi S., Kwiatkowski Jr. T. J., Servadio A., Beaudet A. L., McCall A. E., Duvick L. A., Ranum L. P. W. und Zoghbi Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet.* 4: 221-225
- Parrish J. E., Oostra B. A., Verkerk A. J. M. H., Richards C. S., Reynolds J., Spikes A. S., Shaffer L. G. und Nelson D. L. (1994) Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. *Nature Genet.* 8: 229-235
- Paulson H. L., Fischbeck K. H. (1996) Trinucleotide repeats in neurogenetic disorders. *Annu. Rev. Neurosci.* 19: 79-107
- Pauws E., van Kampen A. H. C., von de Graaf S. A. R., de Vijlder J. J. M. und Ris-Stalpers C. (2001) Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.* 29: 1690-1694
- Pearson C. E., Ewel A., Acharya S., Fishel R. A. und Sinden R. R. (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* 6: 1117-1123
- Pelletier R., Krasilnikova M. M., Samadashwily G. M., Lahue R., Mirkin S. M. (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* 23: 1349-1357
- Pham C. T., MacIvor D. M., Hug B. A., Heusel J. W., Ley T. J. (1996) Long-range disruption of gene expression by a selectable marker cassette. *Proc. Natl. Acad. Sci. USA* 93: 13090-13095
- Pieretti M., Zhang F. P., Fu Y. H., Warren S. T., Oostra B. A., Caskey C. T. und Nelson D. L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66: 817-822
- Pesole G., Gissi C., Grillo G., Licciulli F. L., Liuni S., Saccone C. (2000) Analysis of oligonucleotide AUG start codon context in eukaryotic mRNAs. *Gene* 261: 85-91
- Pesole G., Liuni S., Grillo G., Licciulli F., Mignone F., Gissi C. und Saccone C. (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions in eukaryotic mRNAs. Update 2002. *Nucleic Acids Res.* 30: 335-340
- Potter H., Weir L., Leder P. (1984) Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci. USA* 81: 7161-7165
- Primerano B., Tassone F., Hagerman R. J., Hagerman P., Amaldi F., Bagni C. (2002) Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *RNA* 8: 1482-1488
- Rajewsky K., Gu H., Kuhn R., Betz U. A., Müller W., Roes J., Schwenk F. (1996) Conditional gene targeting. *J. Clin. Invest.* 98: 600-603

- Ramírez-Solis R., Rivera-Pérez J., Wallace J. D., Wims M., Zheng H. und Bradley A. (1992) Genomic DNA microextraction: a method to screen numerous samples. *Anal. Biochem.* 201: 331-335
- Ranum L. P., Day J. W. (2002) Dominantly inherited, non-coding microsatellite expansion disorders. *Curr. Opin. Genet. Dev.* 12: 266-271
- Rat Genome Sequencing Consortium <http://www.hgsc.bcm.tmc.edu/projects/rat/>
- Razin A. und Shemer R. (1995) DNA methylation in early development. *Hum. Mol. Genet.* 4: 1751-1755
- Remus R., Kämmer C., Heller H., Schmitz B., Schell G., Doerfler W. (1999) Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. *J. Virol.* 73: 1010-1022
- Remus R. (2000) Verteilung der fünften DNA-Base 5-Methyl-Desoxycytidin im Säugergenom: DNA-Methylierungsmuster endogener retroviraler Sequenzen des Syrischen Hamsters (*Mesocricetus auratus*) und in der Promotoren menschlicher Erythrozytenmembranskelettproteine (*Homo sapiens*). Dissertation. Köln, 2000
- Reyniers E., Vits L., de Boulle K., van Roy B., van Velzen D., de Graaff E., Verkerk A. J. M. H., Jorens H. Z. J., Darby J. K., Oostra B. und Willems P. J. (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nature Genet.* 4: 143-146
- Richard G.-F., Hennequin C., Thierry A., Dujon B. (1999) Trinucleotide repeats and other microsatellites in yeast. *Res. Microbiol.* 150: 589-602
- Richards R. I., Holman K., Friend K., Kremer E., Hillen D., Staples A., Brown W. T., Goonewardena P., Tarleton J., Schwartz C. und Sutherland G. R. (1992) Evidence of founder chromosomes in fragile X syndrome. *Nature Genet.* 1: 257-260
- Richards R. I. und Sutherland G.R. (1992) Dynamic mutations: a new class of mutations causing human disease. *Cell* 70: 709-712
- Richards R. I., Holman K., Yu S. und Sutherland G. R. (1993) Fragile X syndrome unstable element, p(CGG)_n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Hum. Mol. Genet.* 2: 1429-1435
- Riggs A. D., Jones P. A. (1983) 5-methylcytosine, gene regulation and cancer. *Adv. Cancer Res.* 40: 1-30
- Ritchie R. J., Knight S. J. L., Hirst M. C., Grewal P. K., Bobrow M., Cross G. S. und Davies K. E. (1994) The cloning of FRAXF: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. *Hum. Mol. Genet.* 3: 2115-2121
- Robertson E., Bradley A., Kuehn M. und Evans M. (1986) Germ-line transmission of genes introduced into culture pluripotential cells by retroviral vector. *Nature* 323: 445-448
- Rockman M. V., Wray G. A. (2002) Abundant raw material for *cis*-regulatory evolution in humans. *Mol. Biol. Evol.* 19:1991-2004

- Rolfmeier M. L., Dixon M. J., Pessoa-Brandão, Pelletier R., Miret J. J. und Lahue R. S. (2001) Cis-Elements governing trinucleotide repeat instability in *Saccharomyces cerevisiae*. *Genetics* 157: 1569-1579
- Rosser T. C., Johnson T. R., Warren S. T. (2002) A cerebellar FMR1 riboCGG binding protein. *Am. J. Hum. Genet.* 71 (Suppl.) abstract 1978
- Ruan H., Hill J. R., Fatemie-Nainie S., Morris D. R. (1994) Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Influence of the structure of the 5' transcript leader on regulation by the upstream open reading frame. *J. Biol. Chem.* 269: 17905-17910
- Rubinsztein D. C., Barton D. E., Davison B. C., Ferguson-Smith M. A. (1993) Analysis of the huntingtin gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. *Hum. Mol. Genet.* 2: 1713-1715
- Rubinsztein D. C., Leggo J., Coetzee G. A., Irvine R. A., Buckley M., Ferguson-Smith M. A. (1995a) Sequence variation and size ranges of CAG repeats in the Machado-Joseph disease, spinocerebellar ataxia type 1 and androgen receptor genes. *Hum. Mol. Genet.* 4: 1585-1590
- Rubinsztein D. C., Amos W., Leggo J., Goodburn S., Jain S., Li S.-H., Margolis R. L., Ross C. A. und Ferguson-Smith M. A. (1995b) Microsatellite evolution - evidence for directionality and variation in rate between species. *Nature Genet.* 10: 337-343
- Russel G. J., Walter P. M. B., Elton R. A., Subak-Sharpe J. H. (1976) Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J. Mol. Biol.* 108: 1-23
- Saiki R. K., Gefand D. H., Stoffel S., Scharf S. J., Higuchi R., Horn G. T., Mullis K. B., Erlich H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 266: 1208-1212
- Salat U., Bardoni B., Wöhrle D., Steinbach P. (2000) Increase of FMRP expression, raised levels of FMR1 mRNA, and clonal selection in proliferating cells with unmethylated fragile X repeat expansions: a clue to the sex bias in the transmission of full mutations? *J. Med. Genet.* 37: 842-850
- Samadashwily G. M., Raca G., Mirkin S. M. (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* 17: 298-304
- Sambrook J., und Russell D.W. (2001) *Molecular Cloning. A laboratory manual.* 3. edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Sanger F., Nicklen S. und Coulson A.T. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467
- Sato T., Oyake M., Nakamura K., Nakao K., Fukusima Y., Onodera O., Igarashi S., Takano H., Kikugawa K., Ishida Y., Shimohata T., Koide R., Ikeuchi T., Tanaka H., Futamura N., Matsumura R., Takayanagi T., Tanaka F., Sobue G., Komure O., Takahashi M., Sano A., Ichikawa Y., Goto J., Kanazawa I., et al. (1999) Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats comparable with those in DRPLA patients. *Hum. Mol. Genet.* 8: 99-106

- Sauer B., Henderson N. (1989) Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* 17: 147-161
- Sauer B. (1996) Multiplex Cre/lox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. *Nucleic Acids Res.* 24: 4608-4613
- Sauer B. (1998) Inducible gene targeting in mice using the Cre/lox system. *Methods* 14: 381-392
- Schlötterer C. und Tautz D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: 211-215
- Schmidt E. E., Taylor D. S., Prigge J. R., Barnett S., Capecchi M. R. (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. USA* 97: 13702-13707
- Schuler G. D. (1997) Pieces of the puzzle: expressed sequence tags and the catalogue of human genes. *J. Mol. Med.* 75: 694-698
- Schwartzberg P. L., Goff S. P., Robertson E. (1989) Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. *Science* 246: 799-803
- Schwenk F., Baron U., Rajewsky K. (1995) A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* 23: 5080-5081
- Sheets M. D., Ogg S. C. und Wickens M. P. (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. *Nucleic Acids Res.* 18: 5799-5805
- Sherman S. L., Morton P. A., Jacobs P. A. und Turner G. (1984) The marker (X) chromosome: a cytogenetic and genetic analysis. *Ann. Hum. Genet.* 48: 21-37
- Sherman S. L., Jacobs P. A., Morton P. A., Froster-Iskenius U., Howard-Peebles P., Nielsen K. B., Partington M. W., Sutherland G. R., Turner G. und Watson M. (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* 69: 289-299
- Sherman S. L. (2002) Epidemiology. In Hagerman R. J. und Hagerman P. J. (eds.), *Fragile X syndrome: Diagnosis, Treatment and Research*, Johns Hopkins University Press, Baltimore, pp. 136-168
- Shimizu M., Gellibolian R., Oostra B. A. und Wells R. D. (1996) Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J. Mol. Biol.* 258: 614-626
- Shiota K., Yanagimachi R. (2002) Epigenetics by DNA methylation for development of normal and cloned animals. *Differentiation* 69: 162-166
- Sinden R. R. (1994) *DNA Structure and Function*. Academic, San Diego, USA

- Sinden R. R., Potamen V. N., Oussatcheva E. A., Pearson C. E., Lyubchenko Y. L. und Shlyakhtenko L. S. (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci.* 27: 53-65
- Siomi H., Siomi M. C., Nussbaum R. L., Dreyfuss G. (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell.* 74: 291-298
- Smeets H. J. M., Smits A. P. T., Verheij C. E., Theelen J. P. G., Willemsen R., van de Burgt I., Hoogeveen A. T., Oosterwijk J. C. und Oostra B. A. (1995) Normal phenotype in two brothers with a full FMR1 mutation. *Hum. Mol. Genet.* 4: 2103-2108
- Smit A. F. A. und Green P. at: <http://repeatmasker.genome.washington.edu>
- Smith A. G., Nichols J., Robertson M. und Rathjen P. D. (1992) Differentiation inhibiting activity (DIA/LIF) and mouse development. *Dev. Biol.* 151: 339-351
- Smith G. P. (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harb. Symp. Quant. Biol.* 507-513
- Smith S. S. (1994) Biological implications of the mechanism of action of human DNA (cytosine-5) methyltransferase. *Prog. Nucleic Acid Res. Mol. Biol.* 49: 65-111
- Snow K., Doud L. K., Hagerman R., Pergolizzi R. G., Erster S. H., Thibodeau S. N. (1993) Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *Am. J. Hum. Genet.* 53: 1217-1228
- Southern E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517
- Spiro C., Pelletier R., Rolfsmeier M. L., Dixon M. J., Lahue R. S., Gupta G., Park M. S., Chen X., Mariappan S. V., McMurray C. T. (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* 4: 1079-1085
- Sternberg N. und Hamilton D. (1981) Bacteriophage P1 site specific recombination I. Recombination between loxP sites. *J. Mol. Biol.* 150: 467-486
- Stöger R., Kajimura T. M., Brown W. T. und Laird C. D. (1997) Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. *Hum. Mol. Genet.* 6: 1791-1801
- Strausberg R.L., Feingold E. A., Klausner R. D., Collins F. S. (1999) The mammalian gene collection. *Science* 286: 455-457
- Streisinger G., Okada Y., Emrich J., Newton J., Tsugita A., Terzaghi E., und Inouye M. (1966) Frameshift mutations and the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* 77-84
- Sturzeneker R., Bevilacqua R. A. U., Haddad L. A., Simpson A. J. G. und Pena S. D. J. (2000) Microsatellite instability in tumors as a model to study the process of microsatellite mutations. *Hum. Mol. Genet.* 9: 347-352

- Subak-Sharpe H., Burk R. R., Crawford L. V., Morrison J. M., Hay J., Keir H. M. (1966) An approach to evolutionary relationships of mammalian DNA viruses through analysis of the pattern of nearest neighbour base sequences. *Cold Spring Harbor Symp. Quant. Biol.* 31: 737-748
- Sung Y. J., Conti J., Currie J. R., Brown W. T., Denman R. B. (2000) RNAs that interact with the fragile X syndrome RNA binding protein FMRP. *Biochem. Biophys. Res. Commun.* 275: 973-980
- Sutcliffe J. S., Nelson D. L., Zhang F., Pieretti M., Caskey C. T., Saxe D. und Warren S. T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* 1: 397-400
- Sullivan A. K., Crawford D. C., Scott E. H., Leslie M. L. und Sherman S. L. (2002) Paternally transmitted FMR1 alleles are less stable than maternally transmitted alleles in the common and intermediate size range. *Am. J. Hum. Genet.* 70: 1532-1544
- Sutherland G. R. und Richards R. I. (1995) The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.* 5: 323-327
- Sutherland G. R. und Richards R. I. (1999) Fragile sites- cytogenetic similarity with molecular diversity *Am. J. Hum. Genet.* 64: 354-359
- Sutherland G. R. und Baker E. (2000) The clinical significance of fragile sites on human chromosomes. *Clin. Genet.* 58: 157-161
- Suzuki Y., Ishihara D., Sasaki M., Nakagawa H., Hata H., Tsunoda T., Watanabe M., Komatsu T., Ota T., Isogai T., Suyama A. und Sugano S. (2000) Statistical analysis of the 5'-untranslated region of human mRNA using „oligo-capped“ cDNA libraries. *Genomics* 64: 286-297
- Suzuki Y., Tsunoda T.; Jun S., Taira H., Mizushima-Sugano J., Hata J., Ota T., Isogai T., Tanaka T., Nakamura Y., Suyama A., Sakaki Y., Morishita S., Okubo K. und Sugano S. (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* 11: 677-684
- Taillandier E., Taboury J. A., Adams S., Liquier J. (1984) Left-handed helical structure of poly[d(A-C)].poly[d(G-T)] studied by infrared spectroscopy. *Biochemistry* 23: 5703-5706
- Tapscott S. J., Klesert T. R., Widrow R. J., Stöger R. und Laird C. D. (1998) Fragile X syndrome and myotonic dystrophy: parallels and paradoxes. *Curr. Opin. Genet. Dev.* 8: 245-253
- Tassone F., Hagerman R. J., Ikle D., Dyer P. N., Lampe M., Willemsen R., Oostra B. A. und Taylor A. K. (1999) FMRP expression as a potential prognostic indicator in fragile X syndrome. *Am. J. Med. Genet.* 84: 250-261
- Tassone F., Hagerman R. J., Loesch D. Z., Lachiewicz A., Taylor A. K., Hagerman P. J. (2000a) Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am. J. Med. Genet.* 94: 232-236
- Tassone F., Hagerman R. J., Chamberlain W. D., Hagerman P. J. (2000b) Transcription of the FMR1 gene in individuals with fragile X syndrome. *Am. J. Med. Genet.* 97: 195-203

- Tassone F., Hagerman R. J., Taylor A. K., Gane L. W., Godfrey T. E. und Hagerman P. J. (2000c) Elevated levels of FMR1 mRNA in carrier males: A new mechanism of involvement in the fragile X syndrome. *Am. J. Hum. Genet.* 66: 6-15
- Tassone F., Hagerman R. J., Taylor A. K., Hagerman P. J. (2001) A majority of fragile X males with methylated, full mutation alleles have significant levels of FMR1 messenger RNA. *J. Med. Genet.* 38: 453-456
- Tassone F., Sahota P., Pleasant N., Laird C., Hagerman P. J. (2002) Methylation analysis of the promoter region in fragile X males with hypermethylated, full mutation alleles. *Am. J. Hum. Genet.* 71 (Suppl.) abstract 930
- Tautz D., Trick M. und Dover G. A. (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656
- Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17: 6463-6471
- Tautz D., Schlötterer C. (1994) Simple sequences. *Curr. Opin. Genet. Dev.* 4: 832-837
- The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science* 282: 2012–2018
- The Dutch-Belgian Fragile-X-Consortium (1994) FMR1 knockout mice: a model to study fragile X mental retardation. *Cell* 78: 23-33
- The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72: 971-983
- Thibodeau S. N., Bren G., Schaid D. (1993) Microsatellite instability in cancer of the proximal colon. *Science* 260: 816-819
- Thomas C. A. (1971) The genetic organization of chromosomes. *Annu. Rev. Genet.* 5: 237-256
- Thomas K. R., Capecchi M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51: 503-512
- Thyagarajan B., Guimaraes M. J., Groth A. C., Calos M. P. (2000) Mammalian genomes contain active recombinase recognition sites. *Gene* 244: 47-54
- Thomas K. R., Deng C., Capecchi M. R. (1992) High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. *Mol. Cell. Biol.* 12: 2919-2923
- Timchenko L. T. (1999) Myotonic dystrophy: the role of RNA CUG triplet repeats. *Am. J. Hum. Genet.* 64: 360-364
- Torres R. M., Kühn R. (1998) Laboratory protocols for conditional gene targeting. Oxford University Press, Oxford
- Tóth F., Gáspári Z., Jurka J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981

- Turner G., Till R., Daniel A. (1978) Marker chromosomes, mental retardation and macroorchidism. *N. Engl. J. Med.* 299: 1472
- Usdin K. und Woodford K. J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.* 23: 4202-4209
- Utsch B., Becker K., Brock D., Lentze M. J., Bidlingmaier F., Ludwig M. (2002) A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length? *Hum. Genet.* 110: 488-494
- van den Broek W. J., Nelen M. R., Wansink D. G., Coerwinkel M. M., te Riele H., Groenen P. J., Wieringa B. (2002) Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.* 11: 191-198
- Vardimon L., Kressmann A., Cedar H., Maechler M. und Doerfler W. (1982) Expression of a cloned adenovirus gene is inhibited by *in vitro* methylation. *Proc. Natl. Acad. Sci. USA* 79: 1073-1077
- Verheij C., Bakker C. E., de Graaff E., Keulemans J., Willemsen R., Verkerk A. J., Galjaard H., Reuser A. J., Hoogeveen A. T., Oostra B. A. (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. *Nature.* 363: 722-724
- Verkerk A. J. M. H., Pieretti M., Sutcliffe J. S., Fu Y.-H., Kuhl D. P. A., Pizzuti A., Reiner O., Richards S., Victoria M. F., Zhang F., Eussen B. E., van Ommen G.-J. B., Blonden L. A. J., Riggins G. J., Chastain J. L., Kunst C. B., Galjaard H., Caskey C. T., Nelson D. L., Oostra B. A., und Warren S. T. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X Syndrome. *Cell* 65: 905-914
- Wang Y.-H. und Griffith J. (1995) Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics* 25: 570-573
- Wang Z., Fang P., Sachs M. S. (1998) The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol. Cell. Biol.* 18: 7528-7536
- Watson J. D., Cook-Deegan R. M. (1991) Origins of the Human Genome Project. *FASEB J.* 5: 8-11
- Weber J. L., Wong C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2: 1123-1128
- Weissenbach J. (1993) Microsatellite polymorphism and the genetic linkage map of the human genome. *Curr. Opin. Genet. Dev.* 3: 414-417
- Wellauer P. K. und Dawid I. B. (1979) Isolation and sequence organization of human ribosomal DNA. *J. Mol. Biol.* 128: 289-303

- Wells R. D. (1996) Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* 271: 2875-2878
- White P. J., Borts R. H. und Hirst M. C. (1999) Stability of human fragile X (CGG)_n triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. *Mol. Cell. Biol.* 19: 5675-5684
- Wierdl M., Dominska M., Petes T. D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769-779
- Wöhrle D., Hennig I., Vogel W. und Steinbach P. (1993) Mitotic stability of fragile X mutations in differentiated cells indicates early post-conceptual trinucleotide repeat expansion. *Nature Genet.* 4: 140-142
- Wolfe K. H., Sharp P. M. (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.* 37: 441-456
- Xu X., Peng M., Fang Z. und Xu X. (2000) The direction of microsatellite mutations is dependent upon allele length. *Nature Genet.* 24: 396-399
- Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13, 555-556
- Yang Z. und Nielsen R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17: 32-43
- Yano-Yanagisawa H., Li Y., Wang H. und Kohwi Y. (1995) Single stranded DNA-binding proteins isolated from mouse brain recognize specific trinucleotide repeat sequences in vitro. *Nucleic Acids Res.* 23: 2654-2660
- Zhuchenko O., Bailey J., Bonnen P., Ashizawa T., Stockton D. C. A., Dobyns W. B., Subramony S. H., Zoghbi H. Y. und Lee C. C. (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the $\alpha 1A$ -voltage-dependent calcium channel. *Nature Genet.* 15: 62-69
- Zeschnigk M., Schmitz B., Dittrich B., Buiting K., Horsthemke B., Doerfler W. (1997) Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Hum. Mol. Genet.* 6: 387-395
- Zoghbi H. Y., Orr H. T. (1999) Polyglutamine diseases: protein cleavage and aggregation. *Curr. Opin. Neurobiol.* 9: 566-570

I Anhang

I.1 Datenbankeintrag von CGGBP1 - Zugriffsnummer AJ000258

LOCUS HSCGGBP 779 bp RNA PRI 22-JUL-1997
 DEFINITION Homo sapiens trinucleotide repeat 5-d(CGG)n-3ds binding protein p20-CGGBP.
 ACCESSION AJ000258
 NID g2274963
 KEYWORDS CGGBP gene; DNA binding protein; trinucleotide repeat.
 SOURCE human.
 ORGANISM [Homo sapiens](#)
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 779)
 AUTHORS Deissler,H.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-1997) Deissler H., Department for Medical Genetics and Virology, Institute of Genetics, University of Cologne, Weyertal 121, NRW, D-50931 Cologne, GERMANY
 REFERENCE 2 (bases 1 to 779)
 AUTHORS Deissler,H., Wilm,M., Genc,B., Schmitz,B., Ternes,T., Naumann,F., Mann,M. and Doerfler,W.
 TITLE Rapid protein sequencing by tandem mass spectrometry and cDNA cloning of p20-CGGBP. A novel protein that binds to the unstable triplet repeat 5'-d(cgg)n-3' in the human fmrl gene
 JOURNAL J. Biol. Chem. 272 (27), 16761-16768 (1997)
 MEDLINE [97347474](#)
 COMMENT EST clone **ID269133** genbank accession no. N36676 and N24697.
 FEATURES Location/Qualifiers
 source 1..779
 /organism="Homo sapiens"
 /plasmid="pT7T3D"
 /db_xref="taxon:9606"
 /cell_type="melanocytes"
 /chromosome="3"
 /clone_lib="soares melanocytes 2NbHM"
 /lab_host="DH10B"
 /sex="male"
 gene 197..759
 /gene="CGGBP"
 CDS 197..700
 /gene="CGGBP"
 /function="DNA binding protein"
 /note="p20-CGGBP binds highly sequence-specific to the double-stranded triplet repeat 5-d(CGG)n-3. Binding is severely inhibited by cytosine-specific DNA-methylation. p20-CGGBP was isolated from human HeLa cells, three internal peptides were sequenced that were completely contained in the corrected sequence of EST cloneID269133 (genbank acc.-No: N36676, N24697). The derived aa sequence lacks any overall homology to any known protein. p20-CGGBP RNA is expressed in a variety of human tissues. The cDNA-sequence of p20-CGGBP is highly conserved among mammals. Sequenced Peptides: peptide1: aa 4 to 12 FVVTAPPAR peptide2: aa 19 to 21 LYV peptide3: aa 101 to 104 VSVIQ putative nuclear localization signal: aa 69 to 84"

```

/codon_start=1
/evidence=experimental
/product="p20-CGGBP"
/db_xref="PID:e329707"
/db_xref="PID:g2274964"
/translation="MERFVVVTAPPARNRSKTALYVTPLDRVTEFGGELHEDGGKLFCT
SCNVVLNHVRKSAISDHLKSKTHTKRKAEFEEQNVRKKQRPLTASLQCNSTAQTEKVS
VIQDFVKMCLEANI PLEKADHPAVRAFLSRHVKNNGGSI PKSDQLRRAYLPDGYENENQ
LLNSQDC"
polyA_signal 729..732
               /gene="CGGBP"
polyA_site   759
               /gene="CGGBP"
BASE COUNT   242 a    161 c    175 g    201 t
ORIGIN
   1 ggcacgaggg tttcgctctg gagaccattc cctgctaagt atcaagacga aaaaaactgg
  61 aaactaatcc gaatttctgt ggaatgttta atcttctgga tccatgactg tctgatacgt
 121 tggcaattta aagtcctttt gaaagagagt tcatgttacc cagctattct ctaaaccata
 181 tttattttaga gtcagaatgg agcgatttgt agtaacagca ccacctgctc gaaaccgttc
 241 taagactgct ttgtatgtga ctcccctgga tcgagtcact gagtttggag gtgagctgca
 301 tgaagatgga ggaaaactct tctgcacttc ttgcaatgtg gttctgaatc atgttcgcaa
 361 gtctgccatt agtgaccacc tcaagtcaaa gactcatacc aagaggaagg cagaatttga
 421 agagcagaat gtgagaaaga agcagaggcc cctaactgca tctcttcagt gcaacagtac
 481 tgcgcaaaca gagaaagtca gtgttatcca ggactttgtg aaaatgtgcc tggagccaa
 541 catcccactt gagaaggctg atcaccagc agtccgtgct ttcctatctc gccatgtgaa
 601 gaatggaggc tccataccta agtcagacca gctacggagg gcatactctc ctgatggata
 661 tgagaatgag aatcaactcc tcaactcaca agattgttga ctaggaggtt accaccattg
 721 tgatcaagat aaatgtggag tattaaagtt atgtgttgaa aaaaaaaaaa aaaaaaact

```

//

I.2 Verwendete Oligodesoxyribonukleotide

I.2.1 Sequenzierung von m20-9

Name	Position in AY189895 (nt)	Sequenz 5'-3'
M5prime 1	1-21	TTC CCA GAC TCC CTC TCC GTC
mI-t5'	788-808	CCT TCA TCC CGT TGC TCT GCC
mI-t	884-903	AAG AGC TGG GAA GGC GGA AG
mI-b	902-882	TTC CGC CTT CCC AGC TCT TCG
m1t-2	1002-1021	AGC CTG ACA GCG ACG GCC TC
mI-t3	1314-1333	CTG GCT GGA GCT TCG GGA TC
m2-b3-2	1325-1308	AGC TCC AGC CAG CCC CAC
m2b2	1333-1314	GAT CCC GAA GCT CCA GCC AG
mII-t5'	1468-1487	AAG ACA TGA CTT CCC GGC GC
mII-b5'	1489-1472	TGG CGC CGG GAA GTC ATG TC
mII-b	1611-1592	CCA CAG ACG CAT CCA ATC CG
mII-t	1673-1692	CCC GAC TGG AAA TTA AGC TG
mII-b3'	1692-1673	CAG CTT AAT TTC CAG TCG GG
m2t5-2	2062-2081	TGC AGA CAT TTG CAG TGA GG
m2t5-r1	2119-2097	CAG ACA TCA TGA TGA ATC CCA CC
m2ts-3	2551-1571	ATC TTG GGT TTC CAG TGC TTG
m2-t5-r2	2573-2553	TAC AAG CAC TGG AAA CCC AAG
m2t5-4	2770-2792	CTG TGA TGT GAT CCC TGC CAT TC
m4-bS-4	2799-2779	TTG TCC TGA ATG GCA GGG ATC
m2t5-5	3067-3087	ACA TCA AGT TGT GCC CAG CAC
m485-3	3099-3077	AGT GCC CAA TAT GTG CTG GGC AC
m485-r2	3277-3297	TCT ACA TAG TGA GAC TCT GTC
m4b5-2	3660-3640	CTA ACC TTG CCC AAC TTA GAC
m4b5-r1	3789-3808	TAG TAG GGC CGT GTG CTT CC
mIV-t	4229-4247	TCA GAG ACA AAA TGG AAC G
mIV-b5'	4248-4230	TCG TTC CAT TTT GTC TCT G
m4-t2	4732-4751	CAA GAT TGT TGA CGA GGC AG
mIV-b3'	4764-4744	CAA TGG TCA TTA ACT GCC TCG
m4-t3	5155-5177	CTC TCA CAG GAC CTT AAA GAA GC
m4-t2-r1	5177-5155	GCT TCT TTA AGG TCC TGT GAG AG
m4-t4	5547-5567	ACA CTG TGG TTC TGT TTT CTC
m4t2-r2	5603-5583	GTT GTC AGC CAA AAA GTA ACC
m4-t5	5852-5872	ACA CCT CAC TCT TCT GTA AGC
m4t2-r3	6045-6025	AAT CAG GGC ATC TTG CCA CTG
m4-t6	6342-6362	GGT TGT CAA AGT GAC CAC TTG
m4t2-r4	6595-6577	GAT ATC CTA CCA TCT TAG G
m4-t7	6718-6738	TCA ACA TGG CAG AAC AGC TTC
m4t2-r5	7132-7112	AGT TCA CTT TGC CTC TCT CAC
m4-t8	7189-7206	GAC CAA TAG CAG CTC CTG
m4-t9	7653-7673	CCA GAT TCC CAA GCT AAA ACC
G0-SN-bw	7666-7646	GCT TGG GAA TCT GGC CAA TTG
m4-t8-r	7774-7751	ATC CAA TGA AAA CTT GTT TCT AGC

G1-SN-fw	8185-8205	TTT TGT GAG CCA GTG ACT GAC
m4t9-r	8212-8189	ATA AAA TGT CAG TCA CTG GCT CAC
G3-SN-bw	8713-8692	TTG CAC AAA CGT GAG TGG AGG C
G2-SN-fw	8765-8788	GTT CTT AGG AAT CAA CCC ATA CTG
G6-SN-bw	8999-8980	CTT GGC TGG TAG CAA GTC TC
G4-SN-fw	9260-9281	TAT CAT AAC CCA AGC CTC TGG G
G5-SN-bw	9454-9435	CTG ACC ATC TCT CAG CCT CC
G9-SN-bw	9456-9435	ACT GAG CCA TCT CTC AGC CTC C
G8-SN-bw	9770-9749	AAG TTG ACA CAA GAC TGT CTC C
G7-SN-fw	9850-9872	CTG ACT GAT GGA GGT CCC AGT CC
G11-SN-bw	10230-10210	CCA GGG ATG TCC AGT GCT TGC
G10-SN-fw	10245-10265	CTG AAG CAC AAG AAG GTG ACC
G15-SN-fw	10578-10604	CCT GAC ACA AAA CAT CCA AGA AAT CGC
G13-SN-bw	10660-10639	CCA CCA TGC ATG GCT CAA TTC C
G19-bw	10959-10934	CAT GAA AGC TTC TTG TGT GTT TAC AG
G16-fw	11125-11146	CAG ATA GGC CTA GGC AGA TGT C
G17-bw	11146-11125	GAC ATC TGC CTA GGC CTA TCT G
G18-fw	11265-11288	TCT ACA AAT CTA GCC CTA CAG AAG
G14-base-bw	11432-11409	CAA GAC AGG GTT TCT CTG TAT AGC
G14-base-fw	11459-11481	GGA GAA GAT ACT AAA AGG AGA AT
m18-t7-2	12028-12048	CCA CTG TGC CAT CTC TCC AGC
K8-fw	12065-12045	GAA TGG CTA CTC CAG AGG CTG
m18-t7-3	12482-12503	GAG ACA GGG TTT CTC TGT ATA G
K6-SN-fw	12527-12505	AAG TGA GTT TCA GGA CAG CCT GG
K8-SN-bw	12561-12584	TAC CTG CCT TTG CCT CCT GAG TGC
K7-SN-fw	12624-12603	TAA TAC TAG GTG GGC ATG GTG G
K9-SN-fw	12704-12678	AGA GAA ACT CTG TCA TGA ACA CCC TGG
K1-SN-fw	12747-12724	AAC CTG GGC TAC AGG GTG AGT TCC
K2-SN-fw	12786-12766	TTG GGA GGC AGA GGC AGG TGG
K3-SN-fw	12803-12783	ATC TTT AAT CCC AGC ACT TGG
K5-SN-fw	12953-12931	CAA GAT CTG ACG CCA TCT TCT GG
m18-t7-4	12960-12980	ATG GTT GTG AGC CAC CAT GTG
K4-SN-fw	13007-12984	AAG GTC CAG AGT TCA AAT CCC AGC
K0-SN-bw	13176-13197	CTG TTC ATA ACC CAG AAC TTC C

Primer mit der Bezeichnung "SN" wurden von Sebastian Neubert im Rahmen eines Praktikums ausgesucht und für die Sequenzierung von m20-9 verwendet.

I.2.2 Oligodesoxyribonukleotide zur Sequenzierung/Klonierung des Geninaktivierungs-Vektors

Name	Position in AY189895	Sequenz 5'-3'
NEO 3'	easyflox	CCT CTG TTC CAC ATA CAC TTC
TK5'	easyflox	AAA ACC ACA CTG CTC GAG GTG
flox-orf-top	3834-3853	GTG TGT ATC TAT AAT CCC AG
flox-orf-bottom	5099-5080	AGC TCG TAA CCT CAC AAG AAA AAA

I.2.3 Oligodesoxyribonukleotide für RT-PCR und Klonierung/Sequenzierung der Promotorkonstrukte von g20

Name	Position in AF094481	Sequenz 5'-3'
p20-3utr	10904-10885	ATG GTG GTA ACC TCC TAG TC
p20-I	7456-7475	CCC TTA TTC CCT TGC TCC TC
p20-II	8285-8304	CTC TGG AGA CCA TTC CCT GC
p20-ORF	10834-10814	AGA TAT GCC CTC CGT AGC TGG
UTR-T	7400-7422	TTT CCC GCG GGG CTC GAT CGG GCA ACG
UTR-B	10383-10362	CCC AAG CTT TCT GAC TCT AAA TAA ATA TGG
GL1	pGL2bas	TGT ATC TTA TGG TAC TGT AAC TG

I.2.4 Oligodesoxyribonukleotide für das genomische Sequenzieren nach der Bisulfit-Methode

Name	Position in AF094481	Sequenz 5'-3'
p20-C5-L2	7261-7288	TTA TTA TTG GTG AAA TCT GAT AGT TTG TTT G
P20-A5-K	6118-6140	ATA TTG AAG GAT GAG GGG GGT TGG TG
P20-A3-K	6845-6819	TCC RTT AAC CRC CCC RAC TTA TCA CC
p20-A5-L	5932-5961	TTA TAA ATG TAG GTA GAT GAG AAT GTA GAG
P20-A3-L	6897-6868	ACR TCR AAA TCA ACA ATC RAA ATT ACT ACC
p20 B5-L	6696-6720	TTA AGT YGG ATA AGT GGA GGA GGG AG
p20-B5-K	6814-6789	GAT TAG GAG AYG TAG AGA GGT TTA GG
p20-B3-K	7345-7370	CCR CCT ACA CCR CCT ATA ACA AAA CC
P20-B3-L	7569-7544	CTC RAC CTC CCT AAC TCC TTA ATA CC
p20-C5-K	7307-7332	GGT TAA TGT TGT TYG GYG GTT GTT GG
P20-C3-K	7939-7917	CAC CTC CTC CAA CTC CCC AAA CC
P20-C3-L	7984-7960	TAA CAC TTC CCC TCC TAA ATC CTC C
p20-C3-K2	7553-7578	TCR CCR CCR TTA CCC RAT CRA ACC
p20-C3-L2	7579-7606	CCR AAC CCA AAA CCC RCA CCA ACR ATC

I.3 Sequenz – Zugriffsnummer AY189895

Sequenz von m20-9 und des 5'-flankierenden PCR-Produktes

```

!!NA_SEQUENCE 1.0
m20final.seq Length: 14530 July 18, 2002 10:35 Type: N Check: 9207 ..
  1  TTCCAGACT CCCTCTCCGT CCCGCCGAG GAcCGGTCCC GGACACGACG
 51  CAgAGACCTG GGCCAGCAGG CCAGGAGCCA CGCAGAGGCC CGGGAGCCCA
101  GTGACAAGTC GGAGCGGCCG ACGGGCGCCT CACCTGACAG AAGCGTCGGC
151  AGTAATCCCG GCTGCTGATC CCGACGCCGC CGcCACAGCC GCCGCCCGCG
201  GCCCCCCTC TGCCGTGcC CGCAGCTAGG AGCCCCACTG GGCCCGGGG
251  GGACTcCACA ATGGCCACGA GCTCTTCTCC gAGGCGCTCG GCCTCTTGGT
301  CGCTCTcTTC CTCCgCCAtG AGGcTcTCGc CGcCCAGCGC GTACCCAGCC
351  CcTGCGCCGC CGCCGcCGCC GCCGcTGcCG cCGcCGCTGc CTCAGCTAAC
401  CTCCTCCTCC TcCTCTAGTT CCCTCCCCCT TTcCCTCCGC CCCAGCCGGA
451  CAGGGCTACT CGCCTGGGCT CACTTcCGG CTcCGCCAC CCACGCCCCC
501  AGTTTTcAcT GGAGGAAGGA CACCGcCAAT cATAGGCATG CCCGCCCTAC
551  ctTccTATTG GTTACAcTAG CcGTCTGTTT GCCAggCCTG cCATcTGTTG
601  GTcCACACTA CtCGGCGGTT GcTgAGGaCC GCCTACGAGC CCTGCCATAG
651  GGCGCGCCGG AGgACGGGgC GCGGCAGGGG ACACGGCGgC CGCGGCGGAG
701  CTCGCTCGGG CGGCGGCGAC GCGGtGGCA GCGGCAGGAG CAGCGGCTGA
751  AGCAGCGGCG GAAGCAGCGG CGGCCGCAAC GGGATCTCCT TCATCCCGTT
801  GCTCTGCCCT TTCTTCTCTC TCTTTCTCTT CCTGGCGAGG GCTCGTGCAC
851  TTCCCTAAC GGCGGCCCGC TCCTACGGTC CCGAAGAGCT GGGAAAGCGG
901  AAGGTAGGGT tGGCCGtTG TGCGGgCGGc GAGCCAGGC CTTGCGGCCG
951  GGCGGGGAA CCGGGGGCGG GCGCGAgGT GGAGGAGAA GGAGCCGGgC
1001 CAGCTGACA GCgACGGCCT CGGCCGCGC CcgCCgC CCC CTCcCGGGG
1051 ACAGCCACGG AGTCGGCGGc GGTGCGGCGG ACTTCTCCGG GGCCCTCCGA
1101 GGGCGGCTGG GACGCGCCT CGGCCCGGC CCCGGCCCGC AGGCGGGGTC
1151 GCGGCGCCgG GGCGCCACCT CGGCCGCGC GGCGGGAGGA GGCTGTcTAG
1201 ACGTCCGGCG GcCCGCGCCC GCCTTCCCCG CCCCACCGG AGCCTCGGGC
1251 CTGCGGAGCG CGCGGGCCGC CGGGCCACCC GCcCCGgGGC GTGCGAgCCG
1301 AGCTGGcGTG gGGCTGGcTG GAGCTTCGGG ATCGCGCACG gAAGgCGGAg
1351 AGAGACAGAC ACCTGTGTGC AGAGCGtGGC CCGAGCCGCT CTGCCGCGGG
1401 CCTGCGCCCT GCCTTTCCCG CCCCTCCACC ATCTTCCCTT TCTGTTTGT
1451 TTTTGTGTTT TACAGTAAAG ACATGACTTC CCGGCGCCaC CGGAAAAATA
1501 GGGtGCAAGC CCAGGAACTG TTTCCCCACC ACCaCCTGTT GAAAAAcTGA
1551 GTCGAAGGCG TCTCCGGGGT CTGAACGAAC GcGAAGTCGC TCGGATtGGA
1601 TGCGTCTGtG GTTTTGCATC CAGAAAACTG TCaCTACTTA ATATCAAGGA
1651 AAaaCAAACa AACAAACAAA AACCCGACTG GAAATTAAGC TGGtGAGTAT
1701 TTCGACTTGA AAActCAGTT TACTTTTGCT GTAGTAAATG GCTGTtTCGC
1751 TTTTGTtCCC TGGGTGCTCT ATCTTAACTT CCTCATTTGT TTAAAAATAT
1801 TCTGACTTCT GGAAACACTT TTACACGTAA CCAACCTCCT CGCCTCTATG
1851 AGGtGGtTAT TTTcATTGGC AGTATCAAGA ATAGaCCTTT CAAGAGAAAA
1901 GATCAGGAAA GTTCAGGCTT CACGTGGCAC AAGAATTGCA GATCCcATT
1951 CTGTAAGTTT GCAGTCAAAG AAAAAATTCTA ACTGCACCCC TAGTTTTCa
2001 TAACAGCTTT AATtGTTACT AGAATGGTTT TATTTcATAG GTGTAGAGTA
2051 CTTGTGGAAG TTGCAGACAT TTGCAGTGAG gATTTGcTGA GTATTAGGTG
2101 GGATTCATCA TGATGTCTGT CTTTTCTGCC ATGAAAAAcTATGCCCTAA
2151 AACAGTATCT CACACGTAAC CCTTGAGtGT CAAAGATGtG GtGGGGTGA
2201 AATCCaCAGA AATTAAGCAT GTGAAATTCT AGACATTTTG tGCCTTAGAG
2251 AATTTTCATG AATTGCTAAG TCCTACTTTG AGTTCATAGA TTTTGCTCAA
2301 AATACAGaCC AAAGTTGAAA TAATCTTGGt GTTGCCcAGC AAGTAGaCTA
2351 ATGCAGTGGG ATACTCCATG CTTTGTAAATG ttACTTATAA AAaCGAATTC
2401 TTCAGaACAT ACTTGAAGTA TTTCTTCAAC ATGTTTTTAA CAAAAATAT
2451 CAGTATAGTA GTATcTgATG TCTTGATAT TATAAAaAGT TAAATCaCAG
2501 GCATAGAAgA GTAAGTCTTG CTAATCtTTt GTTATCAAAG TATATTTAT
2551 ATCTTGGGTT TCCAGtGCTt GTAATTTtAT ATAGCTGATA GTCCGTGAT
2601 AATTTTCAGA AGTTTGGGGG CCTAGTGAAG TGACTTTGAG AGTCACTAaC
2651 AACaCTGAGT TAGTCCaTGa AACCaCTGTC CTCAGCTCCA AACATGTGCC
2701 AgaACACAGG TGCTCACACA TTTAGaCTTC TGTAAGAAAGT AAGACACGGC
2751 CAGGTATGCT GGCACATGTC TGTgATGTGA TCCCTGCCAT TCAGGACAAC
2801 ATTGAAagCC CGCTGGACAA CAGGGTTGTT CTTGGTGTtT TGAGTTTGT

```

2851 TGTTAGGCTA AATTTAGTGT TAAACTGCTT CTAAATTATA TGGGTTTTGA
 2901 TAAATGGTTA TGGGGGGGTC GTTATTTGAA TTTGGCTTTT TTTTTTTCCT
 2951 GTTTGTTCAT TTTtTGTTTT TGTTTTTGA tTGCTGAGGTC AAACCTAGAA
 3001 TTTTAGCTCAT GcTAAGCAGT tGTTCTGCCT cTAGTGGACA CCCAGCTGTT
 3051 CATTACTTTT TTATAAACAT CAAGTTGTGC CCAGCACATA TTGGGCACTA
 3101 TGTtAAgGGA AATAAAAAATT AGTCTTCAAT GGAAATTGGA AACCaAATGT
 3151 TAAAGAACAT ATTGCCCTTCT CTTGATTATC AACTTTTTTAG AGAAAATACAT
 3201 ATGTTAATAT TTTTCCAGAC ATGGTGTcAT ACAGTTTCAG CCTCAGAATT
 3251 CAGGAGGTGa CAAAGATAGG TCTCTGTCTA CATAGTGAGA CTCTGTCTTT
 3301 GTTTTTGGTT AGCTGGTTTT CTAGACTGGA TTTCTCTGtG TAACCCCTGGC
 3351 TGTcCTGGAA CTCaCTTCAT AGAACAGGCC TGCCCTCTGCC TTCCCAAGTG
 3401 CTGGGATTAA AGGCATGTAC CaCTAGGCAT TGAGGTcCTG TCTTAAATAA
 3451 GTAGTACTTA TTGtCAACTA TATGTAGGCC TCAGGTtGtA ATGTTATtGt
 3501 AAATAGTTAG GCATGCTTTA TGgTtTTTT AGTtTTACT TGTTTTTCTT
 3551 TTTGGAAAGT AAGTCTTACT ACCAGATTTT TGCACAgTTG AATTCTCAaA
 3601 ATCGCTTTCC CATCAAGAAG CTCTTTAACT TGAACACATG TCTAAGTTGG
 3651 GCAAGGtTAG TATTTTTAGT ACTATGAATG CTGCTGTGAT TgTgATTTTC
 3701 TAAACGATAG AGACATTTAG AATACTTTAA TATTTCTTCA GGGGTCTGAT
 3751 TTATTTGTCA TAAAAACAA GGTATTGGGA AGTTGGATTA GTAGGGCCGt
 3801 GTGCTTCCTG TGCAAGTGTG TGGACTCTCT GCAGTGTGTA TCTATAATCC
 3851 CAGTTCTGGG GTCAGTTAGA GAAGCTGTTC AAAAAACA CA AGGTGTGCAC
 3901 ATAGGCAGTA CCACATACAC ACACACGTCA AGGAATGGAC CTAGCTGGAT
 3951 GCTGTCTGATT TGCTTTAGTT TTGTTTTTCTT TGAATATGTC AGATGTTTAG
 4001 AAgACATGTG ACCTATAAAT TGAACATAAAA GTGTTTGTGG GATACGGAGG
 4051 ATCACTGAAT TAAGTTTTGTT TTTCTGTGTTG TTGTTTTTGGG GGGGGTGTAT
 4101 TTATTGTTTT TTTTTTAAAT CTACCCTTAC AAAAGCACTT ATAACTACTT
 4151 TTTATTAATG TTAATCACTT CTAGAATGAA CTATTATCAT AATACTAACA
 4201 TCTTTCTCTT GTCTTAgAAG AACCTTATTC AGAGACAAAA TGGAACGATT
 4251 TGTTGTAACA GCACCACCTG CTCGAAATCG TTCTAAGACT GCTTTGTACG
 4301 TAACCCCTCT GGATCGAGTC ACTGAATTTG GAGGTGAGCT TCATGAGGAT
 4351 GGAGGCAAAC TCTTCTGCAC CTCTTGCAAT GTGGTGTCTGA ATCATGTTcG
 4401 CAAGTCTGACC ATTAGTGACC ACCTCAAGTC AAAGACTCAT ACAAAGAGGA
 4451 AGGCAGAATT TGAAGAGCAA AATGTGAGAA AGAAGCAGAG GCCCTGACT
 4501 GCATCGCTTC AGTGCAACAG TCCTGCGCAA ACAGAGAAAG CCAGTGTtAT
 4551 CCAGGACTTT GTGAAAATGT GCCTGgAAGC CAACATCCCC CTGGAGAAGG
 4601 CCGATCATCC CGCAGTGCgA GCGTTCCTGT CCCGGCATGT GAAGAACGGA
 4651 GGGTCCATAC CTAAGTCAGA CCAGCTGAGG CGAGCATATC TGCCCTGATGG
 4701 CTACGAGAAT GAAAATCAGC TCCTCAGCTC ACAAGATTGT TGACGAGGCA
 4751 GTTAATGACC ATTGTGATCA AGATAAATAA TGTGGAGTAT TAAAGTTAAT
 4801 GTGTTGATTG TGTGGTTcAT TTTTATATTT ATTTTAAATC ATGTGATACA
 4851 GAATAGTTTT GCAATGTGTA TATAGTTTTA GGCAAAAGAA AGAAACCTCA
 4901 CTGCAAAACT TACTGCTAAT TTTAAGTCAC CGATGGTGTG AAGTAAAACT
 4951 TACCTATGGA GTGTATTAGG CTATCTGAAA CCCGGTGGTT ATCGTTAAAA
 5001 CTCATGGATG tTCTCctGAA ATGTTTTGTGc ATGGAAAGAA TTCTTTGCAT
 5051 TGTAGCCCTG TTGCCATGTA TGgTTTTTGT TTTtTCTTGT GAGGTTACGT
 5101 AATGATTTGG ATTGAGAACT AGCCTAGAAA ACTGAAGCTG CTGCCAGGCT
 5151 ATACCTCTCA CAGGACCTTA AAGAAGCTCA CATTCTTTTC AAAGTGGAAT
 5201 AGTAGGCCGG AAATGTTCTT AGAGCTTGAG ATCTCTAGGA GAAATGACTG
 5251 tTCAAATTAC ATTCTACTTG TATGTGAAAT AATTGCTTAA CCTTCACTTC
 5301 CTTGTGGtCG CTTTTAAAGA CATATTTAGT AGCAGCTTCA AGTGACTAAA
 5351 AGACTAAAA CTCTCTTAGT GTCTTTGACA AAAGCCATGG GGGATTTTGC
 5401 AGtGATGAGA TTTTAGTCTT CAACTATGAA CAAAAATTTT CaAACCATAG
 5451 TTTTCATGTT AAGTAGCATC CTGGATTTGC AAActtTTTT ATGCTTTGGT
 5501 ATTAAGACTA TTAAGACAGT CTTGTAGTCT TAATTTAAAT TTAAATACAC
 5551 TGTGGTTCTG TTTTCTCTTT TTACAGGGAA TTGGTTACTT TTTGGCTGAC
 5601 AACAAATAAG AAAATTGCTT TCTATACATT AATTTTTCCA TAAGATTTTA
 5651 TAAGAGTTTA TTTAGGTAAA TAGACTAGAA ATGAGTTTTA AAAGTTTCAC
 5701 CTTTAcTGCT TTGCCTCTGG TGATGCAGGT TTTAAAACTT GGATGCTAAA
 5751 CAGTGCAGAA AACGAGGCAA GAACTTTATA GTAAAAATAA GCAGGATGCT
 5801 GAGTCAGGAA ATAATCAGCT CTGTCAAAGG TCAACCGTGT TTAGGTTTTc
 5851 TACACCTCAC TcTTCTGTAA GCTTATTAGG AGTCATTTAG CTCAGAGAAG
 5901 TCCTcTTCAT CAAATCCGGA TTATATTTTT TAAAAGTTTA GTAACGAAG
 5951 ACTTAgATTA GAGCTTTAAT ATTAAAAAA AATCAAAATAG TTGAAAAACAT

6001	TGAAGTTATA	TTTCTATTCT	AAATCAGTGG	CAAGATGCCC	TGATTGTGAC
6051	TtGGAGTTAT	CTTTGAGGTC	ATTGTGCCAC	AAATTTCCAT	TAAACAAAAGC
6101	CAAAGTTAAT	AGAAAATAAT	TTGTGGGTTT	TGTGCTTTGA	TTTTTTTTTTT
6151	CCCCCGTCAC	AGGGTGTTTC	TACATAGCCC	TGGCAGTCCT	GGAACTCACT
6201	AGAGGCTAGC	CTCAAGACCC	ACCTGCCTCT	TTCCCTCGTG	AGTGTGGGA
6251	TTAATGGCGT	GTACTATCAC	GCCCAGCAAT	AATTTGTGTT	TTTAAGAACA
6301	TCTTGTCTTA	TGGGTCTTAC	AGATTTAACA	TTCCCAATTC	AGGTTGTCAA
6351	AGTGACCACT	TGTTTTTATT	TTAAAAGCCT	GCTATGGAAT	TTAAAAAAA
6401	ATaGtTCTCA	AAAATTGCAG	TTTCTATTtT	CTCACGTTTG	GtTATATTCA
6451	TGTAGCACAT	AAATGCCTAG	ACTAGTTGTT	TACTAACTGC	AATTGGTAAA
6501	ACACAGACCA	GCAGCAAAGt	AAACAATGAT	TGCTTCCACT	TCCTTAAGGA
6551	ATTTGCGGGC	AACCATGTTT	AAATTGCCTA	AGATGGTAGG	ATATCTAAGC
6601	AACTAAAATA	ATTGAGTTTT	GTACCACTAT	TTTCAGAAgA	AAAACGATCA
6651	TTTAATTTTC	CAAGATTCCCT	TGTGATTCCG	GCTATTCCCT	ATTGTCTGGG
6701	ACTATATGCG	CTGTCAGTCA	ACATGGCAGA	ACAGCTTCTA	GCAGTCgAGT
6751	AACATGTTGG	ACGGTGCAC	AGGAATATAA	TTCACCTTTG	CCTTTCTTAC
6801	AGTTATTACT	GATTTACCG	TTACTTAGTT	CCTCTAGAAC	TGTTTTTCCA
6851	GGCTTATTTT	TGAATCTCTA	GGGGCAGAGT	GCATTCTGTT	AGGGTGTAT
6901	CATGTCCTTA	GAATGAGGTT	TCCGTCCCCT	TGAGGGAAAAG	TTAGTAGTAA
6951	TGTTTCTTCA	CCGATTATGA	ACATTTGTTT	CTTGGAACTG	GTAAATTTG
7001	TCTCGTAAAT	AATAaCAGTG	GTTATAGTAG	CCTGTGCAAA	CTgGTAAGA
7051	TTTGCaAAAC	TATGATTGGC	ATGTTAGATA	ACAAATGAGA	AGTGTCTAAG
7101	GCGCGTGTCA	TGTGAGAGAG	GCAAAGTGAA	CTTATGTACA	GAAGCTTTCA
7151	ATCCTCCTtA	TATAGTGGTA	GTTGATGCTT	TACTATTGGA	CCAATAGCAG
7201	CTCCTGTTAA	TTAGTTGGAT	GCATTATCAC	AAGCTTTGTA	ATGTTTCATT
7251	TATAAACTGC	CTTCAGCACA	TGCGTACCTG	TTAATGtTTT	CTCAGTGTcG
7301	TAgGATGTCT	TCCAGACTCT	CACTGTATCC	TGCCAGTAAA	TAGTTCTAGC
7351	TTGTTGAATT	TCCTGCTTAT	TTTTTATTAT	CTGTGGCATA	TAGCAAGCCT
7401	AAAAGATAAT	TGTAATCATT	TTTTTTTTCTT	TTACAGAGTG	AGAAGTTAGA
7451	TTGTGTATGA	AAGCCTATAG	TTATTTTATC	CATTATCTTT	GATTAAGGCT
7501	TTCTAATATT	ATTTAGTAAG	TGTTCCCTCC	AGTTAATGCC	TGTGGCCCTA
7551	CTGAATTTTA	AGGGAACAGa	CAACTTTTAA	GAAATAAACC	TTACTGTTTT
7601	ATACAGTAAG	TTTGTCTGAT	TGGGAATGTT	CTGAAGTTGG	ACATACAATT
7651	GGCCAGATTC	CCAAGCTAAA	ACCTTGGCCT	GAAAGGCTGT	GTACATGAAT
7701	GTTAACAAGT	ACAGGAAAGC	AAATTAAACT	TAAGGtGACC	TAAAACAAAG
7751	GCTAGAAACA	AGTTTTTATT	GGATTCTTAG	TCTTTGTAAG	GGCTTTCTGT
7801	TTGACTTACC	TCAACTGCCA	TTGAGTGAGG	GTAAATCACC	TTAAAAATGAC
7851	TGATTTACCT	GGAAGGGTAG	TTGTAGTTTC	CATCCTTAGC	TTGCTTTTGC
7901	TTTGGTCGGC	ACTGATGAAA	TGTGAAGTGG	CTATTCTAGA	CCTTGGTTGA
7951	TCTCAGCCCC	TACTCCCAGA	CTGCTGAGGT	AACTACAGTC	CGTTACCTAG
8001	TGCTATTCTC	CTTGTATCCA	GAACACTATT	AAAATTTTAT	TATGTTAATT
8051	TTCCGTACTCC	TGTGCTTCTT	TTTTACTTTT	GCTCATGCAC	GTATTTTAGT
8101	ATGAATGTGT	TGCTTTTTTA	AAAGTTTGTG	ATATTTAGAA	ATAAATTTCA
8151	AAAAAGTTTC	TAAAGTTTTT	GAAATCATAT	TTTGTTTTGT	GAGCCAGTGA
8201	CTGACATTTT	ATAATTTCAA	GGTTTTACCT	GCTGAATTTT	GATTTTTCTA
8251	CCAATCATTT	ATTTGTAGTG	TTAACCTTTT	AAGCTACTGT	GACATCTAGG
8301	CTTCAAACCTG	CCCTGTTTGA	AGTCTTCCAA	AACACTAGCC	AGTGGGTAAG
8351	TTCAGTCAAC	ATTTTAATGG	GAACATTGTA	CTTCTACAAG	TCTTATGAAA
8401	GGGCTAGGAT	GTAGCTTAGT	GGTTACCCTG	CAACCTATAA	CTTCAGCTCC
8451	AGAGGATCTG	ATACTCTTAA	GGATTTCTCA	GGCACCAGCA	TGCAAGTGCA
8501	TATAAATTCA	CACAGGCACA	CATAAACAAA	TCTTCAGAAA	TGTCATGAAC
8551	ATCAGTTGAT	CACAAGCAGG	ACTGATCGTT	CTAAAATATA	TATATATTTA
8601	CCTCTTGTTT	ACCTCTTAAT	TTGTTTTCAA	AACTCTCCAA	CTGAGACAAT
8651	AAGATTATCT	GCTAAATTAT	TACATGGCTT	GGCATGTCTC	TGCTCCACT
8701	CACGTTTGTG	CAAGAGAACA	TTGAATTATA	GTGTTTTATT	TTCTGAAACC
8751	AAAATGTTAG	AAAAGTTCTT	AGGAATCAAC	CCATACTGTT	CAGTTCTATG
8801	TTAGAGAACT	GTCTGGGTTT	TCTAAAGAAC	CTCAGTTTAT	TACAACTACA
8851	GTAATTAcTA	AGCAATGAGC	CTCTCCAATA	TATATATAAC	AGAAATGCCT
8901	CCCCTCAGTA	CCCAAGTACA	ATGGCATGCT	GCAAGAAGAT	ATGGCTTTGT
8951	TTcTAAGcTA	AACTATAAGC	ATTTCTCATG	AGACTTGcTA	CCAGCCAAGT
9001	TCTTGTCCAG	ATCCTCAAAA	AGACTTGTTT	TTTTAAACCT	GCAAAAATG
9051	AATATcTGCA	GAAATGTACC	TTTTTAgAAAT	cTGGTATTCT	ATATCACcTA
9101	TAGAATATTG	CTTTCAGATT	AAAATCTATC	ATGCAGTGAC	ATTGAAATCC

9151	TTGCAAAGAA	CCTGcTATAA	GAATTTTAAC	AAATTGTGCA	CAAAAGGCAA
9201	ACACTTATTT	GACTCTTCAG	TATGAAAGAC	aCAAGTGTAT	GTAGACAAAT
9251	ATTAATCATT	ATCATAACCC	AAGCCTCTGG	GTTTATATGT	ATATCATAAT
9301	AAATGTCTAA	AAGTTTTcTA	cTTAAGTATG	AAAGCAGTGC	TTTACTcTGC
9351	ACTcTGTctT	GcTTCTCCT	CTCCAACCCC	CACCCCACCC	CGTTAACCA
9401	TGCTACTTAA	TACCTATTCA	ATAATAATAA	TGATGGAGGC	TGAGAGATGG
9451	CTCAGTGTTC	ACAAGCACTT	GCTCTTGCAG	AAGACTAAAAG	TTTAATTTCCC
9501	AGCATCCAAA	ATACACATAA	CTCCAGTTCC	TGGAACCCAT	CGCACTGTTT
9551	TGACCATCAC	AAGCTCTGCA	TGTACAGGGT	GTACACACAG	AAAAGCAGTC
9601	ACATGCATAA	GAACAAAATT	TAAAGTACAA	TGTTATATAT	TACTTTTAAT
9651	CTAGTAAATT	CCCCCTTTTG	GAGTGTATCC	TATGAATGCA	AATAGCAAGT
9701	GTAAAATATA	TGTA AAAAGGA	GGCATTGTG	GTGGGGAAAAG	GAGAAGGGGG
9751	AGACAGTCTT	GTGTCAACTT	GATTACAAAC	TAGAGTTGAT	CTGAATGGAG
9801	GGAAACTGAG	AAAATGCCTT	CATAAGATCC	AGCTGTAAAG	AATTCtTAAC
9851	TGACTGATGG	AGGTCCCAGT	CCATTGTGAG	TGGTTACATC	TCCCTGGACTG
9901	TGGGCTTCAG	TTCTATAAGA	AAGCAGACTG	TCTGCCTGGT	CCACCTGGAG
9951	CAGCAGAGGA	AAGCTGCACA	CTATCCTTTC	GTCTCCATCA	TCTGGCCAC
10001	AGCCATGTCT	GCCTTCCCAA	AAAGGCCTGC	AGGCCCCAAA	AACATCCAGG
10051	TCCATAGCTC	TGTCTGCCTT	CCTCGGGAGT	CCTGTCAGCA	TTAGGAACAT
10101	TCAGTCAACA	ACAGGGACAA	CCAGATGGCT	TAAAGCCAGC	ATGAGAACAT
10151	AATCAATCAG	AGCCAGGACA	ATAAGGGACC	ACCAGAGCAC	AGCTATCCTA
10201	CTATAGCAAG	CAAGCACTGG	ACATCCCTGG	ATACTCTAAC	ACAGCTGAAG
10251	CACAAGAAGG	TGACCTTAAA	TTGAGGATGA	TACAGGCCTT	TAAAGACCAT
10301	TGAAGAAACT	ATAAATCCCT	TAAAGAAATA	CAGGAAAAGTA	AAATCAAACA
10351	GTTGACGGAA	ATGAATAAAG	CTGTTTAAGA	ACTGAAAAAT	GGAAATTGAA
10401	GTAATAAAGA	GAaCAATGAA	GGAACCTGG	GTCTGGAAAA	CTTAGGGAAG
10451	AGAACAGGAA	CTATAGATGC	AAGCATCACT	AATAGAATAT	AAGAGATGGA
10501	AGAGAGAATC	TCAGGCATAG	AAGGTAATAT	TGAAGAAATC	GATATATCTG
10551	TCTAAGAAAA	GTTAAATCTA	AAAAGTTCTT	GACACAAAAC	ATCCAAGAAA
10601	TCGCTAACAA	TATGAAAAGA	CCAAACTTAA	GAATAATAGG	AATTgAGCCA
10651	tGCaTGGTGG	CACATGCCTT	TAATCCTAGC	ACTCaGGAGg	CAGAGGCAGG
10701	cgGATTTCTG	AGTTCaAGGC	CAGCCTGGTC	TACAgAGTGA	GTTCcAGGAC
10751	AGCCAGGGCT	ATACAGAGAA	ACCCTGTCTC	GAAAAAACAA	AAAAcAaAAa
10801	AAAAAAaCaA	aaAAAAAaaa	AAaAaAaGAA	TAATaGGAAT	AGAaAATGGA
10851	GAAGATTCCC	ATCCTTAAGG	cCTAGAAAAAT	AGTTCCAACA	AAATCACAGA
10901	AGAAAATTTT	CCCAACATAA	GAAAAGACTA	TGCCTGTAAA	CACACAAgAA
10951	GCTTTCATGC	ACCAAATAGA	TTGGACCAGA	AAAGAAAAAC	CTCCCTTCAC
11001	ATAATAGAAT	ACTAAATCTA	CACAACAAAG	AAAGAATATT	GAAAGTGGCA
11051	AGGGGAAAGG	CCAAGTAACA	GATAAAAAGCA	GACCTATCAG	AACTATACCC
11101	TACTTCTCAA	CAGAGACTCT	AAGACAGATA	GGCCTAGGCA	GATGTCTTCT
11151	AAAACCTGAG	ACCaCAGATG	TCAGTCCAGA	CTACTATTCC	CAGAAAAACT
11201	TTCAGTCATC	ACAGATGGAG	AAAACAAGAT	ATTCCATGGC	AAAAGTAAAT
11251	TTAAACTATA	TTTATCTACA	AATCTAGCCC	TACAGAAGAT	ACTAGTAAGCC
11301	AGGCTTGGTG	GCACATGCCT	TTAATCCCAG	CACTCGGGAG	TCAgAGACAG
11351	GGAGATTTCT	GAGTTCCAGG	ACAGCCTGGT	CTACaAgAGTG	AGTTCCAGGA
11401	CAGCCAGGGC	TATACAGAGA	AACCCTGTCT	TGAAAAAAAc	cAAgAgGGGG
11451	GGAGGGGGGG	AGAAGATACT	AaAAGGAaAa	TTCCAACCCa	AAGAGGTGAA
11501	CAGcTCTCAa	GGAAACaAaG	GTAaTGAATC	ACCTcaCACC	AGCaaACCaa
11551	aGGAAGGaaA	ACACACACAC	ACACACaCAC	aCACACACAC	ACTCACACTA
11601	cTaTACTACTG	TaCTGCTACT	GcTACTACTA	CTACTACTGC	TACTACCATC
11651	AAAAGAACAg	AAATTAACAA	TCGTTGGTAA	TTAATATcTc	TTAACATCAA
11701	TGGACTCAAT	TCCCCAATTA	AAAAATACAg	GCTAACAgAA	TGGGTGAgAA
11751	AACAGGATCC	ATACAAGAAA	CACACCTAAG	CAACAAAAGAT	AGACATTATC
11801	TCAGAATAAAA	GGGCTGGAAA	AAGATTTTCC	AAGCAAAACAG	ACCCAAGAAG
11851	CAAGTTGGAG	TAGCCATTTT	TTTAAAGATC	TATTTATTTT	ATCTATAGGA
11901	ATACACTGTA	GCTGTCTTCA	GACACACCAA	AAGAGGACAT	TGGGTCCCAC
11951	TACAGATAGT	TGTGAGCTAC	CAAGTGGTTG	TTGGAAATTG	AACTCAGGAT
12001	CTCTGGAAGA	GCAGTCAGTG	TTcTTAACCA	CTGTGCCATC	TCTCCAGCCT
12051	CTGGAGTAGC	CATTCTAATA	TCTTATAAAA	TAGACTTTCA	ACTAAAAATTA
12101	ATCAAAACAT	ATGGGGGAAGA	ACATTTTATA	CTCATCAAAG	AAAAACTACA
12151	TGAAGATGAC	ATcTTAATTc	TGAACATCTA	TACCCAAAAT	CAAGTACACC
12201	CACAGTTGTA	GAAGGTAAAT	TACTAAAGGT	TAAATCATAT	ATTGAGCCCC
12251	ACACAGAAAT	CGCTGTGGAA	ACTTAAGCTG	AATAAACCCCT	TTCTCCAAC

12301 ATGAAGTAAT TTATTAAAAC CAGTTTGCAT GACCATGCTA TCCAAATAAA
12351 AAGAATAACA aGTTACATAA TTCATCCTCT TTATTATAGA AATATTAAAC
12401 AtTGGTTTTT TTTGtTTGTT tGTTTTGtTG TTTTGGGGtT GGGtTGTTt
12451 TTTTTtTTT TTTTTttTGG TTTTTTTTTT TGAGACAGGG TTTCTCTGTA
12501 TAGCCCAGGC TGTCCCTGAAA CTCACTTTTG TAGACCAGGC TGGCCTTGAA
12551 CTCAGAAATC TACCTGCCTT TGCCCTCTGA GTGCTGGGAT TAAAGTCGTA
12601 TGCCACCATG CCCACCTAGT ATTAAACATT GTATATATTT AATTTGTTAA
12651 TTTGTATTAT aAAATCTTTT GGGGGAGCCA GggTgTTCat GACAGaGTTT
12701 CTCTGTgTAG CCCTGGCTGT CCTGGAActC ACCCTGTAGC CCAGGTGGC
12751 CTTGAACTCA GAAATCCACC TGCCCTCTGCC TCCCAAGTGC TGGGATTAAA
12801 GATGTGGGCC ACCACTGCCC AGCATATTAC AAAATCTTAA TTGACAAAAA
12851 TATCTGTAaa CATTCTTTTT TTTTTTTAAA GATTTATTTA TTTATTATAT
12901 GTAAGTACAC TGTAGCTGTC TTCAGACACT CCAGAAGATG GCGTCAGATC
12951 TTGTTACAGA TGgTTGTGAG CCACCATGTG GTTGCTGGGA TTTGAACTCT
13001 GGACCTTCAG AAGAGCAGTA GGGTGCTCTT ATCCaCTGAG CCATCTCAC
13051 AGCCCTCTG TAAACATTTT TATACAATTT AGTTACACAT TTGTAAAAGTT
13101 AAGCCTTTAT TGCTCAAGTT ATAATTAAAT AGGTAAAGGG GAGTCAATTT
13151 CAAATTTATC TTTAGACTTG TTATTCTGTT CATAACCCAG AACTTCCAAT
13201 TGTTTACAGT ACTTGTATAG TAGTCTAAAA TAATTCAGTG GTTCTTATAA
13251 CCAaCAGAAG TTGAACCACT GCTAAAGCCT TAGAGACACA TTTTAAAAATC
13301 TGAAAATTAA TAGAAACAAG ATGAATGTCA CACAATACCA TTGAATAAAT
13351 CTTTATAACT TGCATAACT TCTATAAATA TCCAATATAT ACAAAGAACT
13401 CAAGAAGCTG AACTCCAGAA AAATCAAATA ACCCATTA AAAAAATGGTG
13451 TACAGACTA AACAAAGAAT TCTCAACTGA GAAATACCA AATGGCTGAAA
13501 AGCACCTGAA AAAATGTTCA ACATCCTTAA TCATCAGGGA CATGCAAATC
13551 AAAACAACCC TGAGATTCCA CCTCACACCA GCCAGAATGG CTAAGATCAA
13601 AAATTTAGGT GACAGCAGAT GCTGGCAAGG ATGTGGTGAA AGAGGAAAAAC
13651 TCCTCCATTG TTGGTGGGAT TGCAGGCTTG TACAACCACT CTGGAAATTA
13701 GTCTGGTGGT TCCTGGTTCT GGTGGTGGAC ATAGTACTAC CGGAAGATCC
13751 AGCAATACCT CTCCTGGGCA TATATCCAGA AGATGTTCCA ACTTGTAATA
13801 AGAACACATG TTCCaCTATG TTCATAGCAG CCTTATTTAT AATAGCCAGA
13851 AGCTGGAAAG AACCAGCTG TCCCTCAACA GAGGAATGGA TACAGAAAAAT
13901 GTGGTACATT TACATAATGG AGTACTACTC AGCTATTAAA AACAATGAAT
13951 TTATGAAATT CTTAGGCAAA TGGATGTATC TGGAGGATAT CATCCTAAGT
14001 GAGATAACCC AATCACAAAA GAACACACaT GATATGCACT CACTGATAAG
14051 TGAATATTAG CCCAGAACT TAGAATACCC aAGATACAAT TTGCAAAAACA
14101 CATGAACTC AAGAAGAAGG AAGACCAAAG TGTGGATACT TTGTTCTCTC
14151 TTAGAATAGG GAACAAAATG CCCATGGAGG GAATTACAGA GACAAAGTTT
14201 GGAGCTgAGA CGGAAGGAAG GATCATCCAG TGACTACCCC CcCcCCaCaC
14251 CCCaCCCGGG GATCCATCCC ATAAaCAGCC aCCaAACCCA GACaCTATTG
14301 CATATGCCAG CAAGACTTTG CTGACAGGaC CATGATATAG CTGtCTCTTG
14351 TGAGGCTCTG CCAGTGCCTG GCAAATACAg AAGTGGATGC TCaCAGTCAT
14401 CTATtGGATG GAACaCAGTA TAAaaCAGGG CCCCCAATGG AGGAGCTAGA
14451 GAAAGTACCC AAGGAGCTGA AGGGGTcTGC AACCCtATAG GAGGAaCAaC
14501 AATGTGaACT AACCAGTaCC CCCAGAGCTC

I.4 Plasmidkarten m20-9 und m20-10

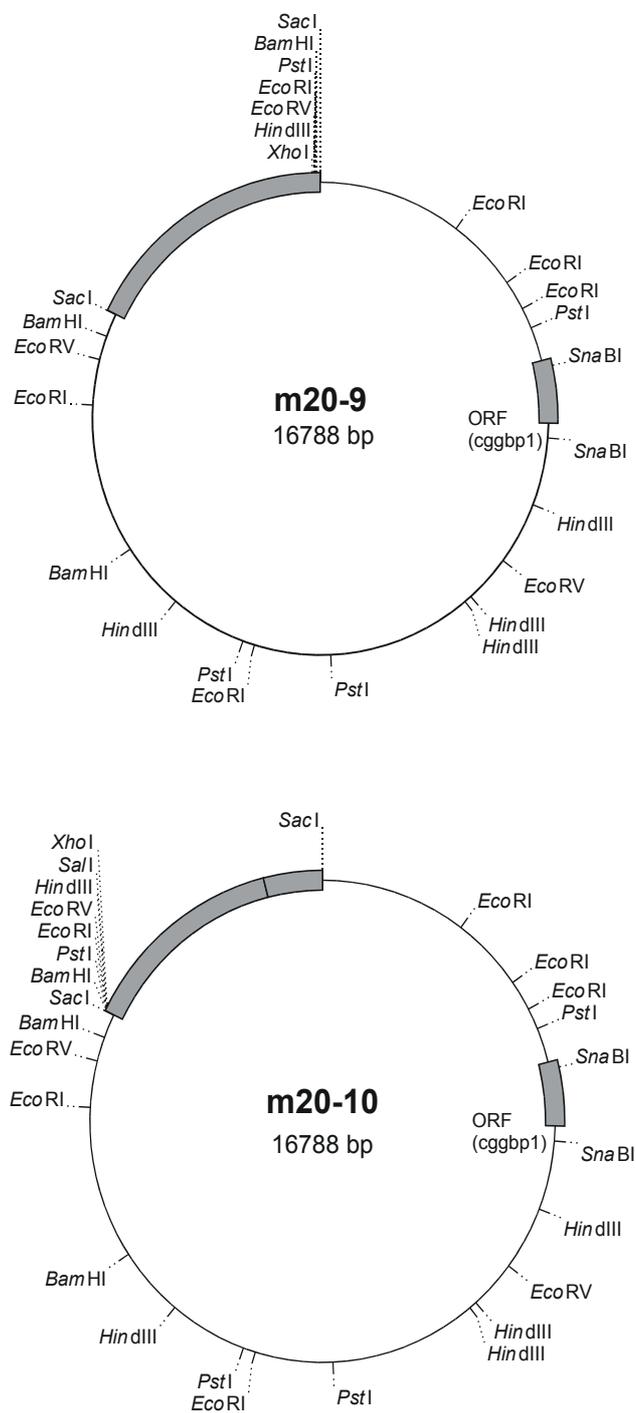


Abbildung 42 Plasmidkarten m20-9 und m20-10

I.5 EST-Klone

Ratte

aa818700	at006274	ai234918	aw141947
aa956676	be116174	bg672621	bq193203
ai113032	bf556503	bi299882	bq195795
ai705483	bg668182	bi302853	bq202932
ai709726	bg373237	ai602770	h32354
ai043734	at006274		

Rind

aw655036	bf601211	bm434423	bg692301
aw660355	bf601520	bm445532	bf041197
be486362	bf601819	bm446344	bf039502
bf045147	bf605085	bm433721	bi681463
bf072994	bi681464	aw356324	aw487673
bf600319			

Schwein

aw354157	bf710955	bi337251	bf193335
bf193334	bf711880	bm659775	bi186614

Huhn

bu109329	bu234304		
----------	----------	--	--

Pferd

bm781074

Hund

bu745953	bu745954	bu751134	
----------	----------	----------	--

Rhesusaffe

cb228654	cb309042	cb230595	
----------	----------	----------	--

I.6 Quellinformation der Northern Blots der Firma Clontech

Die Quellinformation der Northern Blots wurde, soweit verfügbar, hier aufgelistet. Es wurden nur gesunde Gewebe verwendet.

Gewebe	Geschlecht	Alter (Jahre)	Ethnischer Hintergrund	Todesursache	Anzahl der Individuen
Bauchspeicheldrüse	männlich	19–75	kaukasisch	plötzlicher Tod	9
Nebennierenmark	"	17–68	"	"	65
Schilddrüse	"	8–78	"	Unfall	19
Nebennierenrinde	"	22–67	"	"	25
Hoden	"	17–61	"	"	19
Thymus	"	19–35	"	"	26
Dünndarm	"	25+30	"	plötzlicher Tod	2
Magen	"	23–61	"	"	5
Kleinhirn	männlich/ weiblich	16-70		"	24
Großhirnrinde	weiblich	35	"	"	1
Stammhirn	männlich/ weiblich	16-75	"	"	110
Rückenmark	"	15-66	"	"	49
Okzipitallappen	"	22-70	"	"	15
Frontallappen	"	16-75	"	"	26
Temporallappen	"	"	"	"	26
Putamen	"	"	"	"	110
		Alter (Woche)			
fötale Gehirn	männlich	21–30	"	spontaner Abort	10
fötale Lunge	"	20–27	"	"	38
fötale Leber	"	18–24	"	"	32
fötale Niere	"	21–30	"	"	7

Mausembryonen	Mausstamm	Gewebe	Anzahl der Mäuse
Tag 7	BALB/c	ganze Embryonen	mehrere Tausend
Tag 11	"	"	"
Tag 15	"	"	"
Tag 17	"	"	200

I.7 Liste der zitierten Internetseiten

<http://www.amaxa.com/products/protocols>
http://www.expasy.org/cgi-bin/ft_aligner?psa=PS50171 (Interpro-Datenbank)
<http://www.ensembl.org>
<http://www.fugu-sg.org/>
<http://www.hgsc.bcm.tmc.edu/projects/rat/>
<http://www.ncbi.nlm.nih.gov>
<http://www.genomatix.de/>
<http://www.ncbi.nlm.nih.gov>
<http://www.ncbi.nlm.nih.gov/BLAST/>
<http://repeatmasker.genome.washington.edu>
http://www.sanger.ac.uk/Projects/D_rerio/

I.8 Zugriffsnummern der zitierten U1-Zink-Finger-Proteine

Unter cross reference sind die Links zu den beschriebenen Proteinen angegeben.

<http://www.expasy.org/cgi-bin/prosite-search-ac?PS50171>

CIZ1_HUMAN (Q9ULV3), MAT3_HUMAN (P43243), MAT3_RAT (P43244),
RU1C_HUMAN (P09234), RU1C_MOUSE (Q62241), RU1C_XENLA (Q03369),
RU1C_YEAST (Q05900), S3A2_HUMAN (Q15428), S3A2_MOUSE (Q62203),
S3A3_HUMAN (Q12874), S3A3_MOUSE (Q9D554), PR09_YEAST (P19736)

Der Sequenzvergleich der Zink-Finger-Proteine wurde folgender Internetseite entommen:

http://www.expasy.org/cgi-bin/ft_aligner?psa=PS50171

I.9 Kurzzusammenfassung

Das kernlokalisierte 5'-(CGG)_n-3'-bindende Protein (CGGBP1 oder p20) aus menschlichen Zellen inhibierte den Promotor des FMR1 (*fragile X mental retardation gene 1*) Gens indem es spezifisch mit der 5'-(CGG)_n-3'-Wiederholung in der 5'-nicht-translatierten Region des FMR1-Gens interagiert. Es wurde hier die genomische Organisation und die Expression der CGGBP1-Gene des Menschen und der Maus untersucht. Das menschliche Gen bestand aus 4 Exons. Ein 14 kb großes DNA-Fragment mit den 3 Exons des CGGBP1-Gens der Maus wurde isoliert und komplett sequenziert. Das offene Leseraster der CGGBP1-Gene war nicht durch Introns unterbrochen. Aufgrund von alternativer Polyadenylierung wurden zwei Transkripte von 1,2 und 4,5 kb in allen bisher untersuchten adulten und fötalen Geweben des Menschen und in embryonalen Zellen der Maus abgelesen. Das Mengenverhältnis der beiden Transkripte variierte in adulten und fötalen Geweben des Menschen.

Die CGGBP1-Gene des Menschen und der Maus waren hochkonserviert und wiesen Sequenzidentitäten zwischen 78 und 99% sowohl in den translatierten als auch in den nicht-translatierten Regionen des Gens und im Promotor auf. Die Nukleotidsequenzen von anderen Säugetieren und von Hühnern waren in Datenbanken zugänglich und im Vergleich zum menschlichen Gen ebenfalls hoch konserviert. In Transfektionsstudien erwies sich eine 235 bp lange Region der 5'-flankierenden Sequenz des CGGBP1-Gens des Menschen als ausreichend für eine starke Promotoraktivität. Eine *in-vitro*-Methylierung aller CpG-Dinukleotide inaktivierte den CGGBP1-Promotor, der in mononukleären Blutzellen und in HeLa-Zellen komplett unmethyliert war. Außerdem wurde hier gezeigt, daß CGGBP1 an kurze und unterbrochene 5'-(CGG)_n-3'-Wiederholungen ($n \geq 5$) bindet. Um zukünftig die Erforschung der Funktion des CGGBP1-Gens durch Geninaktivierungs-Experimente in der Maus zu ermöglichen, wurden zudem eine Strategie zur konditionalen Inaktivierung des Gens entworfen und der benötigte Geninaktivierungs-Vektor fertiggestellt.

I.10 Abstract

The nuclear 5'-(CGG)_n-3'-binding protein (CGGBP1 or p20) from human cells had been shown to influence the expression of the FMR1 (*fragile X mental retardation*) gene by specifically interacting with the 5'-(CGG)_{n>8}-3' repeat in its 5'-untranslated region. Here, the genomic and mRNA organizations of the human and murine 5'-(CGG)_n-3'-binding protein 1 (CGGBP1) genes were studied to explore the regulation of the CGGBP1 gene. The human gene consists of 4 exons. A 14 kb fragment of genomic DNA with the 3 exons of the murine CGGBP1 gene was isolated and the complete nucleotide sequence was determined. The open reading frames of the human and the murine genes were not interrupted by introns. Due to alternative polyadenylation, mRNAs of approximately 1.2 and 4.5 kilobases were transcribed at varying ratios in human and murine cells as well as in embryonic, fetal and adult tissues. The human and the murine genes, including promoters and large parts of the untranslated regions exhibit 78% to 99% identity. Nucleotide sequences from other mammalian species and of birds were available from GenBank and displayed a high percentage of sequence identity as well. A sequence of about 200 nucleotides upstream of the gene was essential for promoter activity in transfection experiments in HeLa cells. The *in vitro* methylation of 18 CpG dinucleotides inactivated the promoter which was unmethylated in human peripheral white blood and HeLa cells as shown by bisulfite genomic sequencing. Furthermore, CGGBP1 was shown to bind to 5'-(CGG)_n-3'-repeats with $n \geq 5$ and to interrupted repeats. For future analysis of a CGGBP1 knock out mouse, a conditional gene targeting strategy was developed and the appropriate targeting construct was completed.

I.11 Lebenslauf

Name Frauke Naumann
Nationalität deutsch
Geboren 07.04.1972 in Hamburg

Schulbildung

1978-1982 Grundschule in Tübingen und Erlangen
1982-1991 Gymnasium in Erlangen
Juli, 1991 Abitur

Studium

1991-1993 Grundstudium der Biologie an der Heinrich Heine Universität in Düsseldorf
1993-1998 Hauptstudium der Biologie an der Albertus Magnus Universität zu Köln
1996-1998 Diplomarbeit im Institut für Genetik der Universität zu Köln unter der Anleitung von Prof. Dr. W. Doerfler. Thema: "Genomische Organisation des Gens für das 5'-d(CGG)_n-3'-bindende Protein p20 des Menschen"
1998 Diplom

Dissertation

1999-2003 Institut für Genetik, Universität zu Köln

Berufstätigkeit

ab 1999 wissenschaftliche Mitarbeiterin im Institut für Genetik, Universität zu Köln

Köln, den 19.3.2003

I.12 Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich der Tabellen, Karten und Abbildungen - , die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Walter Doerfler betreut worden.

Teilpublikationen:

Gene structure and expression of the 5'-(CGG)_n-3'-binding protein (CGGBP1)

Frauke Naumann, Ralph Remus, Birgit Schmitz und Walter Doerfler

Genomics, 2003, in press

Ich versichere, daß ich alle Angaben wahrheitsgemäß nach dem besten Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obrigen Angaben betreffenden Veränderungen, dem Dekanat unverzüglich mitzuteilen.

Köln, 5.8.03

Danksagung

Mein besonderer Dank gilt Walter Doerfler, der mir die Möglichkeit gegeben hat, in seinem Labor zu arbeiten und diese Arbeit fertigzustellen. Ich bin ihm für die immer konstruktiven Diskussionen, hilfreichen Ratschläge und sein stetes Interesse an meiner Thematik sehr dankbar. Außerdem möchte ich mich dafür bedanken, daß ich wissenschaftliche Tagungen im Aus- und Inland besuchen konnte.

Bei Birgit Schmitz bedanke ich mich für eine sehr gute und produktive Zusammenarbeit im letzten Jahr dieser Doktorarbeit. Auch vorher schon hatten sie und Gudrun Schell mir durch die Durchführung der Sequenzierungsreaktionen viel Arbeit abgenommen. Hanna Mansi-Wothke hat durch das Ansetzen der Zellkulturmedien und durch ihre Umsicht bei der Nachbestellung notwendiger Zellkulturartikel die Transfektionsanalysen sehr erleichtert – vielen Dank!

Den Mitarbeitern der Arbeitsgruppe Rajewsky, v.a. Angela Egert und Manolis Pasparakis danke ich für ihre Anleitung bei der embryonalen Stammzellkultur und den hilfreichen Diskussionen bei der Entwicklung der Geninaktivierungs-Strategie. Der Arbeitsgruppe von Ulrich Müller – und selbstverständlich auch ihm selbst – danke ich für die Kooperation bei der Herstellung der knock-out-Maus - viel Spaß in San Diego!

Tomi Domazet und Arne Nolte aus der Arbeitsgruppe von Prof. Diethard Tautz danke ich für Kommentare und Hilfestellung bei den evolutionären Aspekten in dieser Dissertation.

Ohne die Zuverlässigkeit von Ingeborg Bläser, Elisabeth Jansen und Monika Schmidt, die die Autoklavierungsarbeiten und das Spülen aller Geräte erledigten, wäre vieles schwerer gewesen und hätte wesentlich länger gedauert – vielen Dank!

Mein Dank gilt auch Susanne Scheffler und Petra Böhm für die Unterstützung beim Überwinden bürokratischer Hürden.

Indrikis Muiznieks möchte ich für hilfreiche Ratschläge und Diskussionen während seiner Besuche in unserem Institut danken!

Meinen Laborkollegen Laurence Mangel, Ralph Remus und Jörg Schröer danke ich für die großartige Atmosphäre in unserem Labor – merci beaucoup und herzlichen Dank! Die wissenschaftlichen Diskussionen, der Beistand mit Rat und Tat bei technischen Fragen waren der Anfang von echten Freundschaften. Jörg Schröer hatte mich 1996 in das molekularbiologische Arbeiten eingeführt. Ich verdanke ihm nicht nur einen guten Start! Die Korrekturen des Manuskriptes per FedEx aus USA haben auch sehr viel geholfen! Ralph Remus hat unermüdlich das Manuskript korrigiert und mich mit seinen Kommentaren sehr oft zum Lachen gebracht – ein Segen beim Übertragen der Korrekturen. Vielen, vielen Dank!

Auch Marianna Hösel ist eine sehr gute Freundin geworden. Mit ihre unendlichen Hilfsbereitschaft im Laboralltag hat sie sehr zur Fertigstellung dieser Dissertation beigetragen. Im Endspurt der Dissertation war sie einfach unersetzlich – Bolshoe spasibo!!!

Dennis Webb hat – mit unverwüthlich guter Laune - meine Computerprobleme zu allen Tages und Nachtzeiten gelöst, wofür ich ihm sehr dankbar bin.

Den anderen Mitarbeitern der 4.Etage möchte ich mich für eine gute Zusammenarbeit bedanken.

Auch Ewa Kraft – podziękowanie serdecznie ! - und Anika Wiese bin ich für ihre spontane Hilfe und Aufmunterungen bei den letzten Korrekturen der Dissertation sehr dankbar!

Meiner Familie und meinen Freunden und danke ich für ihre stete Unterstützung und tatkräftige Hilfe in allen Lebenslagen. Dieser Rückhalt hat entscheidend zum Gelingen dieser Arbeit beigetragen.

Danke Euch allen!

Köln, 19.3.03